INTRODUCTION TO SMOOTH MANIFOLDS

by John M. Lee

University of Washington
Department of Mathematics

John M. Lee

Introduction to Smooth Manifolds

Version 3.0
December 31, 2000

John M. Lee
University of Washington
Department of Mathematics
Seattle, WA 98195-4350
USA
lee@math.washington.edu
http://www.math.washington.edu/~lee

Preface

This book is an introductory graduate-level textbook on the theory of smooth manifolds, for students who already have a solid acquaintance with general topology, the fundamental group, and covering spaces, as well as basic undergraduate linear algebra and real analysis. It is a natural sequel to my earlier book on topological manifolds [Lee00].

This subject is often called "differential geometry." I have mostly avoided this term, however, because it applies more properly to the study of smooth manifolds endowed with some extra structure, such as a Riemannian metric, a symplectic structure, a Lie group structure, or a foliation, and of the properties that are invariant under maps that preserve the structure. Although I do treat all of these subjects in this book, they are treated more as interesting examples to which to apply the general theory than as objects of study in their own right. A student who finishes this book should be well prepared to go on to study any of these specialized subjects in much greater depth.

The book is organized roughly as follows. Chapters 1 through 4 are mainly definitions. It is the bane of this subject that there are so many definitions that must be piled on top of one another before anything interesting can be said, much less proved. I have tried, nonetheless, to bring in significant applications as early and as often as possible. The first one comes at the end of Chapter 4, where I show how to generalize the classical theory of line integrals to manifolds.

The next three chapters, 5 through 7, present the first of four major foundational theorems on which all of smooth manifolds theory rests—the inverse function theorem—and some applications of it: to submanifold theory, embeddings of smooth manifolds into Euclidean spaces, approximation of continuous maps by smooth ones, and quotients of manifolds by group actions.

The next four chapters, 8 through 11, focus on tensors and tensor fields on manifolds, and progress from Riemannian metrics through differential forms, integration, and Stokes's theorem (the second of the four foundational theorems), culminating in the de Rham theorem, which relates differential forms on a smooth manifold to its topology via its singular cohomology groups. The proof of the de Rham theorem I give is an adaptation of the beautiful and elementary argument discovered in 1962 by Glen E. Bredon [Bre93].

The last group of four chapters, 12 through 15, explores the circle of ideas surrounding integral curves and flows of vector fields, which are the smooth-manifold version of systems of ordinary differential equations. I prove a basic version of the existence, uniqueness, and smoothness theorem for ordinary differential equations in Chapter 12, and use that to prove the fundamental theorem on flows, the third foundational theorem. After a technical excursion into the theory of Lie derivatives, flows are applied to study foliations and the Frobenius theorem (the last of the four foundational theorems), and to explore the relationship between Lie groups and Lie algebras.

The Appendix (which most readers should read first, or at least skim) contains a very cursory summary of prerequisite material on linear algebra and calculus that is used throughout the book. One large piece of prerequisite material that should probably be in the Appendix, but is not yet, is a summary of general topology, including the theory of the fundamental group and covering spaces. If you need a review of that, you will have to look at another book. (Of course, I recommend [Lee00], but there are many other texts that will serve at least as well!)

This is still a work in progress, and there are bound to be errors and omissions. Thus you will have to be particularly alert for typos and other mistakes. Please let me know as soon as possible when you find any errors, unclear descriptions, or questionable statements. I'll post corrections on the Web for anything that is wrong or misleading.

I apologize in advance for the dearth of illustrations. I plan eventually to include copious drawings in the book, but I have not yet had time to generate them. Any instructor teaching from this book should be sure to draw all the relevant pictures in class, and any student studying from them should make an effort to draw pictures whenever possible.

Acknowledgments. There are many people who have contributed to the development of this book in indispensable ways. I would like to mention especially Judith Arms and Tom Duchamp, both of whom generously shared their own notes and ideas about teaching this subject; Jim Isenberg and Steve Mitchell, who had the courage to teach from these notes while they

were still in development, and who have provided spectacularly helpful suggestions for improvement; and Gary Sandine, who after having found an early version of these notes on the Web has read them with incredible thoroughness and has made more suggestions than anyone else for improving them, and has even contributed several first-rate illustrations, with a promise of more to come.

Happy reading!

John M. Lee Seattle

Contents

1
Smooth Manifolds

This book is about smooth manifolds. In the simplest terms, these are spaces that locally look like some Euclidean space $\mathbb{R}^n$, and on which one can do calculus. The most familiar examples, aside from Euclidean spaces themselves, are smooth plane curves such as circles and parabolas, and smooth surfaces $\mathbb{R}^3$ such as spheres, tori, paraboloids, ellipsoids, and hyperboloids. Higher-dimensional examples include the set of unit vectors in $\mathbb{R}^{n+1}$ (the n-sphere) and graphs of smooth maps between Euclidean spaces.

You are probably already familiar with manifolds as examples of topological spaces: A topological manifold is a topological space with certain properties that encode what we mean when we say that it "locally looks like" $\mathbb{R}^n$. Such spaces are studied intensively by topologists.

However, many (perhaps most) important applications of manifolds involve calculus. For example, the application of manifold theory to geometry involves the study of such properties as volume and curvature. Typically, volumes are computed by integration, and curvatures are computed by formulas involving second derivatives, so to extend these ideas to manifolds would require some means of making sense of differentiation and integration on a manifold. The application of manifold theory to classical mechanics involves solving systems of ordinary differential equations on manifolds, and the application to general relativity (the theory of gravitation) involves solving a system of partial differential equations.

The first requirement for transferring the ideas of calculus to manifolds is some notion of "smoothness." For the simple examples of manifolds we described above, all subsets of Euclidean spaces, it is fairly easy to describe the meaning of smoothness on an intuitive level. For example, we might want to call a curve "smooth" if it has a tangent line that varies continuously from point to point, and similarly a "smooth surface" should be one that has a tangent plane that varies continuously from point to point. But for more sophisticated applications, it is an undue restriction to require smooth manifolds to be subsets of some ambient Euclidean space. The ambient coordinates and the vector space structure of $\mathbb{R}^n$ are superfluous data that often have nothing to do with the problem at hand. It is a tremendous advantage to be able to work with manifolds as abstract topological spaces, without the excess baggage of such an ambient space. For example, in the application of manifold theory to general relativity, spacetime is thought of as a 4-dimensional smooth manifold that carries a certain geometric structure, called a *Lorentz metric*, whose curvature results in gravitational phenomena. In such a model, there is no physical meaning that can be assigned to any higher-dimensional ambient space in which the manifold lives, and including such a space in the model would complicate it needlessly. For such reasons, we need to think of smooth manifolds as abstract topological spaces, not necessarily as subsets of larger spaces.

As we will see shortly, there is no way to define a purely topological property that would serve as a criterion for "smoothness," so topological manifolds will not suffice for our purposes. As a consequence, we will think of a smooth manifold as a set with two layers of structure: first a topology, then a smooth structure.

In the first section of this chapter, we describe the first of these structures. A topological manifold is a topological space with three special properties that express the notion of being locally like Euclidean space. These properties are shared by Euclidean spaces and by all of the familiar geometric objects that look locally like Euclidean spaces, such as curves and surfaces.

In the second section, we introduce an additional structure, called a smooth structure, that can be added to a topological manifold to enable us to make sense of derivatives. At the end of that section, we indicate how the two-stage construction can be combined into a single step.

Following the basic definitions, we introduce a number of examples of manifolds, so you can have something concrete in mind as you read the general theory. (Most of the really interesting examples of manifolds will have to wait until Chapter 5, however.) We then discuss in some detail how local coordinates can be used to identify parts of smooth manifolds locally with parts of Euclidean spaces. At the end of the chapter, we introduce an important generalization of smooth manifolds, called manifolds with boundary.

Topological Manifolds

This section is devoted to a brief overview of the definition and properties of topological manifolds. We assume the reader is familiar with the basic properties of topological spaces, at the level of [Lee00] or [Mun75], for example.

Suppose M is a topological space. We say M is a *topological manifold of dimension n* or a *topological n-manifold* if it has the following properties:

- M is a *Hausdorff space*: For every pair of points $p, q \in M$, there are disjoint open subsets $U, V \subset M$ such that $p \in U$ and $q \in V$.
- M is *second countable*: There exists a countable basis for the topology of M.
- M is *locally Euclidean of dimension n*: Every point has a neighborhood that is homeomorphic to an open subset of $\mathbb{R}^n$.

The locally Euclidean property means that for each $p \in M$, we can find the following:

- an open set $U \subset M$ containing p;
- an open set $\widetilde{U} \subset \mathbb{R}^n$; and
- a homeomorphism $\varphi\colon U \to \widetilde{U}$ (i.e, a continuous bijective map with continuous inverse).

Exercise 1.1. Show that equivalent definitions of locally Euclidean spaces are obtained if, instead of requiring U to be homeomorphic to an open subset of $\mathbb{R}^n$, we require it to be homeomorphic to an open ball in $\mathbb{R}^n$, or to $\mathbb{R}^n$ itself.

The basic example of a topological n-manifold is, of course, $\mathbb{R}^n$. It is Hausdorff because it is a metric space, and it is second countable because the set of all open balls with rational centers and rational radii is a countable basis.

Requiring that manifolds share these properties helps to ensure that manifolds behave in the ways we expect from our experience with Euclidean spaces. For example, it is easy to verify that in a Hausdorff space, one-point sets are closed and limits of convergent sequences are unique. The motivation for second countability is a bit less evident, but it will have important consequences throughout the book, beginning with the existence of partitions of unity in Chapter 2.

In practice, both the Hausdorff and second countability properties are usually easy to check, especially for spaces that are built out of other manifolds, because both properties are inherited by subspaces and products, as the following exercises show.

Exercise 1.2. Show that any topological subspace of a Hausdorff space is Hausdorff, and any finite product of Hausdorff spaces is Hausdorff.

Exercise 1.3. Show that any topological subspace of a second countable space is second countable, and any finite product of second countable spaces is second countable.

In particular, it follows easily from these two exercises that any open subset of a topological n-manifold is itself a topological n-manifold (with the subspace topology, of course).

One of the most important properties of second countable spaces is expressed the following lemma, whose proof can be found in [Lee00, Lemma 2.15].

Lemma 1.1. *Let M be a second countable topological space. Then every open cover of M has a countable subcover.*

The way we have defined topological manifolds, the empty set is a topological n-manifold for every n. For the most part, we will ignore this special case (sometimes without remembering to say so). But because it is useful in certain contexts to allow the empty manifold, we have chosen not to exclude it from the definition.

We should note that some authors choose to omit the the Hausdorff property or second countability or both from the definition of manifolds. However, most of the interesting results about manifolds do in fact require these properties, and it is exceedingly rare to encounter a space "in nature" that would be a manifold except for the failure of one or the other of these hypotheses. See Problems 1-1 and 1-2 for a couple of examples.

Coordinate Charts

Let M be a topological n-manifold. A *coordinate chart* (or just a *chart*) on M is a pair (U, φ), where U is an open subset of M and $\varphi\colon U \to \widetilde{U}$ is a homeomorphism from U to an open subset $\widetilde{U} = \varphi(U) \subset \mathbb{R}^n$ (Figure 1.1). If in addition $\widetilde{U}$ is an open ball in $\mathbb{R}^n$, then U is called a *coordinate ball*.

The definition of a topological manifold implies that each point $p \in M$ is contained in the domain of some chart (U, φ). If $\varphi(p) = 0$, we say the chart is *centered at p*. Given p and any chart (U, φ) whose domain contains p, it is easy to obtain a new chart centered at p by subtracting the constant vector $\varphi(p)$.

Given a chart (U, φ), we call the set U a *coordinate domain*, or a *coordinate neighborhood* of each of its points. The map φ is called a (*local*) *coordinate map*, and the component functions of φ are called *local coordinates* on U. We will sometimes write things like "(U, φ) is a chart containing p" as a shorthand for "(U, φ) is a chart whose domain U contains p."

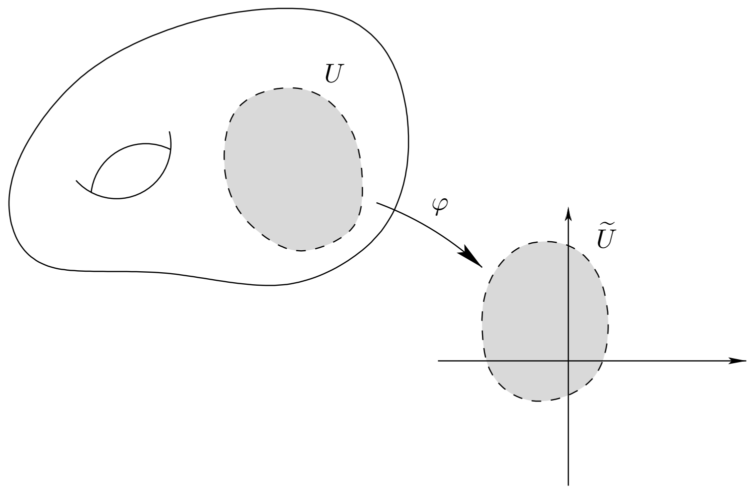

FIGURE 1.1. A coordinate chart.

We conclude this section with a brief look at some examples of topological manifolds.

Example 1.2 (Spheres). Let $\mathbb{S}^n$ denote the (*unit*) *n-sphere*, which is the set of unit-length vectors in $\mathbb{R}^{n+1}$:

$$\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} : |x| = 1\}.$$

It is Hausdorff and second countable because it is a subspace of $\mathbb{R}^n$. To show that it is locally Euclidean, for each index $i = 1, \ldots, n+1$, let U_i^+ denote the subset of $\mathbb{S}^n$ where the ith coordinate is positive:

$$U_i^+ = \{(x^1, \ldots, x^{n+1}) \in \mathbb{S}^n : x^i > 0\}.$$

Similarly, U_i^- is the set where $x^i < 0$.

For each such i, define maps $\varphi_i^\pm \colon U_i^\pm \to \mathbb{R}^n$ by

$$\varphi_i^\pm(x^1, \ldots, x^{n+1}) = (x^1, \ldots, \widehat{x^i}, \ldots, x^{n+1}),$$

where the hat over x^i indicates that x^i is omitted. Each $\varphi_i^\pm$ is evidently a continuous map, being the restriction to $\mathbb{S}^n$ of a linear map on $\mathbb{R}^{n+1}$. It is a homeomorphism onto its image, the unit ball $\mathbb{B}^n \subset \mathbb{R}^n$, because it has a continuous inverse given by

$$(\varphi_i^\pm)^{-1}(u^1, \ldots, u^n) = \left(u^1, \ldots, u^{i-1}, \pm\sqrt{1 - |u|^2}, u^i, \ldots, u^n\right).$$

Since every point in $\mathbb{S}^{n+1}$ is in the domain of one of these $2n+2$ charts, $\mathbb{S}^n$ is locally Euclidean of dimension n and is thus a topological n-manifold.

Example 1.3 (Projective Spaces). The n-dimensional *real projective space*, denoted by $\mathbb{P}^n$ (or sometimes $\mathbb{RP}^n$), is defined as the set of 1-dimensional linear subspaces of $\mathbb{R}^{n+1}$. We give it the quotient topology determined by the natural map $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{P}^n$ sending each point $x \in \mathbb{R}^{n+1} \smallsetminus \{0\}$ to the line through x and 0. For any point $x \in \mathbb{R}^{n+1} \smallsetminus \{0\}$, let $[x] = \pi(x)$ denote the equivalence class of x in $\mathbb{P}^n$.

For each $i = 1, \dots, n+1$, let $\widetilde{U}_i \subset \mathbb{R}^{n+1} \smallsetminus \{0\}$ be the set where $x^i \neq 0$, and let $U_i = \pi(\widetilde{U}_i) \subset \mathbb{P}^n$. Since $\widetilde{U}_i$ is a saturated open set (meaning that it contains the full inverse image $\pi^{-1}(\pi(p))$ for each $p \in \widetilde{U}_i$), U_i is open and $\pi\colon \widetilde{U}_i \to U_i$ is a quotient map. Define a map $\varphi_i\colon U_i \to \mathbb{R}^n$ by

$$\varphi_i[x^1, \dots, x^{n+1}] = \left(\frac{x^1}{x^i}, \dots, \frac{x^{i-1}}{x^i}, \frac{x^{i+1}}{x^i}, \dots, \frac{x^{n+1}}{x^i}\right).$$

This map is well-defined because its value is unchanged by multiplying x by a nonzero constant, and it is continuous because $\varphi_i \circ \pi$ is continuous. (The characteristic property of a quotient map π is that a map f from the quotient space is continuous if and only if the composition $f \circ \pi$ is continuous; see [Lee00].) In fact, φ_i is a homeomorphism, because its inverse is given by

$$\varphi_i^{-1}(u^1, \dots, u^n) = [u^1, \dots, u^{i-1}, 1, u^i, \dots, u^n],$$

as you can easily check. Geometrically, if we identify $\mathbb{R}^n$ in the obvious way with the affine subspace where $x^i = 1$, then $\varphi_i[x]$ can be interpreted as the point where the line $[x]$ intersects this subspace. Because the sets U_i cover $\mathbb{P}^n$, this shows that $\mathbb{P}^n$ is locally Euclidean of dimension n. The Hausdorff and second countability properties are left as exercises.

Exercise 1.4. Show that $\mathbb{P}^n$ is Hausdorff and second countable, and is therefore a topological n-manifold.

Smooth Structures

The definition of manifolds that we gave in the preceding section is sufficient for studying topological properties of manifolds, such as compactness, connectedness, simple connectedness, and the problem of classifying manifolds up to homeomorphism. However, in the entire theory of topological manifolds, there is no mention of calculus. There is a good reason for this: Whatever sense we might try to make of derivatives of functions or curves on a manifold, they cannot be invariant under homeomorphisms. For example, if f is a function on the circle $\mathbb{S}^1$, we would want to consider f to be differentiable if it has an ordinary derivative with respect to the angle θ. But the circle is homeomorphic to the unit square, and because of the corners the homeomorphism and its inverse cannot simultaneously be differentiable. Thus, depending on the homeomorphism we choose, there will either be functions on the circle whose composition with the homeomorphism is not differentiable on the square, or vice versa. (Although this claim may seem plausible, it is probably not obvious at this point how to prove it. After we have developed some more machinery, you will be asked to prove it in Problem 5-11.)

To make sense of derivatives of functions, curves, or maps, we will need to introduce a new kind of manifold called a "smooth manifold." (Throughout this book, we will use the word "smooth" to mean C^∞, or infinitely differentiable.)

From the example above, it is clear that we cannot define a smooth manifold simply to be a topological manifold with some special property, because the property of "smoothness" (whatever that might be) cannot be invariant under homeomorphisms.

Instead, we are going to define a smooth manifold as one with some extra structure in addition to its topology, which will allow us to decide which functions on the manifold are smooth. To see what this additional structure might look like, consider an arbitrary topological n-manifold M. Each point in M is in the domain of a coordinate map $\varphi\colon U \to \widetilde{U} \subset \mathbb{R}^n$. A plausible definition of a smooth function on M would be to say that $f\colon M \to \mathbb{R}$ is smooth if and only if the composite function $f \circ \varphi^{-1}\colon \widetilde{U} \to \mathbb{R}$ is smooth. But this will make sense only if this property is independent of the choice of coordinate chart. To guarantee this, we will restrict our attention to "smooth charts." Since smoothness is not a homeomorphism-invariant property, the way to do this is to consider the collection of all smooth charts as a new kind of structure on M. In the remainder of this chapter, we will carry out the details.

Our study of smooth manifolds will be based on the calculus of maps between Euclidean spaces. If U and V are open subsets of Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, a map $F\colon U \to V$ is said to be *smooth* if each of the component functions of F has continuous partial derivatives of all orders. If in addition F is bijective and has a smooth inverse map, it is called a *diffeomorphism*. A diffeomorphism is, in particular, a homeomorphism. A review of some of the most important properties of smooth maps is given in the Appendix.

Let M be a topological n-manifold. If (U,φ), (V,ψ) are two charts such that $U \cap V \neq \varnothing$, then the composite map $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ (called the *transition map* from φ to ψ) is a composition of homeomorphisms, and is therefore itself a homeomorphism (Figure 1.2). Two charts (U,φ) and (V,ψ) are said to be *smoothly compatible* if either $U \cap V = \varnothing$ or the transition map $\psi \circ \varphi^{-1}$ is a diffeomorphism. (Since $\varphi(U \cap V)$ and $\psi(U \cap V)$ are open subsets of $\mathbb{R}^n$, smoothness of this map is to be inter-

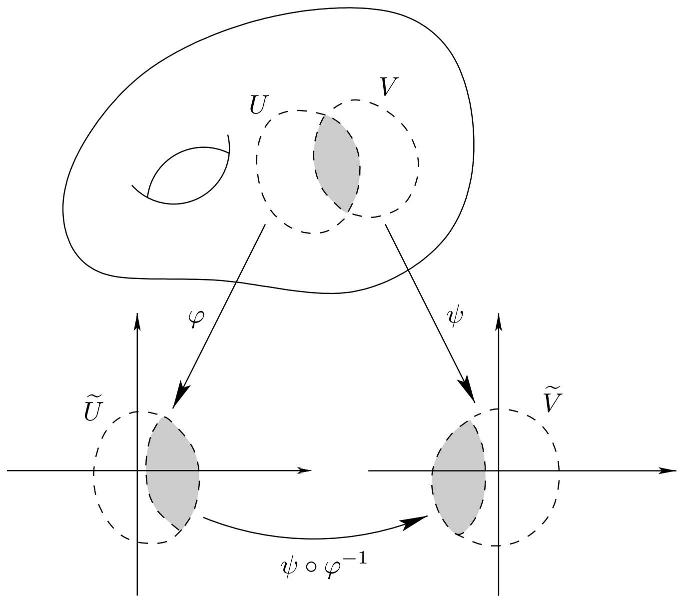

FIGURE 1.2. A transition map.

preted in the ordinary sense of having continuous partial derivatives of all orders.)

We define an *atlas* for M to be a collection of charts whose domains cover M. An atlas $\mathcal{A}$ is called a *smooth atlas* if any two charts in $\mathcal{A}$ are smoothly compatible with each other.

In practice, to show that the charts of an atlas are smoothly compatible, it suffices to check that the transition map $\psi \circ \varphi^{-1}$ is smooth for every pair of coordinate maps φ and ψ, for then reversing the roles of φ and ψ shows that the inverse map $(\psi \circ \varphi^{-1})^{-1} = \varphi \circ \psi^{-1}$ is also smooth, so each transition map is in fact a diffeomorphism. We will use this observation without further comment in what follows.

Our plan is to define a "smooth structure" on M by giving a smooth atlas, and to define a function $f\colon M \to \mathbb{R}$ to be smooth if and only if $f \circ \varphi^{-1}$ is smooth (in the ordinary sense of functions defined on open subsets of $\mathbb{R}^n$) for each coordinate chart (U, φ) in the atlas. There is one minor technical problem with this approach: In general, there will be many possible choices of atlas that give the "same" smooth structure, in that they all determine the same collection of smooth functions on M. For example, consider the following pair of atlases on $\mathbb{R}^n$:

$$\mathcal{A}_1 = \{(\mathbb{R}^n, \mathrm{Id})\}$$
$$\mathcal{A}_2 = \{(B_1(x), \mathrm{Id}) : x \in \mathbb{R}^n\},$$

where $B_1(x)$ is the unit ball around x and Id is the identity map. Although these are different smooth atlases, clearly they determine the same collection of smooth functions on the manifold $\mathbb{R}^n$ (namely, those functions that are smooth in the sense of ordinary calculus).

We could choose to define a smooth structure as an equivalence class of smooth atlases under an appropriate equivalence relation. However, it is more straightforward to make the following definition. A smooth atlas $\mathcal{A}$ on M is *maximal* if it is not contained in any strictly larger smooth atlas. This just means every chart that is smoothly compatible with every chart in $\mathcal{A}$ is already in $\mathcal{A}$. (Such a smooth atlas is also said to be *complete*.)

Now we can define the main concept of this chapter. A *smooth structure* on a topological n-manifold M is a maximal smooth atlas. A *smooth manifold* is a pair $(M, \mathcal{A})$, where M is a topological manifold and $\mathcal{A}$ is a smooth structure on M. When the smooth structure is understood, we usually omit mention of it and just say "M is a smooth manifold." Smooth structures are also called *differentiable structures* or *C^∞ structures* by some authors. We will use the term *smooth manifold structure* to mean a manifold topology together with a smooth structure.

We emphasize that a smooth structure is an additional piece of data that must be added to a topological manifold before we are entitled to talk about a "smooth manifold." In fact, a given topological manifold may have many different smooth structures (we will return to this issue in the next chapter). And it should be noted that it is not always possible to find any smooth structure—there exist topological manifolds that admit no smooth structures at all.

It is worth mentioning that the notion of smooth structure can be generalized in several different ways by changing the compatibility requirement for charts. For example, if we replace the requirement that charts be smoothly compatible by the weaker requirement that each transition map $\psi \circ \varphi^{-1}$ (and its inverse) be of class C^k, we obtain the definition of a *C^k structure*. Similarly, if we require that each transition map be real-analytic (i.e., expressible as a convergent power series in a neighborhood of each point), we obtain the definition of a *real-analytic structure*, also called a *C^ω structure*. If M has even dimension $n = 2m$, we can identify $\mathbb{R}^{2m}$ with $\mathbb{C}^m$ and require that the transition maps be complex analytic; this determines a *complex analytic structure*. A manifold endowed with one of these structures is called a *C^k manifold*, *real-analytic manifold*, or *complex manifold*, respectively. (Note that a C^0 manifold is just a topological manifold.) We will not treat any of these other kinds of manifolds in this book, but they play important roles in analysis, so it is useful to know the definitions.

Without further qualification, every manifold mentioned in this book will be assumed to be a smooth manifold endowed with a specific smooth structure. In particular examples, the smooth structure will usually be obvious from the context. If M is a smooth manifold, any chart contained in the given maximal smooth atlas will be called a *smooth chart*, and the corresponding coordinate map will be called a *smooth coordinate map*.

It is generally not very convenient to define a smooth structure by explicitly describing a maximal smooth atlas, because such an atlas contains very many charts. Fortunately, we need only specify *some* smooth atlas, as the next lemma shows.

Lemma 1.4. *Let M be a topological manifold.*

(a) *Every smooth atlas for M is contained in a unique maximal smooth atlas.*

(b) *Two smooth atlases for M determine the same maximal smooth atlas if and only if their union is a smooth atlas.*

Proof. Let $\mathcal{A}$ be a smooth atlas for M, and let $\overline{\mathcal{A}}$ denote the set of all charts that are smoothly compatible with every chart in $\mathcal{A}$. To show that $\overline{\mathcal{A}}$ is a smooth atlas, we need to show that any two charts of $\overline{\mathcal{A}}$ are compatible with each other, which is to say that for any (U,φ), $(V,\psi) \in \overline{\mathcal{A}}$, $\psi\circ\varphi^{-1}\colon \varphi(U\cap V)\to\psi(U\cap V)$ is smooth.

Let $x = \varphi(p) \in \varphi(U\cap V)$ be arbitrary. Because the domains of the charts in $\mathcal{A}$ cover M, there is some chart $(W,\theta)\in\mathcal{A}$ such that $p \in W$. Since every chart in $\overline{\mathcal{A}}$ is smoothly compatible with (W,θ), both the maps $\theta\circ\varphi^{-1}$ and $\psi\circ\theta^{-1}$ are smooth where they are defined. Since $p\in U\cap V\cap W$, it follows that $\psi\circ\varphi^{-1} = (\psi\circ\theta^{-1})\circ(\theta\circ\varphi^{-1})$ is smooth on a neighborhood of x. Thus $\psi\circ\varphi^{-1}$ is smooth in a neighborhood of each point in $\varphi(U\cap V)$. Therefore $\overline{\mathcal{A}}$ is a smooth atlas. To check that it is maximal, just note that any chart that is smoothly compatible with every chart in $\overline{\mathcal{A}}$ must in particular be smoothly compatible with every chart in $\mathcal{A}$, so it is already in $\overline{\mathcal{A}}$. This proves the existence of a maximal smooth atlas containing $\mathcal{A}$. If $\mathcal{B}$ is any other maximal smooth atlas containing $\mathcal{A}$, each of its charts is smoothly compatible with each chart in $\mathcal{A}$, so $\mathcal{B}\subset\overline{\mathcal{A}}$. By maximality of $\mathcal{B}$, $\mathcal{B}=\overline{\mathcal{A}}$.

The proof of (b) is left as an exercise. □

Exercise 1.5. Prove Lemma 1.4(b).

For example, if a topological manifold M can be covered by a single chart, the smooth compatibility condition is trivially satisfied, so any such chart automatically determines a smooth structure on M.

Examples

Before proceeding further with the general theory, let us establish some examples of smooth manifolds.

Example 1.5 (Euclidean spaces). $\mathbb{R}^n$ is a smooth n-manifold with the smooth structure determined by the atlas consisting of the single chart $(\mathbb{R}^n, \mathrm{Id})$. We call this the *standard smooth structure*, and the resulting coordinate map *standard coordinates*. Unless we explicitly specify otherwise, we will always use this smooth structure on $\mathbb{R}^n$.

Example 1.6 (Finite-dimensional vector spaces). Let V be any finite-dimensional vector space. Any norm on V determines a topology, which is independent of the choice of norm (Exercise A.21 in the Appendix). With this topology, V has a natural smooth structure defined as follows. Any (ordered) basis $(E_1, \dots, E_n)$ for V defines a linear isomorphism $E\colon \mathbb{R}^n \to V$ by

$$E(x) = \sum_{i=1}^{n} x^i E_i.$$

This map is a homeomorphism, so the atlas consisting of the single chart (V, E^{-1}) defines a smooth structure. To see that this smooth structure is independent of the choice of basis, let $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ be any other basis and let $\widetilde{E}(x) = \sum_j x^j \widetilde{E}_j$ be the corresponding isomorphism. There is some invertible matrix (A_i^j) such that $E_i = \sum_j A_i^j \widetilde{E}_j$ for each j. The transition map between the two charts is then given by $\widetilde{E}^{-1} \circ E(x) = \widetilde{x}$, where $\widetilde{x} = (\widetilde{x}^1, \dots, \widetilde{x}^n)$ is determined by

$$\sum_{j=1}^{n} \widetilde{x}^j \widetilde{E}_j = \sum_{i=1}^{n} x^i E_i = \sum_{i,j=1}^{n} x^i A_i^j \widetilde{E}_j.$$

It follows that $\widetilde{x}^j = \sum_i A_i^j x^i$. Thus the map from x to $\widetilde{x}$ is an invertible linear map and hence a diffeomorphism, so the two charts are smoothly compatible. This shows that the union of the two charts determined by any two bases is still a smooth atlas, and thus all bases determine the same smooth structure. We will call this the *standard smooth structure* on V.

The Einstein Summation Convention

This is a good place to pause and introduce an important notational convention that we will use throughout the book. Because of the proliferation of summations such as $\sum_i x^i E_i$ in this subject, we will often abbreviate such a sum by omitting the summation sign, as in

$$E(x) = x^i E_i.$$

We interpret any such expression according to the following rule, called the *Einstein summation convention*: If the same index name (such as i in the expression above) appears twice in any term, once as an upper index and once as a lower index, that term is understood to be summed over all possible values of that index, generally from 1 to the dimension of the space in question. This simple idea was introduced by Einstein to reduce the complexity of the expressions arising in the study of smooth manifolds by eliminating the necessity of explicitly writing summation signs.

Another important aspect of the summation convention is the positions of the indices. We will always write basis vectors (such as E_i) with lower indices, and components of a vector with respect to a basis (such as x^i) with upper indices. These index conventions help to ensure that, in summations that make mathematical sense, any index to be summed over will typically appear twice in any given term, once as a lower index and once as an upper index.

To be consistent with our convention of writing components of vectors with upper indices, we need to use upper indices for the coordinates of a point $(x^1, \ldots, x^n) \in \mathbb{R}^n$, and we will do so throughout this book. Although this may seem awkward at first, in combination with the summation convention it offers enormous advantages when working with complicated indexed sums, not the least of which is that expressions that are not mathematically meaningful often identify themselves quickly by violating the index convention. (The main exceptions are the Euclidean dot product $x \cdot y = \sum_i x^i y^i$, in which i appears twice as an upper index, and certain expressions involving matrices. We will always explicitly write summation signs in such expressions.)

More Examples

Now we continue with our examples of smooth manifolds.

Example 1.7 (Matrices). Let $\mathrm{M}(m \times n, \mathbb{R})$ denote the space of $m \times n$ matrices with real entries. It is a vector space of dimension mn under matrix addition and scalar multiplication. Thus $\mathrm{M}(m \times n, \mathbb{R})$ is a smooth mn-dimensional manifold. Similarly, the space $\mathrm{M}(m \times n, \mathbb{C})$ of $m \times n$ complex matrices is a vector space of dimension $2mn$ over $\mathbb{R}$, and thus a manifold of dimension $2mn$. In the special case $m = n$ (square matrices), we will abbreviate $\mathrm{M}(n \times n, \mathbb{R})$ and $\mathrm{M}(n \times n, \mathbb{C})$ by $\mathrm{M}(n, \mathbb{R})$ and $\mathrm{M}(n, \mathbb{C})$, respectively.

Example 1.8 (Open Submanifolds). Let U be any open subset of $\mathbb{R}^n$. Then U is a topological n-manifold, and the single chart (U, Id) defines a smooth structure on U.

More generally, let M be a smooth n-manifold and $U \subset M$ any open subset. Define an atlas on U by

$$\mathcal{A}_U = \{\text{smooth charts } (V, \varphi) \text{ for } M \text{ such that } V \subset U\}.$$

It is easy to verify that this is a smooth atlas for U. Thus any open subset of a smooth n-manifold is itself a smooth n-manifold in a natural way. We call such a subset an *open submanifold* of M.

Example 1.9 (The General Linear Group). The *general linear group* $\mathrm{GL}(n, \mathbb{R})$ is the set of invertible $n \times n$ matrices with real entries. It is an n^2-dimensional manifold because it is an open subset of the n^2-dimensional vector space $\mathrm{M}(n, \mathbb{R})$, namely the set where the (continuous) determinant function is nonzero.

Example 1.10 (Matrices of Maximal Rank). The previous example has a natural generalization to rectangular matrices of maximal rank. Suppose $m < n$, and let $\mathrm{M}_m(m \times n, \mathbb{R})$ denote the subset of $\mathrm{M}(m \times n, \mathbb{R})$ consisting of matrices of rank m. If A is an arbitrary such matrix, the fact that $\operatorname{rank} A = m$ means that A has some nonsingular $m \times m$ minor. By continuity of the determinant function, this same minor has nonzero determinant on some neighborhood of A in $\mathrm{M}(m \times n, \mathbb{R})$, which implies that A has a neighborhood contained in $\mathrm{M}_m(m \times n, \mathbb{R})$. Thus $\mathrm{M}_m(m \times n, \mathbb{R})$ is an open subset of $M(m \times n, \mathbb{R})$, and therefore is itself an mn-dimensional manifold. A similar argument shows that $\mathrm{M}_n(m \times n, \mathbb{R})$ is an mn-manifold when $n < m$.

Exercise 1.6. If k is an integer between 0 and $\min(m, n)$, show that the set of $m \times n$ matrices whose rank is at least k is an open submanifold of $\mathrm{M}(m \times n, \mathbb{R})$.

Example 1.11 (Spheres). We showed in Example 1.2 that the n-sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ is a topological n-manifold. We put a smooth structure on $\mathbb{S}^n$ as follows. For each $i = 1, \ldots, n+1$, let $(U_i^\pm, \varphi_i^\pm)$ denote the coordinate chart we constructed in Example 1.2. For any distinct indices i and j, the transition map $\varphi_j^\pm \circ (\varphi_i^\pm)^{-1}$ is easily computed. In the case $i < j$, we get

$$\varphi_j^\pm \circ (\varphi_i^\pm)^{-1}(u^1, \ldots, u^n) = \left(u^1, \ldots, \widehat{u^i}, \ldots, \pm\sqrt{1 - |u|^2}, \ldots, u^n\right),$$

and a similar formula holds when $i > j$. When $i = j$, an even simpler computation gives $\varphi_i^\pm \circ (\varphi_i^\pm) = \mathrm{Id}_{\mathbb{B}^n}$. Thus the collection of charts $\{(U_i^\pm, \varphi_i^\pm)\}$ is a smooth atlas, and so defines a smooth structure on $\mathbb{S}^n$. We call this its *standard smooth structure*. The coordinates defined above will be called *graph coordinates*, because they arise from considering the sphere locally as the graph of the function $u^i = \pm\sqrt{1 - |u|^2}$.

Exercise 1.7. By identifying $\mathbb{R}^2$ with $\mathbb{C}$ in the usual way, we can think of the unit circle $\mathbb{S}^1$ as a subset of the complex plane. An *angle function* on a subset $U \subset \mathbb{S}^1$ is a continuous function $\theta\colon U \to \mathbb{R}$ such that $e^{i\theta(p)} = p$ for all $p \in U$. Show that there exists an angle function θ on an open subset $U \subset \mathbb{S}^1$ if and only if $U \neq \mathbb{S}^1$. For any such angle function, show that (U, θ) is a smooth coordinate chart for $\mathbb{S}^1$ with its standard smooth structure.

Example 1.12 (Projective spaces). The n-dimensional real projective space $\mathbb{P}^n$ is a topological n-manifold by Example 1.3. We will show that the coordinate charts (U_i, φ_i) constructed in that example are all smoothly compatible. Assuming for convenience that $i > j$, it is straightforward to compute that

$$\varphi_j \circ \varphi_i^{-1}(u^1, \dots, u^n) = \left(\frac{u^1}{u^j}, \dots, \frac{u^{j-1}}{u^j}, \frac{u^{j+1}}{u^j}, \dots, \frac{u^{i-1}}{u^j}, \frac{1}{u^j}, \frac{u^{i+1}}{u^j}, \dots, \frac{u^n}{u^j}\right),$$

which is a diffeomorphism from $\varphi_i(U_i \cap U_j)$ to $\varphi_j(U_i \cap U_j)$.

Example 1.13 (Product Manifolds). Suppose $M_1, \dots, M_k$ are smooth manifolds of dimensions $n_1, \dots, n_k$ respectively. The product space $M_1 \times \cdots \times M_k$ is Hausdorff by Exercise 1.2 and second countable by Exercise 1.3. Given a smooth chart (U_i, φ_i) for each M_i, the map $\varphi_1 \times \cdots \times \varphi_k\colon U_1 \times \cdots \times U_k \to \mathbb{R}^{n_1+\cdots+n_k}$ is a homeomorphism onto its image, which is an open subset of $\mathbb{R}^{n_1+\cdots+n_k}$. Thus the product set is a topological manifold of dimension $n_1 + \cdots + n_k$, with charts of the form $(U_1 \times \cdots \times U_k, \varphi_1 \times \cdots \times \varphi_k)$. Any two such charts are smoothly compatible because, as is easily verified,

$$(\psi_1 \times \cdots \times \psi_k) \circ (\varphi_1 \times \cdots \times \varphi_k)^{-1} = (\psi_1 \circ \varphi_1^{-1}) \times \cdots \times (\psi_k \circ \varphi_k^{-1}),$$

which is a smooth map. This defines a natural smooth manifold structure on the product, called the *product smooth manifold structure*. For example, this yields a smooth manifold structure on the n-dimensional torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$.

In each of the examples we have seen so far, we have constructed a smooth manifold structure in two stages: We started with a topological space and checked that it was a topological manifold, and then we specified a smooth structure. It is often more convenient to combine these two steps into a single construction, especially if we start with a set or a topological space that is not known a priori to be a topological manifold. The following lemma provides a shortcut.

Lemma 1.14 (One-Step Smooth Manifold Structure). *Let M be a set, and suppose we are given a collection $\{U_\alpha\}$ of subsets of M, together with an injective map $\varphi_\alpha\colon U_\alpha \to \mathbb{R}^n$ for each α, such that the following properties are satisfied.*

(*i*) *For each* α, $\widetilde{U}_\alpha = \varphi_\alpha(U_\alpha)$ *is an open subset of* $\mathbb{R}^n$.

(*ii*) *For each* α *and* β, $\varphi_\alpha(U_\alpha \cap U_\beta)$ *and* $\varphi_\beta(U_\alpha \cap U_\beta)$ *are open in* $\mathbb{R}^n$.

(*iii*) *Whenever* $U_\alpha \cap U_\beta \neq \varnothing$, $\varphi_\beta \circ \varphi_\alpha^{-1}\colon \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$ *is smooth.*

(*iv*) *Countably many of the sets* U_α *cover* M.

(*v*) *Whenever* p, q *are distinct points in* M, *either there exists some* U_α *containing both* p *and* q *or there exist disjoint sets* U_α, U_β *with* $p \in U_\alpha$ *and* $q \in U_\beta$.

Then M *has a unique smooth manifold structure such that each* $(U_\alpha, \varphi_\alpha)$ *is a smooth chart.*

Proof. We define the topology by taking the sets of the form $\varphi_\alpha^{-1}(V)$, where $V \subset \widetilde{U}_\alpha$ is open, as a basis. To prove that this is a basis for a topology, let $\varphi_\alpha^{-1}(V)$ and $\varphi_\beta^{-1}(W)$ be two such basis sets. Properties (ii) and (iii) imply that $\varphi_\alpha \circ \varphi_\beta^{-1}(W)$ is an open subset of $\varphi_\alpha(U_\alpha \cap U_\beta)$, and therefore also of $\widetilde{U}_\alpha$. Thus if p is any point in $\varphi_\alpha^{-1}(V) \cap \varphi_\beta^{-1}(W)$, then

$$\varphi_\alpha^{-1}(V \cap \varphi_\alpha \circ \varphi_\beta^{-1}(W)) = \varphi_\alpha^{-1}(V) \cap \varphi_\beta^{-1}(W)$$

is a basis open set containing p. Each of the maps φ_α is then a homeomorphism (essentially by definition), so M is locally Euclidean of dimension n. If $\{U_{\alpha_i}\}$ is a countable collection of the sets U_α covering M, each of the sets U_{α_i} has a countable basis, and the union of all these is a countable basis for M, so M is second countable, and the Hausdorff property follows easily from (v). Finally, (iii) guarantees that the collection $\{(U_\alpha, \varphi_\alpha)\}$ is a smooth atlas. It is clear that this topology and smooth structure are the unique ones satisfying the conclusions of the lemma. $\square$

Example 1.15 (Grassmann Manifolds). Let V be an n-dimensional real vector space. For any integer $0 \leq k \leq n$, we let $\mathrm{G}_k(V)$ denote the set of all k-dimensional linear subspaces of V. We will show that $\mathrm{G}_k(V)$ can be naturally given the structure of a smooth manifold of dimension $k(n-k)$. The construction is somewhat more involved than the ones we have done so far, but the basic idea is just to use linear algebra to construct charts for $\mathrm{G}_k(V)$ and then use Lemma 1.14 to show that these charts yield a smooth manifold structure. Since we will give a more straightforward proof that $\mathrm{G}_k(V)$ is a smooth manifold after we have developed more machinery in Chapter 7, you may skip the details of this construction on first reading if you wish.

Let P and Q be any complementary subspaces of V of dimensions k and $(n-k)$, respectively, so that V decomposes as a direct sum: $V = P \oplus Q$. The graph of any linear map $A\colon P \to Q$ is a k-dimensional subspace $\Gamma(A) \subset V$, defined by

$$\Gamma(A) = \{x + Ax : x \in P\}.$$

Any such subspace has the property that its intersection with Q is the zero subspace. Conversely, any subspace with this property is easily seen to be the graph of a unique linear map $A\colon P \to Q$.

Let $L(P, Q)$ denote the vector space of linear maps from P to Q, and let U_Q denote the subset of $\mathrm{G}_k(V)$ consisting of k-dimensional subspaces whose intersection with Q is trivial. Define a map $\psi\colon L(P,Q) \to U_Q$ by

$$\psi(A) = \Gamma(A).$$

The discussion above shows that ψ is a bijection. Let $\varphi = \psi^{-1}\colon U_Q \to L(P,Q)$. By choosing bases for P and Q, we can identify $L(P,Q)$ with $\mathrm{M}((n-k)\times k, \mathbb{R})$ and hence with $\mathbb{R}^{k(n-k)}$, and thus we can think of (U_Q, φ) as a coordinate chart. Since the image of each chart is all of $L(P,Q)$, condition (i) of Lemma 1.14 is clearly satisfied.

Now let (P', Q') be any other such pair of subspaces, and let ψ', φ' be the corresponding maps. The set $\varphi(U_Q \cap U_{Q'}) \subset L(P,Q)$ consists of all $A \in L(P,Q)$ whose graphs intersect both Q and Q' trivially, which is easily seen to be an open set, so (ii) holds. We need to show that the transition map $\varphi' \circ \varphi^{-1} = \varphi' \circ \psi$ is smooth on this set. This is the trickiest part of the argument.

Suppose $A \in \varphi(U_Q \cap U_{Q'}) \subset L(P,Q)$ is arbitrary, and let S denote the subspace $\psi(A) = \Gamma(A) \subset V$. If we put $A' = \varphi' \circ \psi(A)$, then A' is the unique linear map from P' to Q' whose graph is equal to S. To identify this map, let $x' \in P'$ be arbitrary, and note that $A'x'$ is the unique element of Q' such that $x' + A'x' \in S$, which is to say that

$$x' + A'x' = x + Ax \quad \text{for some } x \in P. \tag{1.1}$$

(See Figure 1.3.) There is in fact a unique $x \in P$ for which this holds, characterized by the property that

$$x + Ax - x' \in Q'.$$

If we let $I_A\colon P \to V$ denote the map $I_A(x) = x + Ax$ and let $\pi_{P'}\colon V \to P'$ be the projection onto P' with kernel Q', then x satisfies

$$0 = \pi_{P'}(x + Ax - x') = \pi_{P'} \circ I_A(x) - x'.$$

As long as A stays in the open subset of maps whose graphs intersect both Q and Q' trivially, $\pi_{P'} \circ I_A\colon P \to P'$ is invertible, and thus we can solve this last equation for x to obtain $x = (\pi_{P'} \circ I_A)^{-1}(x')$. Therefore, A' is given in terms of A by

$$A'x' = I_A x - x' = I_A \circ (\pi_{P'} \circ I_A)^{-1}(x') - x'. \tag{1.2}$$

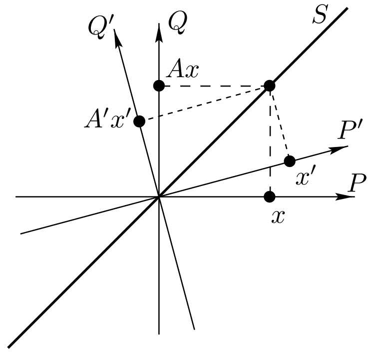

FIGURE 1.3. Smooth compatibility of coordinates on $\mathrm{G}_k(V)$.

If we choose bases (E_i') for P' and (F_j') for Q', the columns of the matrix representation of A' are the components of $A'E_i'$. By (1.2), this can be written

$$A'E_i' = I_A \circ (\pi_{P'} \circ I_A)^{-1}(E_i') - E_i'.$$

The matrix entries of I_A clearly depend smoothly on those of A, and thus so also do those of $\pi_{P'} \circ I_A$. By Cramer's rule, the components of the inverse of a matrix are rational functions of the matrix entries, so the expression above shows that the components of $A'E_i'$ depend smoothly on the components of A. This proves that $\varphi' \circ \varphi^{-1}$ is a smooth map, so the charts we have constructed satisfy condition (iii) of Lemma 1.14.

To check the countability condition (iv), we just note that $\mathrm{G}_k(V)$ can in fact be covered by *finitely* many of the sets U_Q: For example, if $(E_1, \dots, E_n)$ is any fixed basis for V, any partition of the basis elements into two subsets containing k and $n-k$ elements determines appropriate subspaces P and Q, and any subspace S must have trivial intersection with Q for at least one of these partitions (see Exercise A.4). Thus $\mathrm{G}_k(V)$ is covered by the finitely many charts determined by all possible partitions of a fixed basis. Finally, the Hausdorff condition (v) is easily verified by noting that for any two k-dimensional subspaces $P, P' \subset V$, it is possible to find a subspace Q of dimension $n-k$ whose intersections with both P and P' are trivial, and then P and P' are both contained in the domain of the chart determined by, say, (P, Q).

The smooth manifold $\mathrm{G}_k(V)$ is called the *Grassmann manifold* of k-planes in V, or simply a *Grassmannian*. In the special case $V = \mathbb{R}^n$, the Grassmannian $\mathrm{G}_k(\mathbb{R}^n)$ is often denoted by some simpler notation such as $\mathrm{G}_{k,n}$ or $\mathrm{G}(k,n)$. Note that $\mathrm{G}_1(\mathbb{R}^{n+1})$ is exactly the n-dimensional projective space $\mathbb{P}^n$.

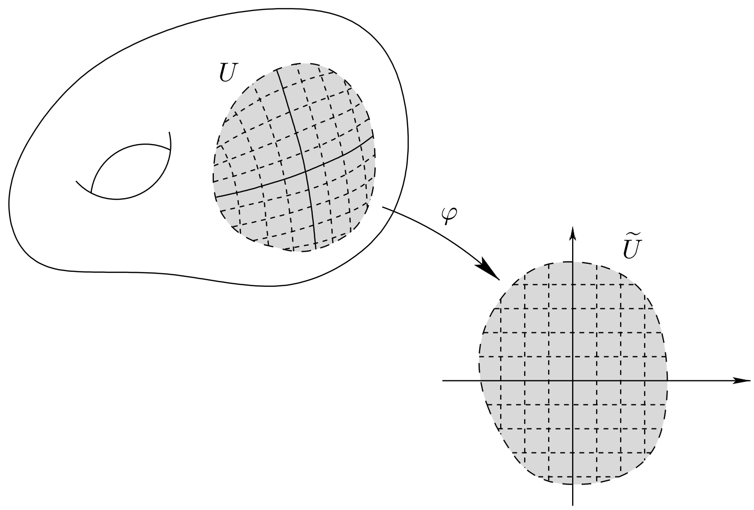

FIGURE 1.4. A coordinate grid.

Exercise 1.8. Let $0 < k < n$ be integers, and let $P, Q \subset \mathbb{R}^n$ be the subspaces spanned by $(e_1, \dots, e_k)$ and $(e_{k+1}, \dots, e_n)$, respectively, where e_i is the ith standard basis vector. For any k-dimensional subspace $S \subset \mathbb{R}^n$ that has trivial intersection with Q, show that the coordinate representation $\varphi(S)$ constructed in the preceding example is the unique $(n-k) \times k$ matrix B such that S is spanned by the columns of the matrix $\binom{I_k}{B}$, where I_k denotes the $k \times k$ identity matrix.

Local Coordinate Representations

Here is how one usually thinks about local coordinate charts on a smooth manifold. Once we choose a chart (U, φ) on M, the coordinate map $\varphi \colon U \to \widetilde{U} \subset \mathbb{R}^n$ can be thought of as giving an *identification* between U and $\widetilde{U}$. Using this identification, we can think of U simultaneously as an open subset of M and (at least temporarily while we work with this chart) as an open subset of $\mathbb{R}^n$. You can visualize this identification by thinking of a "grid" drawn on U representing the inverse images of the coordinate lines under φ (Figure 1.4). Under this identification, we can represent a point $p \in M$ by its coordinates $(x^1, \dots, x^n) = \varphi(p)$, and think of this n-tuple as *being* the point p. We will typically express this by saying "$(x^1, \dots, x^n)$ is the (local) coordinate representation for p" or "$p = (x^1, \dots, x^n)$ in local coordinates."

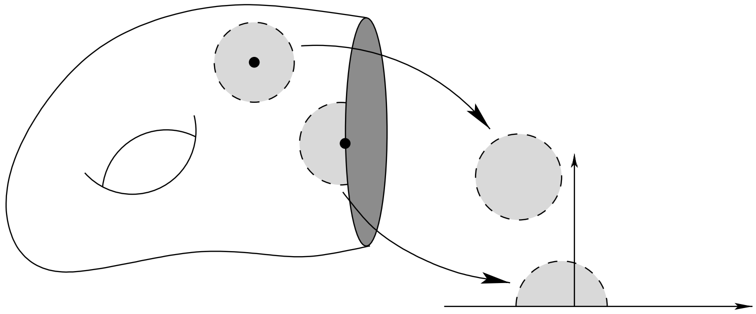

FIGURE 1.5. A manifold with boundary.

Another way to look at it is that by means of our identification $U \leftrightarrow \widetilde{U}$, we can think of φ as the identity map and suppress it from the notation. This takes a bit of getting used to, but the payoff is a huge simplification of the notation in many situations. You just need to remember that the identification depends heavily on the choice of coordinate chart.

For example, if $M = \mathbb{R}^2$, let $U = \{(x,y) : x > 0\}$ be the open right half-plane, and let $\varphi\colon U \to \mathbb{R}^2$ be the *polar coordinate map* $\varphi(x,y) = (r,\theta) = (\sqrt{x^2+y^2}, \arctan y/x)$. We can write a given point $p \in U$ either as $p = (x,y)$ in standard coordinates or as $p = (r,\theta)$ in polar coordinates, where the two coordinate representations are related by $(r,\theta) = (\sqrt{x^2+y^2}, \arctan y/x)$ and $(x,y) = (r\cos\theta, r\sin\theta)$.

Manifolds With Boundary

For some purposes, we will need the following generalization of manifolds. An n-dimensional *topological manifold with boundary* is a second countable Hausdorff space in which every point has a neighborhood homeomorphic to an open subset of the closed n-dimensional upper half space $\mathbb{H}^n = \{(x^1,\dots,x^n) \in \mathbb{R}^n : x^n \geq 0\}$ (Figure 1.5). An open subset $U \subset M$ together with a homeomorphism φ from U to an open subset of $\mathbb{H}^n$ is called a *generalized chart* for M.

The boundary of $\mathbb{H}^n$ in $\mathbb{R}^n$ is the set of points where $x^n = 0$. If M is a manifold with boundary, a point that is in the inverse image of $\partial\mathbb{H}^n$ under some generalized chart is called a *boundary point* of M, and a point that is in the inverse image of $\operatorname{Int}\mathbb{H}^n$ is called an *interior point*. The *boundary* of M (the set of all its boundary points) is denoted ∂M; similarly its *interior* is denoted $\operatorname{Int} M$.

Be careful to observe the distinction between this use of the terms "boundary" and "interior" and their usage to refer to the boundary and interior of a subset of a topological space. A manifold M with boundary may have nonempty boundary in this new sense, irrespective of whether it has a boundary as a subset of some other topological space. If we need to emphasize the difference between the two notions of boundary, we will use the terms *topological boundary* or *manifold boundary* as appropriate.

To see how to define a smooth structure on a manifold with boundary, recall that a smooth map from an arbitrary subset $A \subset \mathbb{R}^n$ is defined to be one that extends smoothly to an open neighborhood of A (see the Appendix). Thus if U is an open subset of $\mathbb{H}^n$, a smooth map $F\colon U \to \mathbb{R}^k$ is a map that extends to a smooth map $\widetilde{F}\colon \widetilde{U} \to \mathbb{R}^k$, where $\widetilde{U}$ is some open subset of $\mathbb{R}^n$ containing U. If F is such a map, by continuity all the partial derivatives of F at points of $\partial\mathbb{H}^n$ are determined by their values in $\mathbb{H}^n$, and therefore in particular are independent of the choice of extension. It is a fact (which we will neither prove nor use) that $F\colon U \to \mathbb{R}^k$ has such a smooth extension if and only if F is continuous, $F|_{U\cap \operatorname{Int}\mathbb{H}^n}$ is smooth, and each of the partial derivatives of $F|_{U\cap \operatorname{Int}\mathbb{H}^n}$ has a continuous extension to $U \cap \mathbb{H}^n$.

For example, let $\mathbb{B}^2 \subset \mathbb{R}^2$ denote the unit disk, let $U = \mathbb{B}^2 \cap \mathbb{H}^2$, and define $f\colon U \to \mathbb{R}$ by $f(x, y) = \sqrt{1 - x^2 - y^2}$. Because f extends to all of $\mathbb{B}^2$ (by the same formula), f is a smooth function on U. On the other hand, although $g(x, y) = \sqrt{y}$ is continuous on U and smooth in $U \cap \operatorname{Int}\mathbb{H}^2$, it has no smooth extension to any neighborhood of U in $\mathbb{R}^2$ because $\partial g/\partial y \to \infty$ as $y \to 0$. Thus g is not a smooth function on U.

Given a topological manifold with boundary M, we define an atlas for M as before to be a collection of generalized charts whose domains cover M. Two such charts (U, φ), (V, ψ) are smoothly compatible if $\psi \circ \varphi^{-1}$ is smooth (in the sense just described) wherever it is defined. Just as in the case of manifolds, a smooth atlas for M is an atlas all of whose charts are smoothly compatible with each other, and a smooth structure for M is a maximal smooth atlas.

It can be shown using homology theory that the interior and boundary of a topological manifold with boundary are disjoint (see [Lee00, Problem 13-9], for example). We will not need this result, because the analogous result for smooth manifolds with boundary is much easier to prove (or will be, after we have developed a bit more machinery). A proof is outlined in Problem 5-19.

Since any open ball in $\mathbb{R}^n$ admits a diffeomorphism onto an open subset of $\mathbb{H}^n$, a smooth n-manifold is automatically a smooth n-manifold with boundary (whose boundary is empty), but the converse is not true: A manifold with boundary is a manifold if and only if its boundary is empty. (This will follow from the fact that interior points and boundary points are distinct.)

Problems

1-1. Let X be the set of all points $(x, y) \in \mathbb{R}^2$ such that $y = \pm 1$, and let M be the quotient of X by the equivalence relation generated by $(x, -1) \sim (x, 1)$ for all $x \neq 0$. Show that M is locally Euclidean and second countable, but not Hausdorff. [This space is called the *line with two origins.*]

1-2. Show that the disjoint union of uncountably many copies of $\mathbb{R}$ is locally Euclidean and Hausdorff, but not second countable.

1-3. Let $N = (0, \dots, 0, 1)$ be the "north pole" and $S = -N$ the "south pole." Define *stereographic projection* $\sigma \colon \mathbb{S}^n \smallsetminus \{N\} \to \mathbb{R}^n$ by

$$\sigma(x^1, \dots, x^{n+1}) = \frac{(x^1, \dots, x^n)}{1 - x^{n+1}}.$$

Let $\widetilde{\sigma}(x) = \sigma(-x)$ for $x \in \mathbb{S}^n \smallsetminus \{S\}$.

(a) Show that σ is bijective, and

$$\sigma^{-1}(u^1, \dots, u^n) = \frac{(2u^1, \dots, 2u^n, |u|^2 - 1)}{|u|^2 + 1}.$$

(b) Compute the transition map $\widetilde{\sigma} \circ \sigma^{-1}$ and verify that the atlas consisting of the two charts $(\mathbb{S}^n \smallsetminus \{N\}, \sigma)$ and $(\mathbb{S}^n \smallsetminus \{S\}, \widetilde{\sigma})$ defines a smooth structure on $\mathbb{S}^n$. (The coordinates defined by σ or $\widetilde{\sigma}$ are called *stereographic coordinates.*)

(c) Show that this smooth structure is the same as the one defined in Example 1.11.

1-4. Let M be a smooth n-manifold with boundary. Show that $\operatorname{Int} M$ is a smooth n-manifold and ∂M is a smooth $(n-1)$-manifold (both without boundary).

1-5. Let $M = \overline{\mathbb{B}^n}$, the closed unit ball in $\mathbb{R}^n$. Show that M is a manifold with boundary and has a natural smooth structure such that its interior is the open unit ball with its standard smooth structure.

2

Smooth Maps

The main reason for introducing smooth structures was to enable us to define smooth functions on manifolds and smooth maps between manifolds. In this chapter, we will carry out that project.

Although the terms "function" and "map" are technically synonymous, when studying smooth manifolds it is often convenient to make a slight distinction between them. Throughout this book, we will generally reserve the term "function" for a map whose range is $\mathbb{R}$ (a *real-valued function*) or $\mathbb{R}^k$ for some $k > 1$ (a *vector-valued function*). The word "map" or "mapping" can mean any type of map, such as a map between arbitrary manifolds.

We begin by defining smooth real-valued and vector-valued functions, and then generalize this to smooth maps between manifolds. We then study diffeomorphisms, which are bijective smooth maps with smooth inverses. If there is a diffeomorphism between two manifolds, we say they are diffeomorphic. The main objects of study in smooth manifold theory are properties that are invariant under diffeomorphisms.

Later in the chapter, we study smooth covering maps, and their relationship to the continuous covering maps studied in topology; and we introduce Lie groups, which are smooth manifold that are also groups in which multiplication and inversion are smooth maps.

At the end of the chapter, we introduce some powerful tools for smoothly piecing together local smooth objects, called bump functions and partitions of unity. They will be used throughout the book for building global smooth objects out of ones that are initially defined only locally.

Smooth Functions and Smooth Maps

If M is a smooth manifold, a function $f\colon M \to \mathbb{R}^k$ is said to be *smooth* if, for every smooth chart (U, φ) on M, the composite function $f \circ \varphi^{-1}$ is smooth on the open subset $\varphi(U) \subset \mathbb{R}^n$. The most important special case is that of smooth real-valued functions $f\colon M \to \mathbb{R}$; the set of all such functions is denoted by $C^\infty(M)$. Because sums and constant multiples of smooth functions are smooth, $C^\infty(M)$ is a vector space. In fact, it is a ring under pointwise multiplication, as you can easily verify.

Although by definition smoothness of f means that its composition with *every* smooth coordinate map is smooth, in practice it suffices to check smoothness in each of the charts of some smooth atlas, as the next lemma shows.

Lemma 2.1. *Suppose $\{(U_\alpha, \varphi_\alpha)\}$ is a smooth atlas for M. If $f\colon M \to \mathbb{R}^k$ is a function such that $f \circ \varphi_\alpha^{-1}$ is smooth for each α, then f is smooth.*

Proof. We just need to check that $f \circ \varphi^{-1}$ is smooth for *any* smooth chart (U, φ) on M. It suffices to show it is smooth in a neighborhood of each point $x = \varphi(p) \in \varphi(U)$. For any $p \in U$, there is a chart $(U_\alpha, \varphi_\alpha)$ in the atlas whose domain contains p. Since (U, φ) is smoothly compatible with $(U_\alpha, \varphi_\alpha)$, the transition map $\varphi_\alpha \circ \varphi^{-1}$ is smooth on its domain of definition, which includes x. Thus $f \circ \varphi^{-1} = (f \circ \varphi_\alpha^{-1}) \circ (\varphi_\alpha \circ \varphi^{-1})$ is smooth in a neighborhood of x. □

Given a function $f\colon M \to \mathbb{R}^k$ and a chart (U, φ) for M, the function $\widehat{f}\colon \varphi(U) \to \mathbb{R}^k$ defined by $\widehat{f}(x) = f \circ \varphi^{-1}(x)$ is called the *coordinate representation* of f. For example, consider $f(x, y) = x^2 + y^2$ on the plane. In polar coordinates, it has the coordinate representation $\widehat{f}(r, \theta) = r^2$. In keeping with our practice of using local coordinates to identify U with a subset of Euclidean space, in cases where it will cause no confusion we will often not even observe the distinction between $\widehat{f}$ and f itself, and write $f(r, \theta) = r^2$ in polar coordinates. Thus, we might say "f is smooth on U because its coordinate representation $f(r, \theta) = r^2$ is smooth."

The definition of smooth functions generalizes easily to maps between manifolds. Let M, N be smooth manifolds, and let $F\colon M \to N$ be any map. We say F is a *smooth map* if, for any smooth charts (U, φ) for M and (V, ψ) for N, the composite map $\psi \circ F \circ \varphi^{-1}$ is smooth from $\varphi(U \cap F^{-1}(V))$ to $\psi(V)$. Note that our previous definition of smoothness of real-valued functions can be viewed as a special case of this one, by taking $N = \mathbb{R}^k$ and $\psi = \mathrm{Id}$.

Exercise 2.1 (Smoothness is Local). Let $F\colon M \to N$ be a map between smooth manifolds, and suppose each point $p \in M$ has a neighborhood U such that $F|_U$ is smooth. Show that F is smooth.

We call $\widehat{F} = \psi \circ F \circ \varphi^{-1}$ the coordinate representation of F with respect to the given coordinates. As with real-valued or vector-valued functions, once we have chosen specific local coordinates in both the domain and range, we can often ignore the distinction between F and $\widehat{F}$.

Just as for functions, to prove that a map is smooth it suffices to show that its coordinate representatives with respect to a particular smooth atlas are smooth. The proof is analogous to that of Lemma 2.1 and is left as an exercise.

Lemma 2.2. *Let M, N be smooth manifolds and let $F\colon M \to N$ be any map. If $\{(U_\alpha, \varphi_\alpha)\}$ and $\{(V_\beta, \psi_\beta)\}$ are smooth atlases for M and N, respectively, and if for each α and β, $\psi_\beta \circ F \circ \varphi_\alpha^{-1}$ is smooth on its domain of definition, then F is smooth.*

Exercise 2.2. Prove Lemma 2.2.

Lemma 2.3. *Any composition of smooth maps between manifolds is smooth.*

Proof. Given smooth maps $F\colon M \to N$ and $G\colon N \to P$, let (U, φ) and (V, ψ) be any charts for M and P respectively. We need to show that $\psi \circ (G \circ F) \circ \varphi^{-1}$ is smooth where it is defined, namely on $\varphi(U \cap (G \circ F)^{-1}(V))$. For any point $p \in U \cap (G \circ F)^{-1}(V)$, there is a chart (W, θ) for N such that $F(p) \in W$. Smoothness of F and G means that $\theta \circ F \circ \varphi^{-1}$ and $\psi \circ G \circ \theta^{-1}$ are smooth where they are defined, and therefore $\psi \circ (G \circ F) \circ \varphi^{-1} = (\psi \circ G \circ \theta^{-1}) \circ (\theta \circ F \circ \varphi^{-1})$ is smooth. □

Example 2.4 (Smooth maps).

(a) Consider the n-sphere $\mathbb{S}^n$ with its standard smooth structure. The inclusion map $\iota\colon \mathbb{S}^n \hookrightarrow \mathbb{R}^{n+1}$ is certainly continuous, because it is the inclusion map of a topological subspace. It is a smooth map because its coordinate representation with respect to any of the graph coordinates of Example 1.11 is

$$\begin{aligned}\widehat{\iota}(u^1, \dots, u^n) &= \iota \circ (\varphi_i^{\pm})^{-1}(u^1, \dots, u^n) \\ &= \left(u^1, \dots, u^{i-1}, \pm\sqrt{1 - |u|^2}, u^i, \dots, u^n\right),\end{aligned}$$

which is smooth on its domain (the set where $|u|^2 < 1$).

(b) The quotient map $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{P}^n$ is smooth, because its coordinate representation in terms of the coordinates for $\mathbb{P}^n$ constructed in Example 1.12 and standard coordinates on $\mathbb{R}^{n+1} \smallsetminus \{0\}$ is

$$\begin{aligned}\widehat{\pi}(x^1, \dots, x^{n+1}) &= \varphi_i \circ \pi(x^1, \dots, x^{n+1}) = \varphi_i[x^1, \dots, x^{n+1}] \\ &= \left(\frac{x^1}{x^i}, \dots, \frac{x^{i-1}}{x^i}, \frac{x^{i+1}}{x^i}, \dots, \frac{x^{n+1}}{x^i}\right).\end{aligned}$$

(c) Define $p\colon \mathbb{S}^n \to \mathbb{P}^n$ as the restriction of $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{P}^n$ to $\mathbb{S}^n \subset \mathbb{R}^{n+1} \smallsetminus \{0\}$. It is a smooth map, because it is the composition $p = \pi \circ \iota$ of the maps in the preceding two examples.

Exercise 2.3. Let $M_1, \dots, M_k$ and N be smooth manifolds. Show that a map $F\colon N \to M_1 \times \cdots \times M_k$ is smooth if and only if each of the "component maps" $F_i = \pi_i \circ F\colon N \to M_i$ is smooth. (Here $\pi_i\colon M_1 \times \cdots \times M_k \to M_i$ is the projection onto the ith factor.)

The definitions of smooth functions and smooth maps on a manifold with boundary are exactly the same as for manifolds; you can work out the details for yourself.

Diffeomorphisms

A *diffeomorphism* between manifolds M and N is a smooth map $F\colon M \to N$ that has a smooth inverse. We say M and N are *diffeomorphic* if there exists a diffeomorphism between them. Sometimes this is symbolized by $M \approx N$. For example, if $\mathbb{B}^n$ denotes the open unit ball in $\mathbb{R}^n$, the map $F\colon \mathbb{B}^n \to \mathbb{R}^n$ given by $F(x) = x/(1 - |x|^2)$ is easily seen to be a diffeomorphism, so $\mathbb{B}^n \approx \mathbb{R}^n$.

Exercise 2.4. Show that "diffeomorphic" is an equivalence relation.

More generally, $F\colon M \to N$ is called a *local diffeomorphism* if every point $p \in M$ has a neighborhood U such that $F(U)$ is open in N and $F\colon U \to F(U)$ is a diffeomorphism. It is clear from the definition that a local diffeomorphism is, in particular, a local homeomorphism and therefore an open map.

Exercise 2.5. Show that a map $F\colon M \to N$ is a diffeomorphism if and only if it is a bijective local diffeomorphism.

Just as two topological spaces are considered to be "the same" if they are homeomorphic, two smooth manifolds are essentially indistinguishable if they are diffeomorphic. The central concern of smooth manifold theory is the study of properties of smooth manifolds that are preserved by diffeomorphisms.

One question that naturally arises is to what extent a smooth structure on a given topological manifold might be unique. There are really two different questions here: The first is whether a given manifold M admits distinct smooth structures, and the second is whether it admits smooth structures that are not diffeomorphic to each other.

Let us begin by addressing the first question. It is easy to see that two smooth structures $\mathcal{A}_1, \mathcal{A}_2$ on a given manifold M are the same if and only if the identity map of M is a diffeomorphism from $(M, \mathcal{A}_1)$ to $(M, \mathcal{A}_2)$.

In general, a given topological manifold will admit very many distinct smooth structures. For example, consider the two homeomorphisms $\varphi\colon \mathbb{R} \to \mathbb{R}$ and $\psi\colon \mathbb{R} \to \mathbb{R}$ given by

$$\begin{aligned}\varphi(x) &= x,\\ \psi(x) &= x^3.\end{aligned}$$

Each of the atlases $\{(\mathbb{R}, \varphi)\}$ and $\{(\mathbb{R}, \psi)\}$ determines a smooth structure on $\mathbb{R}$. (Since there is only one chart in each case, the smooth compatibility condition is trivially satisfied.) These two charts are not smoothly compatible with each other, because $\varphi \circ \psi^{-1}(y) = y^{1/3}$ is not smooth at the origin. Therefore the two smooth structures on $\mathbb{R}$ determined by these atlases are distinct. Using similar ideas, it is not hard to construct many different smooth structures on any given manifold.

The second question, whether two given smooth structures are diffeomorphic to each other, is more subtle. Consider the same two smooth structures on $\mathbb{R}$, and for the moment let $\mathbb{R}_\varphi$ denote $\mathbb{R}$ with the smooth structure determined by φ (this is just the standard smooth structure) and $\mathbb{R}_\psi$ the same topological manifold but with the smooth structure determined by ψ. It turns out that these two manifolds are diffeomorphic to each other. Define a map $F\colon \mathbb{R}_\varphi \to \mathbb{R}_\psi$ by $F(x) = x^{1/3}$. The coordinate representation of this map is $\widehat{F}(t) = \psi \circ F \circ \varphi^{-1}(t) = t$, which is clearly smooth. Moreover, the coordinate representation of its inverse is $\widehat{F^{-1}}(y) = \varphi \circ F^{-1} \circ \psi^{-1}(y) = y$, which is also smooth, so F is a diffeomorphism. (This is one case in which it is important to maintain the distinction between a map and its coordinate representation!)

It turns out, as you will see later, that there is only one smooth structure on $\mathbb{R}$ up to diffeomorphism (see Problem 12-5). More precisely, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are any two smooth structures on $\mathbb{R}$, there exists a diffeomorphism $F\colon (\mathbb{R}, \mathcal{A}_1) \to (\mathbb{R}, \mathcal{A}_2)$. In fact, it follows from work of Edwin Moise [Moi77] and James Munkres [Mun60] that every topological manifold of dimension less than or equal to 3 has a smooth structure that is unique up to diffeomorphism. The analogous question in higher dimensions turns out to be quite deep, and is still largely unanswered. Even for Euclidean spaces, the problem was not completely solved until late in the twentieth century. The answer is somewhat surprising: As long as $n \neq 4$, $\mathbb{R}^n$ has a unique smooth structure (up to diffeomorphism); but $\mathbb{R}^4$ has uncountably many distinct smooth structures, no two of which are diffeomorphic to each other! The existence of nonstandard smooth structures on $\mathbb{R}^4$ (called *fake* $\mathbb{R}^4 s$) was first proved by Simon Donaldson and Michael Freedman in 1984 as a consequence of their work on the geometry and topology of compact 4-manifolds; the results are described in [DK90] and [FQ90].

For compact manifolds, the situation is even more interesting. For example, in 1963, Michel Kervaire and John Milnor [KM63] showed that, up to diffeomorphism, $\mathbb{S}^7$ has exactly 28 non-diffeomorphic smooth structures.

On the other hand, in all dimensions greater than 3 there are compact topological manifolds that have no smooth structures at all. (The first example was found in 1960 by Kervaire [Ker60].) The problem of identifying the number of smooth structures (if any) on topological 4-manifolds is an active subject of current research.

Smooth Covering Maps

You are probably already familiar with the notion of a *covering map* between topological spaces: This is a surjective continuous map $\pi\colon \widetilde{M} \to M$ between connected, locally path connected spaces, with the property that every point $p \in M$ has a neighborhood U that is *evenly covered*, meaning that each component of $\pi^{-1}(U)$ is mapped homeomorphically onto U by π. In this section, we will assume familiarity with the basic properties of covering maps, as described for example in [Lee00, Chapters 11 and 12].

In the context of smooth manifolds, it is useful to introduce a slightly more restrictive type of covering map. If $\widetilde{M}$ and M are connected smooth manifolds, a *smooth covering map* $\pi\colon \widetilde{M} \to M$ is a smooth surjective map with the property that every $p \in M$ has a neighborhood U such that each component of $\pi^{-1}(U)$ is mapped *diffeomorphically* onto U by π. In this context, we will also say that U is evenly covered. The manifold M is called the *base* of the covering, and $\widetilde{M}$ is called a *covering space* of M.

To distinguish this new definition from the previous one, we will often call an ordinary (not necessarily smooth) covering map a *topological covering map*. A smooth covering map is, in particular, a topological covering map. However, it is important to bear in mind that a smooth covering map is more than just a topological covering map that happens to be smooth—the definition of smooth covering map requires in addition that the restriction of π to each component of the inverse image of an evenly covered set be a diffeomorphism, not just a smooth homeomorphism.

Proposition 2.5 (Properties of Smooth Coverings).

(*a*) *Any smooth covering map is a local diffeomorphism and an open map.*

(*b*) *An injective smooth covering map is a diffeomorphism.*

(*c*) *A topological covering map is a smooth covering map if and only if it is a local diffeomorphism.*

Exercise 2.6. Prove Proposition 2.5.

If $\pi\colon \widetilde{M} \to M$ is any continuous map, a *section* of π is a continuous map $\sigma\colon M \to \widetilde{M}$ such that $\pi \circ \sigma = \mathrm{Id}_M$. A *local section* is a continuous map $\sigma\colon U \to \widetilde{M}$ defined on some open set $U \subset M$ and satisfying the analogous relation $\pi \circ \sigma = \mathrm{Id}_U$. Many of the important properties of smooth covering maps arise from the existence of smooth local sections.

Lemma 2.6 (Local Sections of Smooth Coverings). *Suppose $\pi\colon \widetilde{M} \to M$ is a smooth covering map. Every point of $\widetilde{M}$ is in the image of a smooth local section of π. More precisely, for any $q \in \widetilde{M}$, there is a neighborhood U of $p = \pi(q)$ and a smooth local section $\sigma\colon U \to \widetilde{M}$ such that $\sigma(p) = q$.*

Proof. Let $U \subset M$ be an evenly covered neighborhood of p. If $\widetilde{U}$ is the component of $\pi^{-1}(U)$ containing q, then $\pi|_{\widetilde{U}} : \widetilde{U} \to U$ is by hypothesis a diffeomorphism. It follows that $\sigma = (\pi|_{\widetilde{U}})^{-1} \colon U \to \widetilde{U}$ is a smooth local section of π such that $\sigma(p) = q$. □

One important application of local sections is the following proposition, which gives a very simple criterion for deciding which maps out of the base space of a covering are smooth.

Proposition 2.7. *Suppose $\pi\colon \widetilde{M} \to M$ is a smooth covering map and N is any smooth manifold. A map $F\colon M \to N$ is smooth if and only if $F \circ \pi\colon \widetilde{M} \to M$ is smooth:*

$$\begin{tikzcd} \widetilde{M} \arrow[d, "\pi"'] \arrow[dr, "F \circ \pi"] & \\ M \arrow[r, "F"'] & N. \end{tikzcd}$$

Proof. One direction is obvious by composition. Suppose conversely that $F \circ \pi$ is smooth, and let $p \in M$ be arbitrary. By the preceding lemma, there is a neighborhood U of p and a smooth local section $\sigma\colon U \to \widetilde{M}$, so that $\pi \circ \sigma = \mathrm{Id}_U$. Then the restriction of F to U satisfies

$$F|_U = F \circ \mathrm{Id}_U = F \circ (\pi \circ \sigma) = (F \circ \pi) \circ \sigma,$$

which is a composition of smooth maps. Thus F is smooth on U. Since F is smooth in a neighborhood of each point, it is smooth. □

The next proposition shows that every covering space of a smooth manifold is itself a smooth manifold.

Proposition 2.8. *If M is a smooth manifold and $\pi\colon \widetilde{M} \to M$ is any topological covering map, then $\widetilde{M}$ has a unique smooth manifold structure such that π is a smooth covering map.*

Proof. Because π is, in particular, a local homeomorphism, it is clear that $\widetilde{M}$ is locally Euclidean.

Let p, q be distinct points in $\widetilde{M}$. If $\pi(p) = \pi(q)$ and $U \subset M$ is an evenly covered open set containing $\pi(p)$, then the components of $\pi^{-1}(U)$ containing p and q are disjoint open subsets of $\widetilde{M}$ separating p and q.

On the other hand, if $\pi(p) \neq \pi(q)$, there are disjoint open sets U and V containing $\pi(p)$ and $\pi(q)$, respectively, and then $\pi^{-1}(U)$ and $\pi^{-1}(V)$ are open subsets of $\widetilde{M}$ separating p and q. Thus $\widetilde{M}$ is Hausdorff.

The fibers of π are countable, because the fundamental group of M is countable and acts transitively on each fiber [Lee00, Theorems 8.11 and 11.21]. Thus if $\{U_i\}_{i\in\mathbb{N}}$ is a countable basis for the topology of $\widetilde{M}$, it is easy to check that the set of all components of $\pi^{-1}(U_i)$ as i ranges over $\mathbb{N}$ forms a countable basis for the topology of $\widetilde{M}$, so $\widetilde{M}$ is second countable.

Any point $p \in M$ has an evenly covered neighborhood U. Shrinking U if necessary, we may assume also that it is the domain of a coordinate map $\varphi\colon U \to \mathbb{R}^n$. Letting $\widetilde{U}$ be a component of $\pi^{-1}(U)$ and $\widetilde{\varphi} = \varphi \circ \pi\colon \widetilde{U} \to \mathbb{R}^n$, it is clear that $(\widetilde{U}, \widetilde{\varphi})$ is a chart on $\widetilde{M}$. If two such charts $(\widetilde{U}, \widetilde{\varphi})$ and $(\widetilde{V}, \widetilde{\psi})$ overlap, the transition map can be written

$$\begin{aligned}\widetilde{\psi} \circ \widetilde{\varphi}^{-1} &= (\psi \circ \pi|_{\widetilde{U}\cap\widetilde{V}}) \circ (\varphi \circ \pi|_{\widetilde{U}\cap\widetilde{V}})^{-1} \\ &= \psi \circ \pi|_{\widetilde{U}\cap\widetilde{V}} \circ (\pi|_{\widetilde{U}\cap\widetilde{V}})^{-1} \circ \varphi^{-1} \\ &= \psi \circ \varphi^{-1},\end{aligned}$$

which is smooth. Thus the collection of all such charts defines a smooth structure on $\widetilde{M}$. The uniqueness of this smooth structure is left as an exercise. □

Exercise 2.7. Prove that the smooth structure constructed above on $\widetilde{M}$ is the unique one such that π is a smooth covering map. [Hint: Use the existence of smooth local sections.]

Lie Groups

A *Lie group* is a smooth manifold G that is also a group in the algebraic sense, with the property that the multiplication map $m\colon G \times G \to G$ and inversion map $i\colon G \to G$, given by

$$m(g, h) = gh, \qquad i(g) = g^{-1},$$

are both smooth. Because smooth maps are continuous, a Lie group is, in particular, a *topological group* (a topological space with a group structure such that the multiplication and inversion maps are continuous).

The group operation in an arbitrary Lie group will be denoted by juxtaposition, except in certain abelian groups such as $\mathbb{R}^n$ in which the operation is usually written additively. It is traditional to denote the identity element of an arbitrary Lie group by the symbol e (for German *Einselement*, "unit element"), and we will follow this convention, except in specific examples in which there are more common notations (such as I or I_n for the identity matrix in a matrix group, or 0 for the identity element in $\mathbb{R}^n$).

The following alternative characterization of the smoothness condition is sometimes useful.

Lemma 2.9. *Suppose G is a smooth manifold with a group structure such that the map $G \times G \to G$ given by $(g,h) \mapsto gh^{-1}$ is smooth. Then G is a Lie group.*

Exercise 2.8. Prove Lemma 2.9.

Example 2.10 (Lie Groups). Each of the following manifolds is a Lie group with the indicated group operation.

(a) The *general linear group* $\mathrm{GL}(n,\mathbb{R})$ is the set of invertible $n \times n$ matrices with real entries. It is a group under matrix multiplication, and it is an open submanifold of the vector space $\mathrm{M}(n,\mathbb{R})$ as we observed in Chapter 1. Multiplication is smooth because the matrix entries of a product matrix AB are polynomials in the entries of A and B. Inversion is smooth because Cramer's rule expresses the entries of A^{-1} as rational functions of the entries of A.

(b) The *complex general linear group* $\mathrm{GL}(n,\mathbb{C})$ is the group of complex $n \times n$ matrices under matrix multiplication. It is an open submanifold of $\mathrm{M}(n,\mathbb{C})$ and thus a $2n^2$-dimensional smooth manifold, and it is a Lie group because matrix products and inverses are smooth functions of the real and imaginary parts of the matrix entries.

(c) If V is any real or complex vector space, we let $\mathrm{GL}(V)$ denote the group of invertible linear transformations from V to itself. If V is finite-dimensional, any basis for V determines an isomorphism of $\mathrm{GL}(V)$ with $\mathrm{GL}(n,\mathbb{R})$ or $\mathrm{GL}(n,\mathbb{C})$, with $n = \dim V$, so $\mathrm{GL}(V)$ is a Lie group.

(d) The real number field $\mathbb{R}$ and Euclidean space $\mathbb{R}^n$ are Lie groups under addition, because the coordinates of $x - y$ are smooth (linear!) functions of (x, y).

(e) The set $\mathbb{R}^*$ of nonzero complex numbers is a 1-dimensional Lie group under multiplication. (In fact, it is exactly $\mathrm{GL}(1,\mathbb{R})$, if we identify a 1×1 matrix with the corresponding real number.) The subset $\mathbb{R}^+$ of positive real numbers is an open subgroup, and is thus itself a 1-dimensional Lie group.

(f) The set $\mathbb{C}^*$ of nonzero complex numbers is a 2-dimensional Lie group under complex multiplication, which can be identified with $\mathrm{GL}(1,\mathbb{C})$.

(g) The circle $\mathbb{S}^1 \subset \mathbb{C}^*$ is a smooth manifold and a group under complex multiplication. Using appropriate angle functions as local coordinates on open subsets of $\mathbb{S}^1$, multiplication and inversion have the smooth coordinate expressions $(\theta_1, \theta_2) \mapsto \theta_1 + \theta_2$ and $\theta \mapsto -\theta$, and therefore $\mathbb{S}^1$ is a Lie group, called the *circle group*.

(h) Any product of Lie groups is a Lie group with the product manifold structure and the direct product group structure, as you can easily check.

(i) The n-torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ is an n-dimensional abelian Lie group.

(j) Any finite or countable group with the discrete topology is a zero-dimensional Lie group. We will call any such group a *discrete group*.

Let G be an arbitrary Lie group. Any element $g \in G$ defines maps $L_g, R_g \colon G \to G$, called *left translation* and *right translation*, respectively, by

$$L_g(h) = gh, \qquad R_g(h) = hg.$$

Because L_g can be written as the composition of smooth maps

$$G \xrightarrow{\iota_g} G \times G \xrightarrow{m} G,$$

where $\iota_g(h) = (g, h)$ and m is multiplication, it follows that L_g is smooth. It is actually a diffeomorphism of G, because $L_{g^{-1}}$ is a smooth inverse for it. Similarly, $R_g \colon G \to G$ is a diffeomorphism. Observe that, given any two points $g_1, g_2 \in G$, there is a unique left translation of G taking g_1 to g_2, namely left translation by $g_2 g_1^{-1}$. Many of the important properties of Lie groups follow, as you will see repeatedly in later chapters, from the fact that we can systematically map any point to any other by such a global diffeomorphism.

If G and H are Lie groups, a *Lie group homomorphism* from G to H is a smooth map $F \colon G \to H$ that is also a group homomorphism. It is called a *Lie group isomorphism* if it is also a diffeomorphism, which implies that it has an inverse that is also a Lie group homomorphism. We will sometimes use the abbreviated terms *Lie homomorphism* and *Lie isomorphism* when they will not cause confusion.

Example 2.11 (Lie Group Homomorphisms).

(a) The inclusion map $\mathbb{S}^1 \hookrightarrow \mathbb{C}^*$ is a Lie group homomorphism.

(b) The map $\exp \colon \mathbb{R} \to \mathbb{R}^*$ given by $\exp(t) = e^t$ is smooth, and is a Lie group homomorphism because $e^{(s+t)} = e^s e^t$. (Note that $\mathbb{R}$ is considered as a Lie group under addition, while $\mathbb{R}^*$ is a Lie group under multiplication.) The image of exp is the open subgroup $\mathbb{R}^+$ consisting of positive real numbers, and $\exp \colon \mathbb{R} \to \mathbb{R}^+$ is a Lie group isomorphism with inverse $\log \colon \mathbb{R}^+ \to \mathbb{R}$.

(c) Similarly, $\exp \colon \mathbb{C} \to \mathbb{C}^*$ given by $\exp z = e^z$ is a Lie group homomorphism. It is surjective but not injective, because its kernel consists of the complex numbers of the form $2\pi i k$ where k is an integer.

(d) The covering map $\varepsilon\colon \mathbb{R} \to \mathbb{S}^1$ defined by $\varepsilon(t) = e^{2\pi i t}$ is a Lie group homomorphism whose kernel is the set $\mathbb{Z}$ of integers. Similarly the covering map $\varepsilon^n\colon \mathbb{R}^n \to \mathbb{T}^n$ defined by $\varepsilon^n(t_1, \dots, t_n) = (e^{2\pi i t_1}, \dots, e^{2\pi i t_n})$ is a Lie group homomorphism whose kernel is $\mathbb{Z}^n$.

(e) The determinant map $\det\colon \mathrm{GL}(n,\mathbb{R}) \to \mathbb{R}^*$ is smooth because $\det A$ is a polynomial in the matrix entries of A. It is a Lie group homomorphism because $\det(AB) = (\det A)(\det B)$. Similarly, $\det\colon \mathrm{GL}(n,\mathbb{C}) \to \mathbb{C}^*$ is a Lie group homomorphism.

Representations

One kind of Lie group homomorphism plays a fundamental role in many branches of mathematics. If G is a Lie group and V is a vector space, any Lie group homomorphism $\rho\colon G \to \mathrm{GL}(V)$ is called a *representation* of G. If V is finite-dimensional, then ρ is called a *finite-dimensional representation*. The study of representations of Lie groups is a vast field in its own right, and we can do no more than touch on it here.

Example 2.12 (Lie Group Representations).

(a) The identity map $\mathrm{GL}(n,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{R}) = \mathrm{GL}(\mathbb{R}^n)$ is a representation of $\mathrm{GL}(n,\mathbb{R})$, called the *defining representation*. The defining representation of $\mathrm{GL}(n,\mathbb{C})$ is defined similarly.

(b) The inclusion map $\mathbb{S}^1 \hookrightarrow \mathbb{C}^* \cong \mathrm{GL}(1,\mathbb{C})$ is a representation of the circle group. More generally, the map $\rho\colon \mathbb{T}^n \to \mathrm{GL}(n,\mathbb{C})$ given by

$$\rho(z^1, \dots, z^n) = \begin{pmatrix} z^1 & 0 & \dots & 0 \\ 0 & z^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z^n \end{pmatrix}$$

is a representation of $\mathbb{T}^n$.

(c) Define a map $\tau\colon \mathbb{R}^n \to \mathrm{GL}(n+1,\mathbb{R})$ by sending $X \in \mathbb{R}^n$ to the matrix $\tau(X)$ defined in block form by

$$\tau(X) = \begin{pmatrix} I_n & X \\ 0 & 1 \end{pmatrix},$$

where I_n is the $n \times n$ identity matrix and X is thought of as an $n \times 1$ column matrix. Then a simple computation shows that τ is a representation.

Bump Functions and Partitions of Unity

In this section, we develop bump functions and partitions of unity, which are tools for patching together local smooth objects into global ones. These tools are of central importance in smooth manifold theory and will reappear throughout the book.

All of these tools are based on the existence of smooth functions that are positive in a specified part of a manifold and identically zero in some other part. If f is any real-valued or vector-valued function on a smooth manifold M, the *support* of f, denoted by $\operatorname{supp} f$, is the closure of the set of points where f is nonzero:

$$\operatorname{supp} f = \overline{\{p \in M : f(p) \neq 0\}}.$$

If $\operatorname{supp} f$ is contained in some set U, we say f is *supported in* U. A function f is said to be *compactly supported* if $\operatorname{supp} f$ is a compact set. Clearly every function on a compact manifold is compactly supported.

We begin by defining a smooth function on the real line that is zero for $t \leq 0$ and positive for $t > 0$.

Lemma 2.13. *The function $f \colon \mathbb{R} \to \mathbb{R}$ defined by*

$$f(t) = \begin{cases} e^{-1/t} & t > 0, \\ 0 & t \leq 0 \end{cases}$$

is smooth.

Proof. It is clearly smooth on $\mathbb{R} \smallsetminus \{0\}$, so we need only show that all derivatives of f exist and are continuous at the origin. We begin by noting that f is continuous because $\lim_{t \searrow 0} e^{-1/t} = 0$. In fact, a standard application of l'Hôpital's rule and induction shows that for any integer $k \geq 0$,

$$\lim_{t \searrow 0} \frac{e^{-1/t}}{t^k} = \lim_{t \searrow 0} \frac{t^{-k}}{e^{1/t}} = 0. \tag{2.1}$$

We will show by induction that for $t > 0$, the kth derivative of f is of the form

$$f^{(k)}(t) = \frac{p_k(t)}{t^{2k}} e^{-1/t} \tag{2.2}$$

for some polynomial $p_k(t)$. It is clearly true (with $p_0(t) = 1$) for $k = 0$, so suppose it is true for some $k \geq 0$. By the product rule,

$$\begin{aligned} f^{(k+1)}(t) &= \frac{p_k'(t)}{t^{2k}} e^{-1/t} - \frac{2kp_k(t)}{t^{2k+1}} e^{-1/t} + \frac{p_k(t)}{t^{2k}} \frac{1}{t^2} e^{-1/t} \\ &= \frac{t^2 p_k'(t) - 2ktp_k(t) + p_k(t)}{t^{2k+2}} e^{-1/t}, \end{aligned}$$

which is of the required form. Note that (2.2) and (2.1) imply that

$$\lim_{t \searrow 0} f^{(k)}(t) = 0, \tag{2.3}$$

since a polynomial is continuous at 0.

Finally, we prove that for each $k \geq 0$,

$$f^{(k)}(0) = 0.$$

For $k = 0$, this is true by definition, so assume that it is true for some $k \geq 0$. It suffices to show that f has one-sided derivatives from both sides and that they are equal. Clearly the derivative from the left is zero. Using (2.1) again, we compute

$$f^{(k+1)}(0) = \lim_{t \searrow 0} \frac{\dfrac{p_k(t)}{t^{2k}} e^{-1/t} - 0}{t} = \lim_{t \searrow 0} \frac{p_k(t)}{t^{2k+1}} e^{-1/t} = 0.$$

By (2.3), this implies each $f^{(k)}$ is continuous, so f is smooth. □

Lemma 2.14. *There exists a smooth function $h \colon \mathbb{R} \to [0,1]$ such that $h(t) \equiv 1$ for $t \leq 1$, $0 < h(t) < 1$ for $1 < t < 2$, and $h(t) \equiv 0$ for $t \geq 2$.*

Proof. Let f be the function of the previous lemma, and set

$$h(t) = \frac{f(2-t)}{f(2-t) + f(t-1)}.$$

Note that the denominator is positive for all t, because at least one of the expressions $2-t$ or $t-1$ is always positive. Since $f \geq 0$ always, it is easy to check that $h(t)$ is always between 0 and 1, and is zero when $t \geq 2$. When $t \leq 1$, $f(t-1) = 0$, so $h(t) \equiv 1$ there. □

A function with the properties of h in this lemma is usually called a *cutoff function*.

Lemma 2.15. *There is a smooth function $H \colon \mathbb{R}^n \to [0,1]$ such that $H \equiv 1$ on $\overline{B}_1(0)$ and* $\operatorname{supp} H = \overline{B}_2(0)$.

Proof. Just set $H(x) = h(|x|)$, where h is the function of the preceding lemma. Clearly H is smooth on $\mathbb{R}^n \smallsetminus \{0\}$, because it is a composition of smooth functions there. Since it is identically equal to 1 on $B_1(0)$, it is smooth there too. □

The function H constructed in this lemma is an example of a *bump function*—a smooth function that is equal to 1 on a specified closed set (in this case $\overline{B}_1(0)$) and is supported in a specified open set (in this case any open set containing $\overline{B}_2(0)$). Later, we will generalize this notion to manifolds.

To use bump functions effectively on a manifold, we need to construct some special covers. A collection of subsets $\{U_\alpha\}_{\alpha\in A}$ of a topological space X is said to be *locally finite* if each point $p \in X$ has a neighborhood that intersects at most finitely many of the sets U_α.

Exercise 2.9. Show that an *open* cover $\{U_\alpha\}$ of X is locally finite if and only if each U_α intersects U_β for only finitely many β. Give a counterexample if the sets of the cover are not assumed to be open.

Recall that a subset of a topological space is said to be *precompact* if its closure is compact.

Lemma 2.16. *Every topological manifold admits a countable, locally finite cover by precompact open sets.*

Proof. Let $\{U_i\}$ be any countable cover of M by precompact open sets (for example, small coordinate balls will do). We will define another open cover $\{V_j\}$ with the same properties that also satisfies

$$\overline{V}_{j-1} \subset V_j, \quad j \geq 2. \tag{2.4}$$

Let $V_1 = U_1$. Assume by induction that precompact open sets V_j have been defined for $j = 1, \dots, k$ satisfying $U_j \subset V_j$ and (2.4). Because $\overline{V}_k$ is compact and covered by $\{U_i\}$, there is some m_k such that $\overline{V}_k \subset U_1 \cup \cdots \cup U_{m_k}$. By increasing m_k if necessary, we may assume that $m_k > k$. Let $V_{k+1} = U_1 \cup \cdots \cup U_{m_k}$; then clearly (2.4) is satisfied with $j = k+1$, $U_{k+1} \subset V_{k+1}$, and $\overline{V}_{k+1} = \overline{U}_1 \cup \cdots \cup \overline{U}_{m_k}$ is compact. Since $U_j \subset V_j$ for each j, $\{V_j\}$ is a cover of M.

Now just set $W_j = V_j \smallsetminus \overline{V}_{j-2}$. Since $\overline{W}_j$ is a closed subset of the compact set $\overline{V}_j$, it is compact. If $p \in M$ is arbitrary, then $p \in W_k$, where k is the smallest integer such that $p \in V_k$. Clearly W_k has nonempty intersection only with W_{k-1} and W_{k+1}, so the cover $\{W_j\}$ is locally finite. □

Given an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in A}$ of a topological space, another open cover $\mathcal{V} = \{V_\beta\}_{\beta\in B}$ is called a *refinement* of $\mathcal{U}$ if for each $V_\beta \in \mathcal{V}$ there exists some $U_\alpha \in \mathcal{U}$ such that $V_\beta \subset U_\alpha$. A topological space is said to be *paracompact* if every open cover admits a locally finite refinement.

The key topological result we need is that every manifold is paracompact. In fact, for future use we will show something stronger: that every open cover admits a locally finite refinement of a particularly nice type. We say an open cover $\{W_i\}$ of M is *regular* if it satisfies the following properties:

(i) The cover $\{W_i\}$ is countable and locally finite.

(ii) For each i, there exists a diffeomorphism $\psi_i\colon W_i \to B_3(0) \subset \mathbb{R}^n$.

(iii) The collection $\{U_i\}$ still covers M, where $U_i = \psi_i^{-1}(B_1(0))$.

Proposition 2.17. *Let M be a smooth manifold. Every open cover of M has a regular refinement. In particular, M is paracompact.*

Proof. Let $\mathfrak{X}$ be any open cover of M, and let $\{V_j\}$ be a countable, locally finite cover of M by precompact open sets. For each $p \in M$, let (W_p, ψ_p) be a coordinate chart centered at p such that

- $\psi_p(W_p) = B_3(0)$;
- W_p is contained in one of the open sets of $\mathfrak{X}$; and
- if $p \in V_j$, then $W_p \subset V_j$ as well.

(The last condition is possible because of the local finiteness of $\{V_j\}$.) Let $U_p = \psi_p^{-1}(B_1(0))$.

For each k, the collection $\{U_p : p \in \overline{V}_k\}$ is an open cover of $\overline{V}_k$. By compactness, $\overline{V}_k$ is covered by finitely many of these sets. Call the sets $U_{k,1}, \dots, U_{k,m(k)}$, and let $(W_{k,1}, \psi_{k,1}), \dots, (W_{k,m(k)}, \psi_{k,m(k)})$ denote the corresponding coordinate charts. The collection of all the sets $\{W_{k,i}\}$ as k and i vary is clearly a countable open cover that refines $\mathfrak{X}$ and satisfies (ii) and (iii) above. To show it is a regular cover, we need only show it is locally finite.

For any given k, each set $W_{k,i}$ is by construction contained in some V_j such that $\overline{V}_k \cap V_j \neq 0$. The compact set $\overline{V}_k$ is covered by finitely many V_j's, and each such V_j intersects at most finitely many others, so there are only finitely many values of j for which $W_{k,i}$ and $W_{j,i}$ can have nonempty intersection. Since there are only finitely many sets $W_{j,i}$ for each j, the cover we have constructed is locally finite. □

Now let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be an arbitrary open cover of a smooth manifold M. A *partition of unity subordinate to* $\mathcal{U}$ is a collection of smooth functions $\{\varphi_\alpha \colon M \to \mathbb{R}\}_{\alpha \in A}$, with the following properties:

(i) $0 \le \varphi_\alpha(x) \le 1$ for all $\alpha \in A$ and all $x \in M$;

(ii) $\operatorname{supp} \varphi_\alpha \subset U_\alpha$;

(iii) the set of supports $\{\operatorname{supp} \varphi_\alpha\}_{\alpha \in A}$ is locally finite; and

(iv) $\sum_{\alpha \in A} \varphi_\alpha(x) = 1$ for all $x \in M$.

Because of the local finiteness condition (iii), the sum in (iv) actually has only finitely many nonzero terms in a neighborhood of each point, so there is no issue of convergence.

Theorem 2.18 (Existence of Partitions of Unity). *If M is a smooth manifold and $\mathfrak{X} = \{X_\alpha\}_{\alpha \in A}$ is any open cover of M, there exists a partition of unity subordinate to $\mathfrak{X}$.*

Proof. Let $\{W_i\}$ be a regular refinement of $\mathcal{X}$. For each i, let $\psi_i \colon W_i \to B_3(0)$ be the diffeomorphism whose existence is guaranteed by the definition of a regular cover, and let

$$\begin{aligned} U_i &= \psi_i^{-1}(B_1(0)), \\ V_i &= \psi_i^{-1}(B_2(0)). \end{aligned}$$

For each i, define a function $f_i \colon M \to \mathbb{R}$ by

$$f_i = \begin{cases} H \circ \psi_i & \text{on } W_i, \\ 0 & \text{on } M \smallsetminus \overline{V}_i, \end{cases}$$

where $H \colon \mathbb{R}^n \to \mathbb{R}$ is the bump function of Lemma 2.15. On the set $W_i \smallsetminus \overline{V}_i$ where the two definitions overlap, both definitions yield the zero function, so f_i is well defined and smooth, and $\operatorname{supp} f_i \subset W_i$.

Define new functions $g_i \colon M \to \mathbb{R}$ by

$$g_i(x) = \frac{f_i(x)}{\sum_j f_j(x)}.$$

Because of the local finiteness of the cover $\{W_i\}$, the sum in the denominator has only finitely many nonzero terms in a neighborhood of each point and thus defines a smooth function. Because $f_i \equiv 1$ on U_i and every point of M is in some U_i, the denominator is always positive, so g_i is a smooth function on M. It is immediate from the definition that $0 \le g_i \le 1$ and $\sum_i g_i \equiv 1$.

Finally, we need to re-index our functions so that they are indexed by the same set A as our open cover. For each i, there is some index $a(i) \in A$ such that $W_i \subset X_{a(i)}$. For each $\alpha \in A$, define $\varphi_\alpha \colon M \to \mathbb{R}$ by

$$\varphi_\alpha = \sum_{i : a(i) = \alpha} g_i.$$

Each φ_α is smooth and satisfies $0 \le \varphi_\alpha \le 1$ and $\operatorname{supp} \varphi_\alpha \subset X_\alpha$. Moreover, the set of supports $\{\operatorname{supp} \varphi_\alpha\}_{\alpha \in A}$ is still locally finite, and $\sum_\alpha \varphi_\alpha \equiv \sum_i g_i \equiv 1$, so this is the desired partition of unity. □

Now we can extend the notion of bump functions to arbitrary closed sets in manifolds.

Corollary 2.19 (Existence of Bump Functions). *Let M be a smooth manifold. For any closed set $A \subset M$ and any open set U containing A, there exists a smooth function $\varphi \colon M \to \mathbb{R}$ such that $\varphi \equiv 1$ on A and $\operatorname{supp} \varphi \subset U$.*

Proof. Let $U_0 = U$ and $U_1 = M \smallsetminus A$, and let $\{\varphi_0, \varphi_1\}$ be a partition of unity subordinate to the open cover $\{U_0, U_1\}$. Because $\varphi_1 \equiv 0$ on A and therefore $\varphi_0 = \sum_i \varphi_i = 1$ there, the function φ_0 has the required properties. □

Any function with the properties described in this lemma is called a *bump function* for A supported in U.

As our first application, we will prove an important lemma regarding the extension of smooth functions from closed subsets. Suppose M is a smooth manifold We say that a function defined on an arbitrary subset $A \subset M$ is *smooth on* A if it admits a smooth extension to some open set U containing A.

Lemma 2.20 (Extension Lemma). *Let M be a smooth manifold, and suppose f is a smooth function defined on a closed subset $A \subset M$. For any open set U containing A, there exists a smooth function $\widetilde{f} \in C^\infty(M)$ such that $\widetilde{f}|_A = f|_A$ and $\operatorname{supp} \widetilde{f} \subset U$.*

Proof. The fact that f is smooth on A means by definition that f extends to a smooth function, still denoted by f, on some neighborhood W of A. Replacing W by $W \cap U$, we may assume that $W \subset U$. Letting φ be a bump function for A supported in W, we can define

$$\widetilde{f}(p) = \begin{cases} \varphi(p) f(p), & p \in W, \\ 0, & p \in M \smallsetminus \operatorname{supp} \varphi. \end{cases}$$

This function is clearly a smooth extension of f whose support is contained in W and thus in U. □

Exercise 2.10. Show that the assumption that A is closed is necessary in the extension lemma, by giving an example of a smooth function on a nonclosed subset of a manifold that admits no smooth extension to the whole manifold.

The extension lemma, by the way, illustrates an essential difference between smooth manifolds and real-analytic manifolds, because it is decidedly false on the latter. This follows from the fact that real-analytic functions that agree on an open set must agree on their whole common domain (assuming it is connected).

Problems

2-1. Compute the coordinate representation for each of the following maps, using stereographic coordinates for spheres (see Problem 1-3); use this to conclude that each map is smooth.

(a) $A\colon \mathbb{S}^n \to \mathbb{S}^n$ is the antipodal map $A(x) = -x$.

(b) $F\colon \mathbb{S}^3 \to \mathbb{S}^2$ is given by $F(z,w) = (z\overline{w}+w\overline{z}, iw\overline{z}-iz\overline{w}, z\overline{z}-w\overline{w})$, where we think of $\mathbb{S}^3$ as the subset $\{(w,z) : |w|^2 + |z|^2 = 1\}$ of $\mathbb{C}^2$.

2-2. Let $M\overline{B^n}$, the closed unit ball in $\mathbb{R}^n$, thought of as a smooth manifold with boundary. Show that the inclusion map $M \hookrightarrow \mathbb{R}^n$ is smooth.

2-3. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be the atlases for $\mathbb{R}$ defined by $\mathcal{A}_1 = \{(\mathbb{R}, \mathrm{Id})\}$, and $\mathcal{A}_2 = \{(\mathbb{R}, \psi)\}$, where $\psi(x) = x^3$. Let $f\colon \mathbb{R} \to \mathbb{R}$ be any function. Determine necessary and sufficient conditions on f so that it will be:

(a) a smooth map $(\mathbb{R}, \mathcal{A}_2) \to (\mathbb{R}, \mathcal{A}_1)$;

(b) a smooth map $(\mathbb{R}, \mathcal{A}_1) \to (\mathbb{R}, \mathcal{A}_2)$.

2-4. For any topological space M, let $C(M)$ denote the vector space of continuous functions $f\colon M \to \mathbb{R}$. If $F\colon M \to N$ is a continuous map, define $F^*\colon C(N) \to C(M)$ by $F^*(f) = f \circ F$.

(a) Show that F^* is linear.

(b) If M and N are smooth manifolds, show that F is smooth if and only if $F^*(C^\infty(N)) \subset C^\infty(M)$.

(c) If $F\colon M \to N$ is a homeomorphism between smooth manifolds, show that it is a diffeomorphism if and only if $F^*\colon C^\infty(N) \to C^\infty(M)$ is an isomorphism.

Thus in a certain sense the entire smooth structure of M is encoded in the space $C^\infty(M)$.

3
The Tangent Bundle

One of the key tools in our study of smooth manifolds will be the idea of *linear approximation.* This is probably familiar from your study of calculus in Euclidean spaces, where for example a function of one variable can be approximated by its tangent line, a curve in $\mathbb{R}^n$ by its tangent vector, or a map from $\mathbb{R}^n$ to $\mathbb{R}^m$ by its total derivative (see the Appendix).

In order to make sense of linear approximations on manifolds, we will need to introduce the notion of the tangent space to a manifold at a point. Because of the abstractness of the definition of a smooth manifold, this takes some work, which we carry out in this chapter.

We begin by studying a much more concrete object: geometric tangent vectors in $\mathbb{R}^n$, which can be thought of as "arrows" attached to a particular point in $\mathbb{R}^n$. Because the definition of smooth manifolds is built around the idea of identifying which functions are smooth, the property of geometric tangent vectors is its action on smooth functions as a "directional derivative." The key observation about geometric tangent vectors, which we prove in the first section of this chapter, is that the process of taking directional derivatives gives a natural one-to-one correspondence between geometric tangent vectors and linear maps from $C^\infty(\mathbb{R}^n)$ to $\mathbb{R}$ satisfying the product rule (called "derivations"). With this as motivation, we then *define* tangent vectors on a smooth manifold as derivations of $C^\infty(M)$.

In the second section of the chapter, we connect the abstract definition of to our concrete geometric picture by showning that any coordinate chart (U, φ) gives a natural isomorphism between the vector space of tangent vectors at $p \in M$ with the space of geometric tangent vectors at $\varphi(p) \in \mathbb{R}^n$. Thus any coordinate chart yields a basis for each tangent space. Later in the chapter, we describe how to do concrete computations in such a basis. This also leads to the definition of a natural smooth manifold structure on the union of all the tangent spaces at all points of the manifold, called the tangent bundle. Using this structure, we define vector fields, which are continuous or smooth functions that attach a tangent vector to each point on the manifold.

Tangent Vectors

Imagine a manifold in Euclidean space. For concreteness, let us take it to be the unit sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$. What do we mean by a "tangent vector" at a point of $\mathbb{S}^n$? Before we can answer this question, we have to come to terms with a dichotomy in the way we think about an element of $\mathbb{R}^n$. On one hand, we usually think of it as a *point* in space, whose only property is its location, expressed by the coordinates $(x^1, \ldots, x^n)$. On the other hand, when doing calculus we sometimes think of it instead as a *vector*, which is an object that has magnitude and direction, but whose location is irrelevant. A vector $v = v^i e_i$ can be visualized as an arrow with its initial point anywhere in $\mathbb{R}^n$; what is relevant from the vector point of view is only which direction it points and how long it is.

What we really have in mind when we work with tangent vectors is a separate copy of $\mathbb{R}^n$ at each point. When we talk about the set of vectors tangent to the sphere at a point a, for example, we are imagining them as living in a copy of $\mathbb{R}^n$ with its origin translated to a.

Geometric Tangent Vectors

Here is a preliminary definition of tangent vectors in Euclidean space. Let us define the *geometric tangent space* to $\mathbb{R}^n$ at the point $a \in \mathbb{R}^n$, denoted by $\mathbb{R}^n_a$, to be the set $\{a\} \times \mathbb{R}^n$. More explicitly,

$$\mathbb{R}^n_a = \{(a, v) : v \in \mathbb{R}^n\}.$$

A *geometric tangent vector* in $\mathbb{R}^n$ is an element of this space. As a matter of notation, we will abbreviate (a, v) as v_a (or sometimes $v|_a$ if it is clearer, for example if v itself has a subscript). We think of v_a as the vector v with its initial point at a. This set $\mathbb{R}^n_a$ is a real vector space (obviously isomorphic to $\mathbb{R}^n$ itself) under the natural operations

$$\begin{aligned} v_a + w_a &= (v + w)_a, \\ c(v_a) &= (cv)_a. \end{aligned}$$

The vectors $e_i|_a$, $i = 1, \ldots, n$, are a basis for $\mathbb{R}^n_a$. In fact, as a vector space, $\mathbb{R}^n_a$ is essentially the same as $\mathbb{R}^n$ itself; the only reason we add the index a

is so that the geometric tangent spaces $\mathbb{R}^n_a$ and $\mathbb{R}^n_b$ at distinct points a and b will be disjoint sets.

With this definition we can, for example, identify the tangent space to $\mathbb{S}^n$ at a point $a \in \mathbb{S}^n$ as the subspace of $\mathbb{R}^n_a$ consisting of those vectors that are orthogonal to the radial unit vector through a, with respect to the usual inner product on $\mathbb{R}^n$ transported to $\mathbb{R}^n_a$ via the natural isomorphism $\mathbb{R}^n \cong \mathbb{R}^n_a$.

The problem with this definition, however, is that it gives us no clue as to how we might set about defining tangent vectors on an arbitrary smooth manifold, where there is no ambient Euclidean space. So we need to look for another characterization of tangent vectors that might make sense on a manifold.

The only things we have to work with on smooth manifolds so far are smooth functions, smooth maps, and coordinate charts. Now one thing that a Euclidean tangent vector provides is a means of taking "directional derivatives" of functions. For example, for any geometric tangent vector $v_a \in \mathbb{R}^n_a$, we can define a map $\widetilde{v}_a \colon C^\infty(\mathbb{R}^n) \to \mathbb{R}$ by taking the directional derivative in the direction v at a:

$$\widetilde{v}_a f = D_v f(a) = \left.\frac{d}{dt}\right|_{t=0} f(a + tv). \tag{3.1}$$

This operation is linear and satisfies the product rule:

$$\widetilde{v}_a(fg) = f(a)\widetilde{v}_a(g) + g(a)\widetilde{v}_a(f).$$

If $v_a = v^i e_i|_a$ in terms of the standard basis, then by the chain rule $\widetilde{v}_a f$ can be written more concretely as

$$\widetilde{v}_a f = v^i \frac{\partial f}{\partial x^i}(a).$$

(Here we are using the summation convention as usual, so the expression on the right-hand side is understood to be summed over $i = 1$ to n. This sum is consistent with our index convention if we stipulate that an upper index "in the denominator" is to be regarded as a lower index.) For example, if $v_a = e_j|_a$, then

$$\widetilde{v}_a f = \frac{\partial f}{\partial x^j}(a).$$

With this construction in mind, we make the following definition. A linear map $X \colon C^\infty(\mathbb{R}^n) \to \mathbb{R}$ is called a *derivation* at a if it satisfies the following product rule:

$$X(fg) = f(a)Xg + g(a)Xf. \tag{3.2}$$

Let $T_a(\mathbb{R}^n)$ denote the set of all derivations of $C^\infty(\mathbb{R}^n)$ at a. Clearly $T_a(\mathbb{R}^n)$ is a vector space under the operations

$$(X+Y)f = Xf + Yf$$
$$(cX)f = c(Xf).$$

Lemma 3.1 (Properties of Derivations). *Suppose $a \in \mathbb{R}^n$ and $X \in T_a(\mathbb{R}^n)$.*

(*a*) *If f is a constant function, then $Xf = 0$.*

(*b*) *If $f(a) = g(a) = 0$, then $X(fg) = 0$.*

Proof. It suffices to prove (a) for the constant function $f_1(x) \equiv 1$, for then $f(x) \equiv c$ implies $Xf = X(cf_1) = cXf_1 = 0$ by linearity. For f_1, it follows from the product rule:

$$Xf_1 = X(f_1 f_1) = f_1(a)Xf_1 + f_1(a)Xf_1 = 2Xf_1,$$

which implies that $Xf_1 = 0$. Similarly, (b) also follows from the product rule:

$$X(fg) = f(a)Xg + g(a)Xf = 0 + 0 = 0.$$

□

Now let $v_a \in \mathbb{R}^n_a$ be a geometric tangent vector at a. By the product rule, the map $\widetilde{v}_a\colon C^\infty(\mathbb{R}^n) \to \mathbb{R}$ defined by (3.1) is a derivation. As the following proposition shows, every derivation is of this form.

Proposition 3.2. *For any $a \in \mathbb{R}^n$, the map $v_a \mapsto \widetilde{v}_a$ is an isomorphism from $\mathbb{R}^n_a$ onto $T_a(\mathbb{R}^n)$.*

Proof. The map $v_a \mapsto \widetilde{v}_a$ is linear, as is easily checked. To see that it is injective, suppose $v_a \in \mathbb{R}^n_a$ has the property that $\widetilde{v}_a$ is the zero derivation. Writing $v_a = v^i e_i|_a$ in terms of the standard basis, and taking f to be the jth coordinate function $x^j\colon \mathbb{R}^n \to \mathbb{R}$, thought of as a smooth function on $\mathbb{R}^n$, we find

$$0 = \widetilde{v}_a(x^j) = v^i \left.\frac{\partial}{\partial x^i}(x^j)\right|_{x=a} = v^j.$$

Since this is true for each j, it follows that v_a is the zero vector.

To prove surjectivity, let $X \in T_a(\mathbb{R}^n)$ be arbitrary. Motivated by the computation in the preceding paragraph, we define real numbers $v^1, \ldots, v^n$ by

$$v^i = X(x^i).$$

We will show that $X = \widetilde{v}_a$, where $v_a = v^i e_i|_a$.

To see this, let f be any smooth function on $\mathbb{R}^n$. By the first-order case of Taylor's formula with remainder (Corollary A.22 in the Appendix), there are smooth functions $g_1, \dots, g_n$ defined on $\mathbb{R}^n$ such that $g_i(a) = 0$ and

$$f(x) = f(a) + \frac{\partial f}{\partial x^i}(a)(x^i - a^i) + g_i(x)(x^i - a^i). \tag{3.3}$$

Note that the last term in (3.3) is a sum of functions, each of which is a product of two functions $g_i(x)$ and $(x^i - a^i)$ that vanish at a. Applying X to this formula and using Lemma 3.1, we obtain

$$\begin{aligned} Xf &= X(f(a)) + X\left(\frac{\partial f}{\partial x^i}(a)(x^i - a^i)\right) + X\big(g_i(x)(x^i - a^i)\big) \\ &= 0 + \frac{\partial f}{\partial x^i}(a)(X(x^i) - X(a^i)) + X(g_i(x)(x^i - a^i)) \\ &= \frac{\partial f}{\partial x^i}(a)v^i \\ &= \widetilde{v}_a f. \end{aligned}$$

This shows that $X = \widetilde{v}_a$. □

Corollary 3.3. *For any $a \in \mathbb{R}^n$, the n derivations*

$$\left.\frac{\partial}{\partial x^1}\right|_a, \dots, \left.\frac{\partial}{\partial x^n}\right|_a,$$

defined by

$$\left.\frac{\partial}{\partial x^i}\right|_a f = \frac{\partial f}{\partial x^i}(a),$$

form a basis for $T_a(\mathbb{R}^n)$.

Proof. This follows immediately from the preceding proposition, once we note that $\partial/\partial x^i|_a = \widetilde{e}_i|_a$. □

Tangent Vectors on a Manifold

Now we are in a position to define tangent vectors on a manifold. Let M be a smooth manifold and let p be a point of M. A linear map $X\colon C^\infty(M) \to \mathbb{R}$ is called a *derivation at p* if it satisfies

$$X(fg) = f(p)Xg + g(p)Xf$$

for all $f, g \in C^\infty(M)$. The set of all derivations of $C^\infty(M)$ at p is a vector space called the *tangent space* to M at p, and is denoted by T_pM. An element of T_pM is called a *tangent vector* at p.

The following lemma is the analogue of Lemma 3.1 for manifolds.

Lemma 3.4 (Properties of Tangent Vectors on Manifolds). *Suppose M is a smooth manifold, $p \in M$, and $X \in T_pM$.*

(*a*) *If f is a constant function, then $Xf = 0$.*

(*b*) *If $f(p) = g(p) = 0$, then $X(fg) = 0$.*

Exercise 3.1. Prove Lemma 3.4.

In the special case $M = \mathbb{R}^n$, Proposition 3.2 shows that $T_a\mathbb{R}^n$ is naturally isomorphic to the geometric tangent space $\mathbb{R}^n_a$, and thus also to $\mathbb{R}^n$ itself.

Push-Forwards

To relate the abstract tangent space we have defined on a manifold to geometric tangent spaces in $\mathbb{R}^n$, we have to explore the way tangent vectors behave under smooth maps. If M and N are smooth manifolds and $F\colon M \to N$ is a smooth map, for each $p \in M$ we define a map $F_*\colon T_pM \to T_{F(p)}N$, called the *push-forward* associated with F, by

$$(F_*X)(f) = X(f \circ F).$$

Note that if $f \in C^\infty(N)$, then $f \circ F \in C^\infty(M)$, so $X(f \circ F)$ makes sense. The operator F_*X is clearly linear, and is a derivation at $F(p)$ because

$$\begin{aligned}
(F_*X)(fg) &= X((fg) \circ F) \\
&= X((f \circ F)(g \circ F)) \\
&= f \circ F(p)X(g \circ F) + g \circ F(p)X(f \circ F) \\
&= f(F(p))(F_*X)(g) + g(F(p))(F_*X)(f).
\end{aligned}$$

Because the notation F_* does not explicitly mention the point p, we will have to be careful to specify it when necessary to avoid confusion.

Lemma 3.5 (Properties of Push-forwards). *Let $F\colon M \to N$ and $G\colon N \to P$ be smooth maps and let $p \in M$.*

(*a*) *$F_*\colon T_pM \to T_{F(p)}N$ is linear.*

(*b*) *$(G \circ F)_* = G_* \circ F_*\colon T_pM \to T_{G\circ F(p)}P$.*

(*c*) *$(\mathrm{Id}_M)_* = \mathrm{Id}_{T_pM}\colon T_pM \to T_pM$.*

(*d*) *If F is a diffeomorphism, then $F_*\colon T_pM \to T_{F(p)}N$ is an isomorphism.*

Exercise 3.2. Prove Lemma 3.5.

Our first important application of the push-forward will be to use coordinate charts to relate the tangent space to a point on a manifold with the Euclidean tangent space. But there is an important technical issue that we must address first: While the tangent space is defined in terms of smooth functions on the whole manifold, coordinate charts are in general defined only on open subsets. The key point is that the tangent space is really a purely local construction. To see why, we use the extension lemma (Lemma 2.20).

The local nature of tangent vectors on a manifold is expressed in the following proposition.

Proposition 3.6. *Suppose M is a smooth manifold, $p \in M$, and $X \in T_pM$. If f and g are functions on M that agree on some neighborhood of p, then $Xf = Xg$.*

Proof. Setting $h = f - g$, by linearity it suffices to show that $Xh = 0$ whenever h vanishes in a neighborhood W of p.

Let A be the closed subset $M \smallsetminus W$. By the extension lemma, there is a globally defined smooth function $u \in C^\infty(M)$ that is equal to the constant function 1 on A and is supported in $M \smallsetminus \{p\}$. Because $u \equiv 1$ where h is nonzero, the product hu is identically equal to h. Since $h(p) = u(p) = 0$, Lemma 3.4 implies that $Xh = X(hu) = 0$. □

Using this proposition, we can show that the tangent space to an open submanifold can be naturally identified with the tangent space to the whole manifold.

Proposition 3.7. *Let M be a smooth manifold, let $U \subset M$ be an open submanifold, and let $\iota\colon U \hookrightarrow M$ be the inclusion map. For any $p \in U$, $\iota_*\colon T_pU \to T_pM$ is an isomorphism.*

Proof. Let B be a small neighborhood of p such that $\overline{B} \subset U$. First suppose $X \in T_pU$ and $\iota_* X = 0 \in T_pM$. If $f \in C^\infty(U)$ is arbitrary, the extension lemma guarantees that there is a smooth function $\widetilde{f} \in C^\infty(M)$ such that $\widetilde{f} \equiv f$ on $\overline{B}$. Then by Proposition 3.6,

$$Xf = X(\widetilde{f}|_U) = X(\widetilde{f} \circ \iota) = (\iota_* X)\widetilde{f} = 0.$$

Since this holds for every $f \in C^\infty(U)$, it follows that $X = 0$, so ι_* is injective.

On the other hand, suppose $Y \in T_pM$ is arbitrary. Define an operator $X\colon C^\infty(U) \to \mathbb{R}$ by setting $Xf = Y\widetilde{f}$, where $\widetilde{f}$ is any function on all of M that agrees with f on $\overline{B}$. By Proposition 3.6, Xf is independent of the choice of $\widetilde{f}$, so X is well-defined. Then for any $g \in C^\infty(M)$,

$$(\iota_* X)g = X(g \circ \iota) = Y(\widetilde{g \circ \iota}) = Yg,$$

where the last equality follows from the fact that $\widetilde{g \circ \iota}$ agrees with g on B. Therefore, ι_* is also surjective. $\square$

If U is an open set in a smooth manifold M, the isomorphism ι_* between T_pU and T_pM is canonically defined, independently of any choices. From now on we will *identify* T_pU with T_pM for any point $p \in U$. This identification just amounts to the observation that $\iota_* X$ is the the *same derivation* as X, thought of as acting on functions on the bigger manifold M instead of functions on U. Since the action of a derivation on a function depends only on the values of the function in an arbitrarily small neighborhood, this is a harmless identification.

Exercise 3.3. If $F\colon M \to N$ is a local diffeomorphism, show that $F_*\colon T_pM \to T_{F(p)}N$ is an isomorphism for every $p \in M$.

Recall from Chapter 1 that every finite-dimensional vector space has a natural smooth manifold structure that is independent of any choice of basis or norm. The following proposition shows that the tangent space to a vector space can be naturally identified with the vector space itself. Compare this with the isomorphism between $T_a\mathbb{R}^n$ and $\mathbb{R}^n_a$ that we proved in the preceding section.

Proposition 3.8 (The Tangent Space to a Vector Space). *For each finite-dimensional vector space V and each point $a \in V$, there is a natural (basis-independent) isomorphism $V \to T_aV$ such for any linear map $L\colon V \to W$ the following diagram commutes:*

$$\begin{tikzcd} V \arrow[r, "\cong"] \arrow[d, "L"'] & T_aV \arrow[d, "L_*"] \\ W \arrow[r, "\cong"'] & T_{La}W \end{tikzcd} \tag{3.4}$$

Proof. As we did in the case of $\mathbb{R}^n$, for any vector $v \in V$, we define a derivation $\widetilde{v}_a$ at a by

$$\widetilde{v}_a f = \frac{d}{dt} f(a + tv).$$

Clearly this is independent of any choice of basis. The same arguments we used in the case of $\mathbb{R}^n$ show that $\widetilde{v}_a$ is indeed a derivation, and that the map $v \mapsto \widetilde{v}_a$ is an isomorphism.

Now suppose $L\colon V \to W$ is a linear map. Unwinding the definitions and using the linearity of L, we compute

$$\begin{aligned}(L_*\widetilde{v}_a)f &= \widetilde{v}_a(f \circ L)\\ &= \frac{d}{dt}\bigg|_{t=0} f(L(a+tv))\\ &= \frac{d}{dt}\bigg|_{t=0} f(La+tLv)\\ &= \widetilde{Lv}_{La}f,\end{aligned}$$

which is 3.4. □

Using this proposition, we will routinely identify tangent vectors to a finite-dimensional vector space with elements of the space itself whenever convenient.

Computations in Coordinates

Our treatment of the tangent space to a manifold so far might seem hopelessly abstract. To bring it down to earth, we will show how to do computations with tangent vectors and push-forwards in local coordinates.

Let (U, φ) be a smooth coordinate chart on M. Note that φ is, in particular, a diffeomorphism from U to $\varphi(U)$. Thus, combining the results of Proposition 3.7 and Lemma 3.5(d) above, $\varphi_*\colon T_pM \to T_{\varphi(p)}\mathbb{R}^n$ is an isomorphism.

By Corollary 3.3, $T_{\varphi(p)}\mathbb{R}^n$ has a basis consisting of the derivations $\partial/\partial x^i|_{\varphi(p)}$, $i = 1, \ldots, n$. Therefore, the push-forwards of these vectors under $(\varphi^{-1})_*$ form a basis for T_pM. In keeping with our standard practice of treating coordinate maps as identifications, we will use the following notation for these push-forwards:

$$\frac{\partial}{\partial x^i}\bigg|_p = (\varphi^{-1})_* \frac{\partial}{\partial x^i}\bigg|_{\varphi(p)}.$$

Unwinding the definitions, we see that $\partial/\partial x^i|_p$ acts on a smooth function $f\colon U \to \mathbb{R}$ by

$$\begin{aligned}\frac{\partial}{\partial x^i}\bigg|_p f &= \frac{\partial}{\partial x^i}\bigg|_{\varphi(p)} (f \circ \varphi^{-1})\\ &= \frac{\partial \widehat{f}}{\partial x^i}(\widehat{p}),\end{aligned}$$

where $\widehat{f}$ is the coordinate representation of f and $\widehat{p} = (p^1, \ldots, p^n) = \varphi(p)$ is the coordinate representation of p. In other words, $\partial/\partial x^i|_p$ is just the derivation that takes the ith partial derivative of (the coordinate representation of) f at (the coordinate representation of) p. These are called the *coordinate vectors* at p associated with the given coordinate system. In the special case in which $M = \mathbb{R}^n$, the coordinate vectors $\partial/\partial x^i|_a$ correspond to the standard basis vectors $e_i|_a$ under the isomorphism $T_a\mathbb{R}^n \leftrightarrow \mathbb{R}^n_a$.

Since the coordinate vectors form a basis for T_pM, any tangent vector $X \in T_pM$ can be written uniquely as a linear combination

$$X = X^i \left.\frac{\partial}{\partial x^i}\right|_p$$

where we are using the summation convention as usual. The numbers $(X^1, \dots, X^n)$ are called the *components* of X with respect to the given coordinate system.

Next we explore how push-forwards look in coordinates. We begin by considering the special case of a smooth map $F\colon U \to V$, where $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open subsets of Euclidean spaces. For any $p \in \mathbb{R}^n$, we will determine the matrix of $F_*\colon T_p\mathbb{R}^n \to T_{F(p)}\mathbb{R}^m$ in terms of the standard coordinate bases. Using $(x^1, \dots, x^n)$ to denote the coordinates in the domain and $(y^1, \dots, y^m)$ to denote those in the range, we use the chain rule to compute the action of F_* on a typical basis vector as follows:

$$\begin{aligned}
\left(F_* \left.\frac{\partial}{\partial x^i}\right|_p\right) f &= \left.\frac{\partial}{\partial x^i}\right|_p (f \circ F) \\
&= \frac{\partial (f \circ F)}{\partial x^i}(p) \\
&= \frac{\partial f}{\partial y^j}(F(p)) \frac{\partial F^j}{\partial x^i}(p) \\
&= \left(\frac{\partial F^j}{\partial x^i}(p) \left.\frac{\partial}{\partial y^j}\right|_{F(p)}\right) f.
\end{aligned}$$

Thus

$$F_* \left.\frac{\partial}{\partial x^i}\right|_p = \frac{\partial F^j}{\partial x^i}(p) \left.\frac{\partial}{\partial y^j}\right|_{F(p)}. \tag{3.5}$$

In other words, the matrix of F_* in terms of the standard coordinate bases is

$$\begin{pmatrix}
\dfrac{\partial F^1}{\partial x^1}(p) & \cdots & \dfrac{\partial F^1}{\partial x^n}(p) \\
\vdots & \ddots & \vdots \\
\dfrac{\partial F^m}{\partial x^1}(p) & \cdots & \dfrac{\partial F^m}{\partial x^n}(p)
\end{pmatrix}.$$

(Recall that the columns of the matrix of a linear map are the components of the images of the basis vectors.) This matrix is none other than the *Jacobian matrix* of F, which is the matrix representation of the total derivative $DF(a)\colon \mathbb{R}^n \to \mathbb{R}^m$. Therefore, in this special case, $F_*\colon T_p\mathbb{R}^n \to T_{F(p)}\mathbb{R}^m$ corresponds to the total derivative $DF(p)\colon \mathbb{R}^n \to \mathbb{R}^m$, under our usual identification of Euclidean space with its tangent space.

To remember the arrangement of rows and columns in the Jacobian matrix, it is useful to observe that the jth row of this matrix consists of the partial derivatives of the jth component function F^j of F.

Now consider the more general case of a smooth map $F\colon M \to N$ between smooth manifolds. Choosing coordinate charts (U,φ) for M near p and (V,ψ) for N near $F(p)$, we obtain the coordinate representation $\widehat{F} = \psi\circ F\circ \varphi^{-1}\colon \varphi(U\cap F^{-1}(V)) \to \psi(V)$. By the computation above, $\widehat{F}_*$ is represented with respect to the standard coordinate bases by the Jacobian matrix of $\widehat{F}$. Making our usual identifications $U \leftrightarrow \varphi(U)$, $V \leftrightarrow \psi(V)$, and $F \leftrightarrow \widehat{F}$, we see that F_* itself is represented in terms of this coordinate basis by the Jacobian matrix of (the coordinate representative of) F. In fact, the definition of the push-forward was cooked up precisely to give a coordinate-independent meaning to the total derivative of a smooth map.

Because of this, in the differential geometry literature the push-forward of a smooth map $F\colon M \to N$ is sometimes called its differential, its total derivative, or just its derivative, and can also be denoted by such symbols as

$$F'(p), \quad dF, \quad DF, \quad dF|_p, \quad DF(p), \quad \text{etc.}$$

We will stick with the notation F_* for the push-forward of a map between manifolds, and reserve $DF(p)$ for the total derivative of a function between finite-dimensional vector spaces, which in the case of Euclidean spaces we identify with the Jacobian matrix of F.

Change of Coordinates

Suppose (U,φ) and (V,ψ) are two smooth charts on M, and $p \in U\cap V$. Let us denote the coordinate functions of φ by (x^i) and those of ψ by $(\widetilde{x}^i)$. Any tangent vector at p can be represented with respect to either basis $(\partial/\partial x^i|_p)$ or $(\partial/\partial \widetilde{x}^i|_p)$. How are the two representations related?

If we write the transition map $\psi\circ\varphi^{-1}\colon \varphi(U\cap V) \to \psi(U\cap V)$ in the shorthand notation

$$\psi\circ\varphi^{-1}(x) = (\widetilde{x}^1(x), \dots, \widetilde{x}^n(x)),$$

then by (3.5) the push-forward by $\psi\circ\varphi^{-1}$ can be written

$$(\psi\circ\varphi^{-1})_* \left.\frac{\partial}{\partial x^i}\right|_{\varphi(p)} = \frac{\partial \widetilde{x}^j}{\partial x^i}(\varphi(p)) \left.\frac{\partial}{\partial \widetilde{x}^j}\right|_{\psi(p)}.$$

Using the definition of the coordinate vectors at p, we find

$$\begin{aligned}
\left.\frac{\partial}{\partial x^i}\right|_p &= (\varphi^{-1})_* \left.\frac{\partial}{\partial x^i}\right|_{\varphi(p)} \\
&= (\psi^{-1})_*(\psi\circ\varphi^{-1})_* \left.\frac{\partial}{\partial x^i}\right|_{\varphi(p)} \\
&= (\psi^{-1})_* \frac{\partial \widetilde{x}^j}{\partial x^i}(\varphi(p)) \left.\frac{\partial}{\partial \widetilde{x}^j}\right|_{\psi(p)} \\
&= \frac{\partial \widetilde{x}^j}{\partial x^i}(\varphi(p))(\psi^{-1})_* \left.\frac{\partial}{\partial \widetilde{x}^j}\right|_{\psi(p)} \\
&= \frac{\partial \widetilde{x}^j}{\partial x^i}(\widehat{p}) \left.\frac{\partial}{\partial \widetilde{x}^j}\right|_p,
\end{aligned}$$

where $\widehat{p} = \varphi(p)$ is the representation of p in x^i-coordinates. Applying this to the components of a vector $X = X^i \partial/\partial x^i|_p = \widetilde{X}^j \partial/\partial \widetilde{x}^j|_p$, we find that the components of X transform by the rule

$$\widetilde{X}^j = \frac{\partial \widetilde{x}^j}{\partial x^i}(\widehat{p}) X^i. \tag{3.6}$$

In the early days of differential geometry, a tangent vector at a point $p \in M$ was defined as a rule that assigns to each coordinate chart near p an ordered n-tuple $(X^1, \dots, X^n)$, such that the n-tuples assigned to overlapping charts transform according to the rule (3.6). (In fact, many physicists are still apt to define it this way.) Our modern definition, however abstract, has several advantages over this old-fashioned one: For example, we have defined a tangent vector as an actual object (a derivation) and not as an equivalence class; even more importantly, our definition is manifestly coordinate independent, and as we will see makes it easy to define a variety of operations on tangent vectors in a coordinate-independent way.

The Tangent Space to a Manifold With Boundary

Suppose M is an n-dimensional manifold with boundary, and p is a boundary point of M. There are a number of ways one might choose to define the tangent space to M at p. Should it be an n-dimensional vector space, like the tangent space at an interior point? Or should it be $(n-1)$-dimensional, like the boundary? Or should it be an n-dimensional half-space, like the space $\mathbb{H}^n$ on which M is modeled locally? The standard choice is to define T_pM to be an n-dimensional vector space. This may or may not seem like the most geometrically intuitive choice, but it has the advantage of making most of the definitions of geometric objects on a manifold with boundary look exactly the same as those on a manifold.

Thus if M is a manifold with boundary and $p \in M$ is arbitrary, we define the tangent space to M at p in the same way as we defined it for a manifold: T_pM is the space of derivations of $C^\infty(M)$ at p. Similarly, if $F\colon M \to N$ is a smooth map between manifolds with boundary, we define the *push-forward* by F at $p \in M$ to be the linear map $F_*\colon T_pM \to T_{F(p)}N$ defined by the same formula as in the manifold case:

$$(F_*X)f = X(f \circ F).$$

The important fact about these definitions is expressed in the following lemma.

Lemma 3.9. *If M is an n-dimensional manifold with boundary and p is a boundary point of M, then T_pM is n-dimensional, with basis given by the coordinate vectors $(\partial/\partial x^1|_p, \dots, \partial/\partial x^n|_p)$ in any smooth chart.*

Proof. For any coordinate map φ, the push-forward $\varphi_*\colon T_pM \to T_{\varphi(p)}\mathbb{H}^n$ is an isomorphism by the same argument as in the manifold case; thus it suffices to show that for any $a \in \partial\mathbb{H}^n$, $T_a\mathbb{H}^n$ is n-dimensional and spanned by the standard coordinate vectors.

Consider the inclusion map $\iota\colon \mathbb{H}^n \hookrightarrow \mathbb{R}^n$. We will show that $\iota_*\colon T_a\mathbb{H}^n \to T_a\mathbb{R}^n$ is an isomorphism. Suppose $\iota_*X = 0$. Let f be any smooth function defined on a neighborhood of a in $\mathbb{H}^n$, and let $\widetilde{f}$ be any extension of f to a smooth function on an open subset of $\mathbb{R}^n$. (Such an extension exists by the definition of what it means for a function to be smooth on $\mathbb{H}^n$.) Then $\widetilde{f} \circ \iota = f$, so

$$Xf = X(\widetilde{f} \circ \iota) = (\iota_*X)\widetilde{f} = 0,$$

which implies that ι_* is injective. On the other hand, if $Y \in \mathbb{T}_a\mathbb{R}^n$ is arbitrary, define $X \in T_a\mathbb{H}^n$ by

$$Xf = Y\widetilde{f},$$

where $\widetilde{f}$ is any extension of f. Writing $Y = Y^i\partial/\partial x^i|_a$ in terms of the standard basis, this means

$$Xf = Y^i \frac{\partial \widetilde{f}}{\partial x^i}(a).$$

This is well defined because by continuity the derivatives of $\widetilde{f}$ at a are determined by those of f in $\mathbb{H}^n$. It is easy to check that X is a derivation and that $Y = \iota_*X$, so ι_* is surjective. □

Tangent Vectors to Curves

The notion of the tangent vector to a curve in $\mathbb{R}^n$ is familiar from elementary calculus—it is just the vector whose components are the derivatives of the component functions of the curve. In this section, we extend this notion to curves in manifolds.

If M is a smooth manifold, a *smooth curve* in M is a smooth map $\gamma\colon J \to M$, where $J \subset \mathbb{R}$ is an interval. (For the most part, we will be interested in curves whose domains are open intervals, but for some purposes it is useful to allow J to have one or two endpoints; the definitions all make sense in that case if we consider J as a manifold with boundary.) Our definitions lead to a very natural interpretation of tangent vectors to curves in manifolds.

If γ is a smooth curve in M, we define the *tangent vector to* γ at $t_0 \in J$ to be the vector

$$\gamma'(t_0) = \gamma_* \left.\frac{d}{dt}\right|_{t_0} \in T_{\gamma(t_0)}M,$$

where $d/dt|_{t_0}$ is the standard coordinate basis for $T_{t_0}\mathbb{R}$. (As in ordinary calculus, is customary to use d/dt instead of $\partial/\partial t$ when the domain is 1-dimensional.) Other common notations for the tangent vector to γ are

$$\dot{\gamma}(t_0), \quad \frac{d\gamma}{dt}(t_0), \quad \text{and} \quad \left.\frac{d\gamma}{dt}\right|_{t=t_0}.$$

This tangent vector acts on functions by

$$\begin{aligned}\gamma'(t_0)f &= \gamma_* \left.\frac{d}{dt}\right|_{t_0} f\\ &= \left.\frac{d}{dt}\right|_{t_0} (f \circ \gamma)\\ &= \frac{d(f\circ\gamma)}{dt}(t_0).\end{aligned}$$

(If t_0 is an endpoint of J, this still holds provided we interpret the derivative with respect to t as a one-sided derivative.) In other words, $\gamma'(t_0)$ is the derivation obtained by taking the derivative of a function along γ.

Now let (U,φ) be a smooth chart with coordinate functions (x^i). If $\gamma(t_0) \in U$, we can write the coordinate representation of γ as $\gamma(t) = (\gamma^1(t),\dots,\gamma^n(t))$, at least for t sufficiently near t_0, and then the formula for the push-forward in coordinates tells us that

$$\gamma'(t_0) = (\gamma^i)'(t_0) \left.\frac{\partial}{\partial x^i}\right|_{\gamma(t_0)}.$$

This means that $\gamma'(t_0)$ is given by essentially the same formula as it would be in Euclidean space: It is the tangent vector whose components in a coordinate basis are the derivatives of the component functions of γ.

The next lemma shows that every tangent vector on a manifold is the tangent vector to some curve. This gives an alternative, and somewhat more geometric, way of thinking about the tangent space: it is just the set of tangent vectors to smooth curves in M.

Lemma 3.10. *Let M be a smooth manifold and $p \in M$. Every $X \in T_pM$ is the tangent vector to some smooth curve in M.*

Proof. Let (U, φ) be a smooth coordinate chart centered at p, and write $X = X^i \partial/\partial x^i|_p$ in terms of the coordinate basis. Define a curve $\gamma\colon (-\varepsilon, \varepsilon) \to U$ by setting $\gamma(t) = (tX^1, \ldots, tX^n)$ in these coordinates. (Remember, this really means $\gamma(t) = \varphi^{-1}(tX^1, \ldots, tX^n)$.) Clearly this is a smooth curve with $\gamma(0) = p$, and by the computation above $\gamma'(0) = X^i \partial/\partial x^i|_{\gamma(0)} = X$. $\square$

The next proposition shows that tangent vectors to curves behave well under composition with smooth maps.

Proposition 3.11. *Let $F\colon M \to N$ be a smooth map, and let $\gamma\colon J \to M$ be a smooth curve. For any $t_0 \in J$, the tangent vector to the composite curve $F \circ \gamma$ at $t = t_0$ is given by*

$$(F \circ \gamma)'(t_0) = F_* \gamma'(t_0).$$

Proof. Just go back to the definition of the tangent vector to a curve:

$$\begin{aligned}
(F \circ \gamma)'(t_0) &= (F \circ \gamma)_* \left.\frac{d}{dt}\right|_{t_0} \\
&= F_* \gamma_* \left.\frac{d}{dt}\right|_{t_0} \\
&= F_* \gamma'(t_0).
\end{aligned}$$

$\square$

On the face of it, the preceding proposition tells us how to compute the tangent vector to a composite curve in terms of the push-forward map. However, it is often much more useful to turn it around the other way, and use it as a streamlined way to compute push-forwards. Suppose $F\colon M \to N$ is a smooth map, and we need to compute the push-forward map F_* at some point $p \in M$. We can compute $F_* X$ for any $X \in T_pM$ by choosing a smooth curve γ whose tangent vector at $t = 0$ is X, and then

$$F_* X = (F \circ \gamma)'(0).$$

This frequently yields a much more succinct computation of F_*, especially if F is presented in some form other than by giving its coordinate functions. We will see many examples of this technique in later chapters.

Alternative Definitions of the Tangent Space

There are several variations on the definition of the tangent space that you will find in the literature. The most common alternative definition is based on the notion of "germs" of smooth functions, which we now define.

A *smooth function element* on a smooth manifold M is an ordered pair (f, U), where U is an open subset of M and $f\colon U \to \mathbb{R}$ is a smooth function. Given a point $p \in M$, the relation $(f, U) \sim (g, V)$ if $f \equiv g$ on some neighborhood of p is an equivalence relation on the set of all smooth function elements whose domains contain p. The equivalence class of a function element (f, U) is called the *germ of f at p*. The set of all germs of smooth functions at p is denoted by $\mathcal{C}_p^\infty$. It is a real vector space under the operations

$$
\begin{aligned}
[(f,U)] + [(g,V)] &= [(f+g, U \cap V)],\\
c[(f,U)] &= [(cf, U)].
\end{aligned}
$$

(The zero element of this vector space is the equivalence class of the zero function on any neighborhood of a.) One ordinarily denotes the germ of the function element (f, U) simply by $[f]$; there is no need to include the domain U in the notation, because the same germ is represented by the restriction of f to any neighborhood of p. To say that two germs $[f]$ and $[g]$ are equal is simply to say that $f \equiv g$ on some neighborhood of p, however small.

It is common to define T_pM as the set of derivations of $\mathcal{C}_p^\infty$, that is, the set of all linear maps $X\colon \mathcal{C}_p^\infty \to \mathbb{R}$ satisfying a product rule analogous to (3.2). Thanks to Proposition 3.6, it is a simple matter to prove that this space is naturally isomorphic to the tangent space as we have defined it (see Problem 3-7). The germ definition has a number of advantages. One of the most significant is that it makes the local nature of the tangent space clearer, without requiring the use of bump functions. Because there do not exist analytic bump functions, the germ definition of tangent vectors is the only one available on real-analytic or complex-analytic manifolds. The chief disadvantage of the germ approach is simply that it adds an additional level of complication to an already highly abstract definition.

Another common approach to defining T_pM is to define some intrinsic equivalence relation on the set of smooth curves in M starting at p, which amounts to "having the same tangent vector," and defining a tangent vector as an equivalence class of curves. For example, one way to define such an equivalence relation is to declare two curves γ, σ to be equivalent if $(\gamma^i)'(0) = (\sigma^i)'(0)$ in every coordinate chart around p. The computations we did above show that the set of equivalence classes is in one-to-one correspondence with T_pM. This definition has the advantage of being geometrically more intuitive, but it has the serious drawback that the existence of a vector space structure on T_pM is not at all obvious.

Yet another approach to defining the tangent space is based on the classical notion of a tangent vector as an n-tuple that transforms according to a certain rule. More precisely, one defines a tangent vector X at p to be a rule that assigns a vector $(X^1, \dots, X^n) \in \mathbb{R}^n$ to each coordinate chart whose domain contains p, with the property that the vectors related to different charts are related by (3.6).

It is a matter of individual taste which of the various characterizations of T_pM one chooses to take as the definition. We have chosen the space of derivations of $C^\infty(M)$ because it is relatively concrete (tangent vectors are actual derivations of $C^\infty(M)$, with no equivalence classes involved), it makes the vector space structure obvious, and it leads to straightforward definitions of many of the other objects we will be studying.

The Tangent Bundle

For any smooth manifold M, we define the *tangent bundle* of M, denoted by TM, to be the disjoint union of the tangent spaces at all points of M:

$$TM = \coprod_{p\in M} T_pM.$$

We consider an element of this disjoint union to be an ordered pair (p, X), where $p \in M$ and $X \in T_pM$. We will often commit the usual mild sin of identifying T_pM with its image under the canonical injection $X \mapsto (p, X)$, and depending on context will use any of the notations (p, X), X_p, or X for a tangent vector in T_pM, depending on how much emphasis we wish to give to the point p. Define the *projection map* $\pi\colon TM \to M$ by declaring $\pi(p, X) = p$.

The tangent bundle can be thought of simply as a collection of vector spaces; but it is much more than that. The next lemma shows that TM has a natural structure as a smooth manifold in its own right.

Lemma 3.12. *For any smooth n-manifold M, the tangent bundle TM has a natural topology and smooth structure that make it into a smooth $2n$-dimensional manifold such that $\pi\colon TM \to M$ is a smooth map.*

Proof. We begin by defining the maps that will become our smooth charts. Given any chart (U, φ) for M, write the component functions of φ as $\varphi(p) = (x^1(p), \ldots, x^n(p))$ and set $\widetilde{U} = \varphi(U) \subset \mathbb{R}^n$. Define a map $\Phi\colon \pi^{-1}(U) \to \mathbb{R}^{2n}$ by

$$\Phi\left(v^i \left.\frac{\partial}{\partial x^i}\right|_p\right) = (x^1(p), \ldots, x^n(p), v^1, \ldots, v^n).$$

Its image set is $\widetilde{U} \times \mathbb{R}^n$, which is an open subset of $\mathbb{R}^{2n}$. It is a bijection onto its image, because its inverse can be written explicitly as

$$(x, v) \mapsto v^i \left.\frac{\partial}{\partial x^i}\right|_{\varphi^{-1}(x)}.$$

Now suppose we are given two charts (U,φ) and (V,ψ) for M, and let $(\pi^{-1}(U),\Phi)$, $(\pi^{-1}(V),\Psi)$ be the corresponding charts on TM. The sets $\Phi(\pi^{-1}(U)\cap\pi^{-1}(V)) = \varphi(U\cap V)\times\mathbb{R}^n$ and $\Psi(\pi^{-1}(U)\cap\pi^{-1}(V)) = \psi(U\cap V)\times\mathbb{R}^n$ are both open in $\mathbb{R}^{2n}$, and the transition map $\Psi\circ\Phi^{-1}\colon \varphi(U\cap V)\times\mathbb{R}^n \to \psi(U\cap V)\times\mathbb{R}^n$ can be written explicitly using (2.2) as

$$\Psi\circ\Phi^{-1}(x^1,\dots,x^n,v^1,\dots,v^n) = \left(\widetilde{x}^1(x),\dots,\widetilde{x}^n(x),\frac{\partial\widetilde{x}^1}{\partial x^j}(x)v^j,\dots,\frac{\partial\widetilde{x}^n}{\partial x^j}(x)v^j\right).$$

This is clearly smooth.

Choosing a countable cover $\{U_i\}$ of M by coordinate domains, we obtain a countable cover of TM by coordinate domains $\{\pi^{-1}(U_i)\}$ satisfying the conditions (i)–(iv) of Lemma 1.14. To check the Hausdorff condition (v), just note that any two points in the same fiber of π lie in one chart; while if (p,X) and (q,Y) lie in different fibers there exist disjoint coordinate domains U_i, U_j for M such that $p\in U_i$ and $q\in U_j$, and then the sets $\pi^{-1}(U_i)$ and $\pi^{-1}(U_j)$ are disjoint coordinate neighborhoods containing (p,X) and (q,Y), respectively.

To check that π is smooth, we just note that its coordinate representation with respect to charts (U,φ) for M and $(\pi^{-1}(U),\Phi)$ for TM is $\pi(x,v) = x$. □

The coordinates (x^i,v^i) defined in this lemma will be called *standard coordinates* for TM.

Exercise 3.4. Suppose $F\colon M\to N$ is a smooth map. By examining the local expression (3.5) for F_* in coordinates, show that $F_*\colon TM\to TN$ is a smooth map.

Because it is a union of vector spaces glued together in a nice way, the tangent bundle has more structure than just that of a smooth manifold. This kind of structure arises frequently, as we will see later in the book, so we introduce the following definition. Let M be a smooth manifold. A (*smooth*) *vector bundle of rank k over M* is a smooth manifold E together with a smooth surjective map $\pi\colon E\to M$ satisfying:

(i) For each $p\in M$, the set $E_p=\pi^{-1}(p)\subset E$ (called the *fiber* of E over p) is endowed with the structure of a real vector space.

(ii) For each $p\in M$, there exists a neighborhood U of p in M and a diffeomorphism $\Phi\colon \pi^{-1}(U)\to U\times\mathbb{R}^k$ such that the following diagram commutes:

$$\begin{tikzcd} \pi^{-1}(U) \arrow[rr, "\Phi"] \arrow[dr, "\pi"'] & & U \times \mathbb{R}^k \arrow[dl, "\pi_1"] \\ & U, & \end{tikzcd}$$

and the restriction of Φ to E_p is a linear isomorphism from E_p to $\{p\} \times \mathbb{R}^k \cong \mathbb{R}^k$. (Here π_1 is the projection on the first factor.)

The manifold E is called the *total space* of the bundle, M is called its *base*, and π is its *projection.* Each map Φ in the above definition is called a *local trivialization* of E over U. If there exists a local trivialization over the entire manifold M (called a *global trivialization* of E), then E is said to be a *trivial bundle.* If $U \subset M$ is any open set, it is easy to verify that the subset $E|_U = \pi^{-1}(U)$ is again a vector bundle with the restriction of π as its projection map, called the *restriction* of E to U.

One particularly simple example of a rank-k vector bundle over any manifold M is the product manifold $E = M \times \mathbb{R}^k$ with $\pi = \pi_1 \colon M \times \mathbb{R}^k \to M$. This bundle is clearly trivial (with the identity map as a global trivialization), but there are bundles that are not trivial, as we will see later.

Lemma 3.13. *With the smooth structure defined in Lemma 3.12, the tangent bundle to an n-manifold is a smooth vector bundle of rank n.*

Proof. It is easy to check that the coordinate charts Φ defined in Lemma 3.12 serve as local trivializations, once we identify each open coordinate domain U with its image $\widetilde{U} = \varphi(U) \subset \mathbb{R}^n$. □

Before proceeding with our study of the tangent bundle, we introduce some important terminology regarding vector bundles.

Let E be a smooth vector bundle over a smooth manifold M, with projection $\pi \colon E \to M$. A *section* of E is a section of the map π, i.e., a continuous map $\sigma \colon M \to E$ such that $\pi \circ \sigma = \mathrm{Id}_M$. If $U \subset M$ is an open subset, a section of the restricted bundle $E|_U$ is called a *local section* of E. A *smooth section* is a (local or global) section that is smooth as a map between manifolds. The *zero section* is the map $\zeta \colon M \to E$ defined by

$$\zeta(p) = 0 \in E_p \text{ for each } p \in M.$$

Just as for functions, the *support* of a section σ is defined to be the closure of the set $\{p \in M : \sigma(p) \neq 0\}$. A section is said to be *compactly supported* if its support is a compact set.

Exercise 3.5. Show that the zero section of any smooth vector bundle is smooth.

Exercise 3.6. Show that $T\mathbb{R}^n$ is isomorphic to the trivial bundle $\mathbb{R}^n \times \mathbb{R}^n$.

If $U \subset M$ is an open set, a *local frame* for E over U is an ordered k-tuple $(\sigma_1, \dots, \sigma_k)$, where each σ_i is a smooth section of E over U, and such that $(\sigma_1(p), \dots, \sigma_k(p))$ is a basis for the fiber E_p for each $p \in U$. It is called a *global frame* if $U = M$.

If $\pi \colon E \to M$ and $\pi' \colon E' \to M$ are two smooth vector bundles over the same smooth manifold M, a *bundle map* from E to E' is a smooth map $F \colon E \to E'$ such that $\pi' \circ F = \pi$, and with the property that for each $p \in M$, the restricted map $F|_{E_p} \colon E_p \to E'_p$ is linear. A bijective bundle map $F \colon E \to E'$ whose inverse is also a bundle map is called a *bundle isomorphism*. If there exists a bundle isomorphism between E and E', the two bundles are said to be *isomorphic*.

Before leaving the subject of vector bundles, it is worth remarking that all of the definitions work exactly the same on manifolds with boundary, if we reinterpret all the charts as generalized charts. For example, if M is a manifold with boundary, then TM is itself a manifold with boundary, and the coordinate charts we constructed in 3.13 become generalized charts for the total space of TM.

Vector Fields

Let M be a smooth manifold. A *vector field* on M is a section of TM. More concretely, a vector field is a continuous map $Y \colon M \to TM$, usually written $p \mapsto Y_p$, with the property that for each $p \in M$, Y_p is an element of T_pM. (We write the value of Y at p as Y_p instead of $Y(p)$ to avoid conflict with the notation $Y(f)$ for the action of a vector on a function.)

You should think of a vector field on M in the same way as you think of vector fields in Euclidean space: as an arrow attached to each point of M, chosen to be tangent to M and to vary continuously from point to point. If (x^i) are any local coordinates on an open set $U \subset M$, the value of Y at any point $p \in U$ can be written

$$Y_p = Y^i(p) \left.\frac{\partial}{\partial x^i}\right|_p$$

for some numbers $Y^1(p), \dots, Y^n(p)$. This defines n functions $Y^i \colon U \to \mathbb{R}$, called the *component functions* of Y with respect to the given chart.

We will be primarily interested in smooth vector fields, the ones that are smooth as maps from M to TM. The next lemma gives two useful alternative ways to check for smoothness.

Lemma 3.14. *Let M be a smooth manifold, and let $Y \colon M \to TM$ be any map (not necessarily continuous) such that $Y_p \in T_pM$ for each $p \in M$. The following are equivalent.*

(*a*) Y *is smooth.*

(*b*) *The component functions of* Y *in any smooth chart are smooth.*

(*c*) *If* f *is any smooth real-valued function on an open set* $U \subset M$*, the function* $Yf : U \to \mathbb{R}$ *defined by* $Yf(p) = Y_p f$ *is smooth.*

Proof. Given coordinates (x^i) on an open set $U \subset M$, let (x^i, v^i) denote the standard coordinates on $\pi^{-1}(U) \subset TM$ constructed in Lemma 3.12. By definition of the standard coordinates, the coordinate representation of $Y : M \to TM$ on U is

$$\widehat{Y}(x) = (x^1, \ldots, x^n, Y^1(x), \ldots, Y^n(x)),$$

where Y^i is the ith component function of Y in x^i-coordinates. It follows immediately that (a) is equivalent to (b).

Now suppose Y is a map for which (b) holds. If f is any function defined in an open set $U \subset M$, and (x^i) are any coordinates on an open set $W \subset U$, the coordinate representation of Yf on W is

$$\begin{aligned} \widehat{Yf}(x) &= \left(Y^i(x) \left. \frac{\partial}{\partial x^i} \right|_x \right) f \\ &= Y^i(x) \frac{\partial f}{\partial x^i}(x). \end{aligned}$$

Since the functions Y^i are smooth on W by hypothesis, it follows that Yf is smooth. Thus (b) implies (c).

Conversely, suppose (c) holds. If (x^i) are any local coordinates on $U \subset M$, we can think of each coordinate x^i as a smooth function on U, and then $Y^i = Yx^i$ is smooth on U by (c), so (b) holds. □

Example 3.15. If (x^i) are any local coordinates on an open set $U \subset M$, the assignment

$$p \mapsto \left. \frac{\partial}{\partial x^i} \right|_p$$

determines a vector field on U, called the ith *coordinate vector field* and denoted by $\partial/\partial x^i$. Because its component functions are constants, it is smooth by Lemma 3.14(b).

The next lemma shows that every tangent vector at a point is the value of a smooth vector field.

Lemma 3.16. *Let* M *be a smooth manifold. If* $p \in M$ *and* $X \in T_pM$*, there is a smooth vector field* $\widetilde{X}$ *on* M *such that* $\widetilde{X}_p = X$*.*

Proof. Let (x^i) be coordinates on a neighborhood U of p, and let $X^i\partial/\partial x^i|_p$ be the coordinate expression for X. If φ is a bump function supported in U and with $\varphi(p) = 1$, the vector field $\widetilde{X}$ defined by

$$\widetilde{X}_q = \begin{cases} \varphi(q)X^i \left.\dfrac{\partial}{\partial x^i}\right|_q & q \in U, \\ 0 & q \notin U, \end{cases}$$

is easily seen to be a smooth vector field whose value at p is equal to X. □

We will use the notation $\mathcal{T}(M)$ to denote the set of all smooth vector fields on M. It is clearly a real vector space under pointwise addition and scalar multiplication. Moreover, vector fields can be multiplied by smooth functions: If $f \in C^\infty(M)$ and $Y \in \mathcal{T}(M)$, we obtain a new vector field fY by

$$(fY)_p = f(p)Y_p.$$

(Many authors use the notation $\mathcal{X}(M)$ instead of $\mathcal{T}(M)$. However, $\mathcal{T}(M)$ is more amenable to generalization—as a rule, we will use the script letter corresponding to the name of a bundle to denote the its space of smooth sections.)

Exercise 3.7. If $f \in C^\infty(M)$ and $Y \in \mathcal{T}(M)$, show that fY is a smooth vector field.

Exercise 3.8. Show that $\mathcal{T}(M)$ is a module over the ring $C^\infty(M)$.

If M is a smooth manifold, we will use the term *local frame for M* to mean a local frame for TM over some open subset $U \subset M$. Similarly, a *global frame for M* is a global frame for TM. We say M is *parallelizable* if it admits a smooth global frame, which is equivalent to TM being a trivial bundle (see Problem 3-5).

A vector field on a manifold with boundary is defined in exactly the same way as on a manifold. All of the results of this section hold equally well in that case.

Push-forwards of Vector Fields

If $F\colon M \to N$ is a smooth map and Y is a smooth vector field on M, then for each point $p \in M$, we obtain a vector $F_*Y_p \in T_{F(p)}N$ by pushing forward Y_p. However, this does not in general define a vector field on N. For example, if F is not surjective, there is no way to decide what vector to assign to a point $q \in N \smallsetminus F(M)$. If F is not injective, then for some points of N there may be several different vectors obtained as push-forwards of Y from different points of M.

If $F: M \to N$ is smooth and $Y \in \mathcal{T}(M)$, suppose there happens to be a vector field $Z \in \mathcal{T}(N)$ with the property that for each $p \in M$, $F_* Y_p = Z_{F(p)}$. In this case, we say the vector fields Y and Z are *F-related*.

Here is a useful criterion for checking that two vector fields are F-related.

Lemma 3.17. *Suppose $F: M \to N$ is a smooth map, $X \in \mathcal{T}(M)$, and $Y \in \mathcal{T}(N)$. Then X and Y are F-related if and only if for every smooth function f defined on an open subset of N,*

$$X(f \circ F) = (Xf) \circ F. \tag{3.7}$$

Proof. For any $p \in M$,

$$\begin{aligned} X(f \circ F)(p) &= X_p(f \circ F) \\ &= (F_* X_p) f, \end{aligned}$$

while

$$\begin{aligned} (Xf) \circ F(p) &= (Xf)(F(p)) \\ &= X_{F(p)} f. \end{aligned}$$

Thus (3.7) is true for all f if and only if $F_* X_p = X_{F(p)}$ for all p, i.e., if and only if X and X are F-related. □

It is important to remember that for a given vector field Y and map F, there may not be *any* vector field on N that is F-related to Y. There is one special case, however, in which there is always such a vector field, as the next proposition shows.

Proposition 3.18. *Suppose $F: M \to N$ is a diffeomorphism. For every smooth vector field $Y \in \mathcal{T}(M)$, there is a unique smooth vector field on N that is F-related to Y.*

Proof. For $Z \in \mathcal{T}(N)$ to be F-related to Y means that $F_* Y_p = Z_{F(p)}$ for every $p \in M$. If F is a diffeomorphism, therefore, we *define* Z by

$$Z_q = F_*(Y_{F^{-1}(q)}).$$

It is clear that Z, so defined, is the unique vector field that is F-related to Y, and it is smooth because it is equal to the composition

$$N \xrightarrow{F^{-1}} M \xrightarrow{Y} TM \xrightarrow{F_*} TN.$$

(See Exercise 3.4.) □

In the situation of the preceding lemma, we will denote the unique vector field that is F-related to Y by $F_* Y$, and call it the *push-forward* of Y by F. Remember, it is only when F is a diffeomorphism that $F_* Y$ is defined.

Problems

3-1. Suppose M and N are smooth manifolds with M connected, and $F\colon M \to N$ is a smooth map such that $F_*\colon T_pM \to T_{F(p)}N$ is the zero map for each $p \in M$. Show that F is a constant map.

3-2. Let $M_1, \dots, M_k$ be smooth manifolds, and let $\pi_j\colon M_1 \times \cdots \times M_k \to M_j$ be the projection onto the jth factor. For any choices of points $p_i \in M_i$, $i = 1, \dots, k$, show that the map

$$\alpha\colon T_{(p_1,\dots,p_k)}(M_1 \times \cdots \times M_k) \to T_{p_1}M_1 \times \cdots \times T_{p_k}M_k$$

defined by

$$\alpha(X) = (\pi_{1*}X, \dots, \pi_{k*}X)$$

is an isomorphism, with inverse

$$\alpha^{-1}(X_1, \dots, X_k) = (j_{1*}X_1, \dots, j_{k*}X_k),$$

where $j_i\colon M_i \to M_1 \times \cdots \times M_k$ is given by $j_i(q) = (p_1, \dots, p_{i-1}, q, p_{i+1}, \dots, p_k)$. [Using this isomorphism, we will routinely identify T_pM, for example, as a subspace of $T_{(p,q)}(M \times N)$.]

3-3. If a nonempty n-manifold is diffeomorphic to an m-manifold, prove that $n = m$.

3-4. Show that there is a smooth vector field on $\mathbb{S}^2$ that vanishes at exactly one point. [Hint: Try using stereographic projection.]

3-5. Let E be a smooth vector bundle over M. Show that E admits a local frame over an open subset $U \subset M$ if and only if it admits a local trivialization over U, and E admits a global frame if and only if it is trivial.

3-6. Show that $\mathbb{S}^1$, $\mathbb{S}^3$, and $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ are all parallelizable. [Hint: Consider the vector fields

$$\begin{aligned}
X_1 &= -x\frac{\partial}{\partial w} + w\frac{\partial}{\partial x} - z\frac{\partial}{\partial y} + y\frac{\partial}{\partial z}, \\
X_2 &= -y\frac{\partial}{\partial w} + z\frac{\partial}{\partial x} + w\frac{\partial}{\partial y} - x\frac{\partial}{\partial z}, \\
X_3 &= -z\frac{\partial}{\partial w} - y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}.
\end{aligned}$$

on $\mathbb{S}^3$.]

3-7. Let M be a smooth manifold and $p \in M$. Show that T_pM is naturally isomorphic to the space of derivations of $\mathcal{C}_p^\infty$ (the space of germs of smooth functions at p).

4
The Cotangent Bundle

In this chapter, we introduce a construction that is not typically seen in elementary calculus: tangent covectors, which are linear functionals on the tangent space at a point $p \in M$. The space of all covectors at p is a vector space called the cotangent space at p; in linear-algebraic terms, it is the dual space to T_pM. The union of all cotangent spaces at all points of M is a vector bundle called the cotangent bundle.

Whereas tangent vectors give us a coordinate-free interpretation of derivatives of curves, it turns out that derivatives of real-valued functions on a manifold are most naturally interpreted as tangent covectors. Thus we will define the differential of a function as a covector field (a smooth section of the cotangent bundle); it is a sort of coordinate-invariant analogue of the classical gradient.

At the end of the chapter, we define line integrals of covector fields. This allows us to generalize the classical notion of line integrals to manifolds. Then we explore the relationships among three closely related types of covector fields: exact (those that are the differentials of functions), conservative (those whose line integrals around closed curves are zero), and closed (those that satisfy a certain differential equation in coordinates).

Covectors

Let V be a finite-dimensional vector space. (As usual, all of our vector spaces are assumed to be real.) We define a *covector* on V to be a real-valued linear functional on V, that is, a linear map $\omega\colon V \to \mathbb{R}$. The space of all covectors on V is itself a real vector space under the obvious operations of pointwise addition and scalar multiplication. It is denoted by V^* and called the *dual space* to V.

The most important fact about V^* is expressed in the following proposition.

Proposition 4.1. *Let V be a finite-dimensional vector space. If $(E_1, \dots, E_n)$ is any basis for V, then the covectors $(\varepsilon^1, \dots, \varepsilon^n)$, defined by*

$$\varepsilon^i(E_j) = \delta^i_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

form a basis for V^*, called the dual basis to (E_i). Therefore $\dim V^* = \dim V$.

Remark. The symbol δ^i_j used in this proposition, meaning 1 if $i = j$ and 0 otherwise, is called the *Kronecker delta*.

Exercise 4.1. Prove Proposition 4.1.

For example, if (e_i) denotes the standard basis for $\mathbb{R}^n$, we denote the dual basis by $(e^1, \dots, e^n)$ (note the upper indices), and call it the *standard dual basis*. These basis covectors are the linear functions from $\mathbb{R}^n$ to $\mathbb{R}$ given by

$$e^j(v) = e^j(v^1, \dots, v^n) = v^j.$$

In other words, e^j is just the linear functional that picks out the jth component of a vector. In matrix notation, a linear map from $\mathbb{R}^n$ to $\mathbb{R}$ is represented by a $1 \times n$ matrix, i.e., a row matrix. The basis covectors can therefore also be thought of as the linear functionals represented by the row matrices

$$e^1 = (1\ 0\ \dots\ 0), \quad \dots \quad , \quad e^n = (0\ \dots\ 0\ 1).$$

In general, if (E_i) is a basis for V and (ε^j) is its dual basis, then Proposition 4.1 shows that we can express an arbitrary covector $\omega \in V^*$ in terms of the dual basis as

$$\omega = \omega_i \varepsilon^i,$$

where the components ω_i are determined by

$$\omega_i = \omega(E_i).$$

Then the action of ω on a vector $X = X^i E_i$ is

$$\omega(X) = \omega_i X^i. \tag{4.1}$$

We will always write basis covectors with lower indices, and components of a covector with upper indices, because this helps to ensure that mathematically meaningful expressions such as (4.1) will always follow our index conventions: Any index that is to be summed over in a given term appears twice, once as a subscript and once as a superscript.

Suppose V and W are vector spaces and $A\colon V \to W$ is a linear map. We define a linear map $A^*\colon W^* \to V^*$, called the *dual map* or *transpose* of A, by

$$(A^*\omega)(X) = \omega(AX), \quad \text{for } \omega \in W^*,\ X \in V.$$

Exercise 4.2. Show that $A^*\omega$ is actually a linear functional on V, and that A^* is a linear map.

Proposition 4.2. *The dual map satisfies the following properties.*

(*a*) $(A \circ B)^* = B^* \circ A^*$.

(*b*) $\mathrm{Id}^*\colon V^* \to V^*$ *is the identity map of* V^*.

Exercise 4.3. Prove the preceding proposition.

(For those who are familiar with the language of category theory, this proposition can be summarized by saying that the assignment that sends a vector space to its dual space and a linear map to its dual map is a contravariant functor from the category of real vector spaces to itself. See, for example, [Lee00, Chapter 7].)

Aside from the fact that the dimension of V^* is the same as that of V, the next most important fact about the dual space is the following.

Proposition 4.3. *Let V be a finite-dimensional vector space. There is a canonical* (*basis-independent*) *isomorphism between V and its second dual space* $V^{**} = (V^*)^*$.

Proof. Given a vector $X \in V$, define a linear functional $\widetilde{X}$ on V^* by

$$\widetilde{X}(\omega) = \omega(X), \quad \text{for } \omega \in V^*.$$

It is easy to check that $\widetilde{X}(\omega)$ depends linearly on ω, so that $\widetilde{X} \in V^{**}$, and that the map $X \mapsto \widetilde{X}$ is linear from V to V^{**}. To show that it is an isomorphism, it suffices for dimensional reasons to verify that it is injective. Suppose $X \in V$ is not zero. Extend X to a basis $(X = E_1, \ldots, E_n)$ for V, and let $(\varepsilon^1, \ldots, \varepsilon^n)$ denote the dual basis for V^*. Then

$$\widetilde{X}(\varepsilon^1) = \varepsilon^1(X) = \varepsilon^1(E_1) = 1 \neq 0,$$

so $\widetilde{X} \neq 0$. □

Because of this proposition, the real number $\omega(X)$ obtained by applying a covector ω to a vector X is sometimes denoted by either of the notations $\langle \omega, X\rangle$ or $\langle X, \omega\rangle$; both expressions can be thought of either as the action of the covector $\omega \in V^*$ on the vector $X \in V$, or as the action of the covector $\widetilde{X} \in V^{**}$ on the element $\omega \in V^*$.

There is also a symmetry between bases and dual bases of a finite-dimensional vector space V: Any basis for V determines a dual basis for V^*, and conversely any basis for V^* determines a dual basis for $V^{**} = V$. It is easy to check that if (ε^i) is the basis for V^* dual to a basis (E_i) for V, then (E_i) is the basis dual to (ε^i).

Tangent Covectors on Manifolds

Now let M be a smooth manifold. For each $p \in M$, we define the *cotangent space* at p, denoted by T_p^*M, to be the dual space to T_pM:

$$T_p^*M = (T_pM)^*.$$

Elements of T_p^*M are called *tangent covectors* at p, or just covectors at p.

If (x^i) are local coordinates on an open subset $U \subset M$, then for each $p \in U$, the coordinate basis $(\partial/\partial x^i|_p)$ gives rise to a dual basis (ε_p^i). Any covector $\xi \in T_p^*M$ can thus be written uniquely as $\xi = \xi_i \varepsilon_p^i$, where

$$\xi_i = \xi\left(\left.\frac{\partial}{\partial x^i}\right|_p\right).$$

Suppose now that $(\widetilde{x}^j)$ are another set of coordinates whose domain overlaps U, and let $(\widetilde{\varepsilon}_p^j)$ denote the basis for T_p^*M dual to $(\partial/\partial \widetilde{x}^j|_p)$. We can compute the components of the same covector ξ with respect to the new coordinate system as follows. First observe that the computations in Chapter 3 show that the coordinate vector fields transform as follows:

$$\left.\frac{\partial}{\partial x^i}\right|_p = \frac{\partial \widetilde{x}^j}{\partial x^i}(p)\left.\frac{\partial}{\partial \widetilde{x}^j}\right|_p. \tag{4.2}$$

Writing ξ in both systems as

$$\xi = \xi_i \varepsilon_p^i = \widetilde{\xi}_j \widetilde{\varepsilon}_p^j,$$

we can use (4.2) to compute the components ξ_i in terms of $\widetilde{\xi}_j$:

$$\xi_i = \xi\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) = \xi\left(\frac{\partial \widetilde{x}^j}{\partial x^i}(p)\left.\frac{\partial}{\partial \widetilde{x}^j}\right|_p\right) = \frac{\partial \widetilde{x}^j}{\partial x^i}(p)\widetilde{\xi}_j. \tag{4.3}$$

As we mentioned in Chapter 3, in the early days of smooth manifold theory, before most of the abstract coordinate-free definitions we are using were developed, mathematicians tended to think of a tangent vector at a point p as an assignment of an n-tuple of real numbers to each coordinate system, with the property that the n-tuples $(X^1, \dots, X^n)$ and $(\widetilde{X}^1, \dots, \widetilde{X}^n)$ assigned to two different coordinate systems (x^i) and $(\widetilde{x}^j)$ were related by the transformation law that we derived in Chapter 3:

$$\widetilde{X}^j = \frac{\partial \widetilde{x}^j}{\partial x^i}(p) X^i.$$

Similarly, a tangent covector was thought of as an n-tuple $(\xi_1, \dots, \xi_n)$ that transforms, by virtue of (4.3), according to the following slightly different rule:

$$\xi_i = \frac{\partial \widetilde{x}^j}{\partial x^i}(p) \widetilde{\xi}_j. \tag{4.4}$$

Since the transformation law (4.2) for the coordinate partial derivatives follows directly from the chain rule, it can be thought of as fundamental. Thus it became customary to call tangent covectors *covariant vectors* because their components transform in the same way as ("vary with") the coordinate partial derivatives, with the Jacobian matrix $(\partial \widetilde{x}^j / \partial x^i)$ multiplying the objects associated with the "new" coordinates $(\widetilde{x}^j)$ to obtain those associated with the "old" coordinates (x^i). Analogously, tangent vectors were called *contravariant vectors*, because their components transform in the opposite way. (Remember, it was the component n-tuples that were thought of as the objects of interest.) Admittedly, it does not make a lot of sense, but by now the terms are well entrenched, and we will see them again in Chapter 8. Note that this use of the terms covariant and contravariant has nothing to do with the covariant and contravariant functors of category theory.

The Cotangent Bundle

The disjoint union

$$T^*M = \coprod_{p \in M} T_p^*M$$

is called the *cotangent bundle* of M.

Proposition 4.4. *The cotangent bundle of a smooth manifold has a natural structure as a vector bundle of rank n over M.*

Proof. The proof is essentially the same as the one we gave for the tangent bundle. Let $\pi \colon T^*M \to M$ be the natural projection that sends $\xi \in T_p^*M$

to $p \in M$. Given a coordinate chart $(U, (x^i))$ on M, we define a chart $\Phi\colon \pi^{-1}(U) \to \mathbb{R}^{2n}$ by

$$\Phi(\xi_i \varepsilon_p^i) = (x^1(p), \dots, x^n(p), \xi_1, \dots, \xi_n),$$

where $(x^1(p), \dots, x^n(p))$ is the coordinate representation of $p \in M$, and (ε_p^i) is the dual coordinate basis for T_p^*M. (In this situation, we must forego our insistence that coordinate functions have upper indices, because the fiber coordinates ξ_i are already required by our index conventions to have lower indices.)

If two charts $(U, (x^i))$ and $(\widetilde{U}, (\widetilde{x}^j))$ overlap, then clearly $\widetilde{x}^j$ are smooth functions of $(x^1, \dots, x^n)$, and (4.3) can be solved for $\widetilde{\xi}_j$ (by inverting the Jacobian matrix) to show that $\widetilde{\xi}_j$ is a smooth function of $(x^1, \dots, x^n, \xi_1, \dots, \xi_n)$. The arguments of Lemmas 3.12 and 3.13 then apply almost verbatim to give T^*M the structure of a smooth vector bundle over M. $\square$

A section of T^*M is called a *covector field* on M. As we did with vector fields, we will write the value of a covector field σ at a point $p \in M$ as σ_p instead of $\sigma(p)$, to avoid conflict with the notation for the action of a covector on a vector. If (ε_p^i) is the dual coordinate basis for T_pM at each point p in some open set $U \subset M$, then σ can be expressed locally as $\sigma_p = \sigma_i(p)\varepsilon_p^i$ for some functions $\sigma_1, \dots, \sigma_n\colon U \to \mathbb{R}$, called the *component functions* of σ.

A covector field is said to be *smooth* if it is smooth as a map from M to T^*M. Smooth covector fields are called (*differential*) *1-forms*. (The reason for the latter terminology will become clear in Chapter 9, when we define differential k-forms for $k > 1$.)

Just as in the case of vector fields, there are several ways to check for smoothness of a covector field. The proof is quite similar to the proof of the analogous fact for vector fields (Lemma 3.14).

Lemma 4.5. *Let M be a smooth manifold, and let $\sigma\colon M \to T^*M$ be a map* (*not assumed to be continuous*) *such that $\sigma_p \in T_p^*M$ for each $p \in M$. The following are equivalent.*

(*a*) *σ is smooth.*

(*b*) *In any coordinate chart, the component functions σ_i of σ are smooth.*

(*c*) *If X is a smooth vector field defined on any open subset $U \subset M$, then the function $\langle \sigma, X \rangle\colon U \to \mathbb{R}$, defined by*

$$\langle \sigma, X \rangle(p) = \langle \sigma_p, X_p \rangle = \sigma_p(X_p)$$

is smooth.

Exercise 4.4. Prove Lemma 4.5.

An ordered n-tuple of smooth covector fields $(\sigma^1, \ldots, \sigma^n)$ defined on some open set $U \subset M$ is called a *local coframe* on U if (σ^i_p) forms a basis for T^*_pM at each point $p \in U$. If $U = M$, it is called a *global coframe*. (A local coframe is just a local frame for T^*M, in the terminology introduced in Chapter 3.) Given a local frame $(E_1, \ldots, E_n)$ for TM over an open set U, there is a uniquely determined smooth local coframe $(\varepsilon^1, \ldots, \varepsilon^n)$ satisfying $\varepsilon^i(E_j) = \delta^i_j$. It is smooth by part (c) of the preceding lemma. This coframe is called the *dual coframe* to the given frame.

Note that Lemma 4.5(b) implies, in particular, that each coordinate covector field ε^i (whose value at p is the ith dual basis element ε^i_p) is smooth, because each of its component functions is either identically 1 or identically 0. Thus $(\varepsilon^1, \ldots, \varepsilon^n)$ is a smooth local coframe on the coordinate domain, called a *coordinate coframe*.

We denote the set of all smooth covector fields on M by $\mathcal{T}^*(M)$. Given smooth covector fields $\sigma, \tau \in \mathcal{T}^*(M)$, any linear combination $a\sigma + b\tau$ with real coefficients is obviously again a smooth covector field, so $\mathcal{T}^*(M)$ is a vector space. Moreover, just like vector fields, covector fields can be multiplied by smooth functions: If $f \in C^\infty(M)$ and $\sigma \in \mathcal{T}^*(M)$, we define a covector field $f\sigma$ by

$$(f\sigma)_p = f(p)\sigma_p. \tag{4.5}$$

A simple verification using either part (b) or part (c) of Lemma 4.5 shows that $f\sigma$ is smooth. Thus $\mathcal{T}^*(M)$ is a module over $C^\infty(M)$.

Geometrically, we think of a vector field on M as a rule that attaches an arrow to each point of M. What kind of geometric picture can we form of a covector field? The key idea is that a nonzero linear functional $\xi \in T^*_pM$ is completely determined by two pieces of data: its kernel, which is a codimension-1 linear subspace of T_pM (a *hyperplane*), and the set of vectors X for which $\xi(X) = 1$, which is an affine hyperplane parallel to the kernel. (Actually, the set where $\xi(X) = 1$ alone suffices, but it is useful to visualize the two parallel hyperplanes.) The value of $\xi(X)$ for any other vector X is then obtained by linear interpolation or extrapolation.

Thus you can visualize a covector field as defining a pair of affine hyperplanes in each tangent space, one through the origin and another parallel to it, and varying smoothly from point to point. At points where the covector field takes on the value zero, one of the hyperplanes goes off to infinity.

The Differential of a Function

In elementary calculus, the gradient of a smooth function f on $\mathbb{R}^n$ is defined as the vector field whose components are the partial derivatives of f. Unfortunately, in this form, the gradient does not make coordinate-invariant sense.

Exercise 4.5. Let $f(x, y) = x$ on $\mathbb{R}^2$, and let X be the vector field

$$X = \operatorname{grad} f = \frac{\partial}{\partial x}.$$

Compute the coordinate expression of X in polar coordinates (on some open set on which they are defined) using (4.2) and show that it is *not* equal to

$$\frac{\partial f}{\partial r}\frac{\partial}{\partial r} + \frac{\partial f}{\partial \theta}\frac{\partial}{\partial \theta}.$$

The most important use of covector fields is to define a coordinate-invariant analogue of the gradient.

Let f be a smooth function on a manifold M. (As usual, all of this discussion applies to functions defined on an open subset $U \subset M$, simply by replacing M by U throughout.) We define a covector field df, called the *differential* of f, by

$$df_p(X_p) = X_p f \quad \text{for } X_p \in T_pM.$$

Lemma 4.6. *The differential of a smooth function is a smooth covector field.*

Proof. First we need to verify that at each point $p \in M$, $df_p(X_p)$ depends linearly on X_p, so that df_p is indeed a covector at p. This is a simple computation: For any $a, b \in \mathbb{R}$ and $X_p, Y_p \in T_pM$,

$$\begin{aligned} &df_p(aX_p + bY_p) \\ &\qquad = (aX_p + bY_p)f = a(X_p f) + b(Y_p f) = a\, df_p(X_p) + b\, df_p(Y_p). \end{aligned}$$

Next we show that df is smooth. Let (x^i) be local coordinates on an open subset $U \subset M$, and let (ε^i) be the corresponding coordinate coframe on U. Writing df in coordinates as $df_p = A_i(p)\varepsilon^i_p$ for some functions $A_i\colon U \to \mathbb{R}$, the definition of df implies

$$A_i(p) = df_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) = \left.\frac{\partial}{\partial x^i}\right|_p f = \frac{\partial f}{\partial x^i}(p).$$

Since this last expression depends smoothly on p, it follows that the component functions A_i of df are smooth, so df is smooth. □

One consequence of the preceding proof is a formula for the coordinate representation of df:

$$df_p = \frac{\partial f}{\partial x^i}(p)\varepsilon^i_p. \tag{4.6}$$

Thus the components of df in any coordinate system are the partial derivatives of f with respect to those coordinates. Because of this, we can think of df as an analogue of the classical gradient, reinterpreted in a way that makes coordinate-invariant sense on a manifold.

If we apply (4.6) to the special case in which f is one of the coordinate functions $x^j : U \to \mathbb{R}$, we find

$$dx^j_p = \frac{\partial x^j}{\partial x^i}(p)\varepsilon^i_p = \delta^j_i \varepsilon^i_p = \varepsilon^j_p.$$

In other words, *the coordinate covector field* ε^j *is none other than* dx^j*!* Therefore, the formula (4.6) for df_p can be rewritten as

$$df_p = \frac{\partial f}{\partial x^i}(p)dx^i_p,$$

or as an equation between covector *fields* instead of covectors:

$$df = \frac{\partial f}{\partial x^i}dx^i. \tag{4.7}$$

In particular, in the one-dimensional case, this reduces to

$$df = \frac{df}{dx}dx.$$

Thus we have recovered the familiar classical expression for the differential of a function f in coordinates. Henceforth, we will abandon the notation ε^i for the coordinate coframe, and use dx^i instead.

Example 4.7. If $f(x,y) = x^2 y \cos x$ on $\mathbb{R}^2$, then df is given by the formula

$$\begin{aligned} df &= \frac{\partial(x^2 y\cos x)}{\partial x}dx + \frac{\partial(x^2 y \cos x)}{\partial y}dy \\ &= (2xy\cos x - x^2 y \sin x)dx + x^2\cos x\, dy. \end{aligned}$$

Proposition 4.8 (Properties of the Differential). *Let M be a smooth manifold, and let $f, g \in C^\infty(M)$.*

(*a*) *For any constants a, b, $d(af + bg) = a\,df + b\,dg$.*

(*b*) *$d(fg) = f\,dg + g\,df$.*

(*c*) *$d(f/g) = (g\,df - f\,dg)/g^2$ on the set where $g \neq 0$.*

(*d*) *If $J \subset \mathbb{R}$ is an interval containing the image of f, and $h\colon J \to \mathbb{R}$ is a smooth function, then $d(h \circ f) = (h' \circ f)\,df$.*

(*e*) *If f is constant, then $df = 0$.*

Exercise 4.6. Prove Proposition 4.8.

One very important property of the differential is the following characterization of smooth functions with vanishing differentials.

Proposition 4.9 (Functions with Vanishing Differentials). *If f is a smooth function on a smooth manifold M, then $df = 0$ if and only if f is constant on each component of M.*

Proof. It suffices to assume that M is connected and show that $df = 0$ if and only if f is constant. One direction is immediate: If f is constant, then $df = 0$ by Proposition 4.8(e). Conversely, suppose $df = 0$, let $p \in M$, and let $\mathcal{C} = \{q \in M : f(q) = f(p)\}$. If q is any point in $\mathcal{C}$, let U be a connected coordinate domain centered at q. From (4.7) we see that $\partial f/\partial x^i \equiv 0$ in U for each i, so by elementary calculus f is constant on U. This shows that $\mathcal{C}$ is open, and since it is closed by continuity it must be all of M. Thus f is everywhere equal to the constant $f(p)$. $\square$

In elementary calculus, one thinks of df as an approximation for the small change in the value of f caused by small changes in the independent variables x^i. In our present context, df has the same meaning, provided we interpret everything appropriately. Suppose that f is defined and smooth on an open subset $U \subset \mathbb{R}^n$, and let p be a point in U. Recall that dx^i_p is the linear functional that picks out the ith component of a tangent vector at p. Writing $\Delta f = f(p+v) - f(p)$ for $v \in \mathbb{R}^n$, Taylor's theorem shows that Δf is well approximated when v is small by

$$\Delta f \approx \frac{\partial f}{\partial x^i}(p) v^i = \frac{\partial f}{\partial x^i}(p) dx^i_p(v) = df_p(v).$$

In other words, df_p is the linear functional that best approximates Δf near p. The great power of the concept of the differential comes from the facts that we can define df invariantly on any manifold, and can do so without resorting to any vague arguments involving infinitesimals.

The next result is an analogue of Proposition 3.11 for the differential.

Proposition 4.10. *Suppose $\gamma\colon J \to M$ is a smooth curve and $f\colon M \to \mathbb{R}$ is a smooth function. Then the derivative of the real-valued function $f \circ \gamma\colon \mathbb{R} \to \mathbb{R}$ is given by*

$$(f \circ \gamma)'(t) = df_{\gamma(t)}(\gamma'(t)). \tag{4.8}$$

Proof. Directly from the definitions, for any $t_0 \in J$,

$$\begin{aligned}
df_{\gamma(t_0)}(\gamma'(t_0)) &= \gamma'(t_0) f && \text{(definition of } df\text{)} \\
&= \left(\gamma_* \left. \frac{d}{dt} \right|_{t_0} \right) f && \text{(definition of } \gamma'(t)\text{)} \\
&= \gamma_* \left. \frac{d}{dt} \right|_{t_0} (f \circ \gamma) && \text{(definition of } \gamma_*\text{)} \\
&= (f \circ \gamma)'(t_0) && \text{(definition of } (f \circ \gamma)'\text{)},
\end{aligned}$$

which was to be proved. □

It is important to observe that for a smooth real-valued function $f\colon M \to \mathbb{R}$, we have now defined two different kinds of derivative of f at a point $p \in M$. In the preceding chapter, we defined the push-forward f_* as a linear map from T_pM to $T_{f(p)}\mathbb{R}$. In this chapter, we defined the differential df_p as a covector at p, which is to say a linear map from T_pM to $\mathbb{R}$. These are really the same object, once we take into account the canonical identification between $\mathbb{R}$ and its tangent space at any point; one easy way to see this is to note that both are represented in coordinates by the row matrix whose components are the partial derivatives of f.

Similarly, if γ is a smooth curve in M, we have two different meanings for the expression $(f \circ \gamma)'(t)$. On the one hand, $f \circ \gamma$ can be interpreted as a smooth curve in $\mathbb{R}$, and thus $(f \circ \gamma)'(t)$ is its tangent vector at the point $f \circ \gamma(t)$, an element of the tangent space $T_{f\circ\gamma(t)}\mathbb{R}$. Proposition 3.11 shows that this tangent vector is equal to $f_*(\gamma'(t))$. On the other hand, $f \circ \gamma$ can also be considered simply as a real-valued function of one real variable, and then $(f\circ\gamma)'(t)$ is just its ordinary derivative. Proposition 4.10 shows that this derivative is equal to the real number $df_{\gamma(t)}(\gamma'(t))$. Which of these interpretations we choose will depend on the purpose we have in mind.

Pullbacks

As we have seen, a smooth map yields a linear map on tangent vectors called the push-forward. Dualizing this leads to a linear map on covectors going in the opposite direction.

Let $F\colon M \to N$ be a smooth map, and let $p \in M$ be arbitrary. The push-forward map

$$F_*\colon T_pM \to T_{F(p)}N$$

yields a dual map

$$(F_*)^*\colon T^*_{F(p)}N \to T^*_pM.$$

To avoid a proliferation of stars, we write this map, called the *pullback* associated with F, as

$$F^*\colon T^*_{F(p)}N \to T^*_pM.$$

Unraveling the definitions, F^* is characterized by

$$(F^*\xi)(X) = \xi(F_*X), \quad \text{for } \xi \in T^*_{F(p)}N,\ X \in T_pM.$$

When we introduced the push-forward map, we made a point of noting that vector *fields* do not push forward to vector fields, except in the special case of a diffeomorphism. The surprising thing about pullbacks is that they always pull smooth covector fields back to smooth covector fields. Given a smooth map $G\colon M \to N$ and a smooth covector field σ on N, define a covector field $G^*\sigma$ on M by

$$(G^*\sigma)_p = G^*(\sigma_{G(p)}). \tag{4.9}$$

Observe that there is no ambiguity here about what point to pull back from, in contrast to the vector field case. We will prove in Proposition 4.12 below that $G^*\sigma$ is smooth. Before doing so, let us examine two important special cases.

Lemma 4.11. *Let $G\colon M \to N$ be a smooth map, and suppose $f \in C^\infty(N)$ and $\sigma \in \mathcal{T}^*(N)$. Then*

$$G^* df = d(f \circ G); \tag{4.10}$$

$$G^*(f\sigma) = (f \circ G) G^*\sigma. \tag{4.11}$$

Proof. To prove (4.10), we let $X_p \in T_pM$ be arbitrary, and compute

$$\begin{aligned}
(G^* df)_p(X_p) &= (G^* df_{G(p)})(X_p) && \text{(by (4.9))}\\
&= df_{G(p)}(G_* X_p) && \text{(by definition of } G^*)\\
&= (G_* X_p) f && \text{(by definition of } df)\\
&= X_p(f \circ G) && \text{(by definition of } G_*)\\
&= d(f \circ G)_p(X_p) && \text{(by definition of } d(f \circ G)).
\end{aligned}$$

Similarly, for (4.11), we compute

$$\begin{aligned}
(G^*(f\sigma))_p &= G^*((f\sigma)_{G(p)}) && \text{(by (4.9))}\\
&= G^*(f(G(p))\sigma_{G(p)}) && \text{(by (4.5))}\\
&= f(G(p)) G^*(\sigma_{G(p)}) && \text{(because } G^* \text{ is linear)}\\
&= f(G(p))(G^*\sigma)_p && \text{(by (4.9))}\\
&= ((f \circ G) G^*\sigma)_p && \text{(by (4.5))},
\end{aligned}$$

which was to be proved. □

Proposition 4.12. *Suppose $G\colon M \to N$ is smooth, and let σ be a smooth covector field on N. Then $G^*\sigma$ is a smooth covector field on M.*

Proof. Let $p \in M$ be arbitrary, and choose local coordinates (x^i) for M near p and (y^j) for N near $G(p)$. Writing σ in coordinates as $\sigma = \sigma_j dy^j$

for smooth functions σ_j defined near $G(p)$ and using Lemma 4.11 twice, we compute

$$\begin{aligned} G^*\sigma &= G^*(\sigma_j dy^j) \\ &= (\sigma_j \circ G) G^* dy^j \\ &= (\sigma_j \circ G) d(y^j \circ G). \end{aligned}$$

Because this expression is smooth, it follows that $G^*\sigma$ is smooth. □

In the course of the preceding proof, we derived the following formula for the pullback of a covector field with respect to coordinates (x^i) on the domain and (y^j) on the range:

$$G^*\sigma = G^*(\sigma_j dy^j) = (\sigma_j \circ G) d(y^j \circ G) = (\sigma_j \circ G) dG^j, \tag{4.12}$$

where G^j is the jth component function of G in these coordinates. This formula makes the computation of pullbacks in coordinates exceedingly simple, as the next example shows.

Example 4.13. Let $G\colon \mathbb{R}^3 \to \mathbb{R}^2$ be the map given by

$$(u, v) = G(x, y, z) = (x^2 y, y \sin z),$$

and let $\sigma \in \mathcal{T}^*(\mathbb{R}^2)$ be the covector field

$$\sigma = u\, dv + v\, du.$$

According to (4.12), the pullback $G^*\sigma$ is given by

$$\begin{aligned} G^*\sigma &= (u \circ G) d(v \circ G) + (v \circ G) d(u \circ G) \\ &= (x^2 y) d(y \sin z) + (y \sin z) d(x^2 y) \\ &= x^2 y (\sin z\, dy + y \cos z\, dz) + y \sin z (2xy\, dx + x^2\, dy) \\ &= 2xy^2 \sin z\, dx + 2x^2 y \sin z\, dy + x^2 y^2 \cos z\, dz. \end{aligned}$$

In other words, to compute $G^*\sigma$, all you need to do is substitute the component functions of G for the coordinate functions of N everywhere they appear in σ!

This also yields an easy way to remember the transformation law for a covector field under a change of coordinates. Again, an example will convey the idea better than a general formula.

Example 4.14. Let (r, θ) be polar coordinates on, say, the upper half-plane $H = \{(x, y) : y > 0\}$. We can think of the change of coordinates $(x, y) = (r \cos\theta, r \sin\theta)$ as the coordinate expression for the identity map of H, but using (r, θ) as coordinates for the domain and (x, y) for the range. Then the the pullback formula (4.12) tells us that we can compute the polar coordinate expression for a covector field simply by substituting $x = r\cos\theta$, $y = r\sin\theta$. For example,

$$\begin{aligned} x\,dx + y\,dy &= \mathrm{Id}^*(x\,dx + y\,dy) \\ &= (r\cos\theta)d(r\cos\theta) + (r\sin\theta)d(r\sin\theta) \\ &= (r\cos\theta)(\cos\theta\,dr - r\sin\theta\,d\theta) + (r\sin\theta)(\sin\theta\,dr + r\cos\theta\,d\theta) \\ &= (r\cos^2\theta + r\sin^2\theta)dr + (-r^2\cos\theta\sin\theta + r^2\sin\theta\cos\theta)d\theta \\ &= r\,dr. \end{aligned}$$

Line Integrals

Another important application of covector fields is to make coordinate-independent sense of the notion of a line integral.

We begin with the simplest case: an interval in the real line. Suppose $[a,b] \subset \mathbb{R}$ is a compact interval, and ω is a smooth covector field on $[a,b]$. (This means that the component function of ω admits a smooth extension to some neighborhood of $[a,b]$.) If we let t denote the standard coordinate on $\mathbb{R}$, ω can be written $\omega_t = f(t)\,dt$ for some smooth function $f\colon [a,b] \to \mathbb{R}$. The similarity between this and the standard notation $\int f(t)\,dt$ for an integral suggests that there might be a connection between covector fields and integrals, and indeed there is. We define the *integral of* ω *over* $[a,b]$ to be

$$\int_{[a,b]} \omega = \int_a^b f(t)\,dt.$$

The next proposition indicates that this is more than just a trick of notation.

Proposition 4.15 (Diffeomorphism Invariance of the Integral). *Let ω be a smooth covector field on the compact interval $[a,b] \subset \mathbb{R}$. If $\varphi\colon [c,d] \to [a,b]$ is an increasing diffeomorphism (meaning that $t < t'$ implies $\varphi(t) < \varphi(t')$), then*

$$\int_{[c,d]} \varphi^*\omega = \int_{[a,b]} \omega.$$

Proof. If we let s denote the standard coordinate on $[c,d]$ and t that on $[a,b]$, then (4.12) shows that the pullback $\varphi^*\omega$ has the coordinate expression $(\varphi^*\omega)_s = f(\varphi(s))\varphi'(s)\,ds$. Inserting this into the definition of the line integral and using the change of variables formula for ordinary integrals, we obtain

$$\int_{[c,d]} \varphi^*\omega = \int_c^d f(\varphi(s))\varphi'(s)\,ds = \int_a^b f(t)\,dt = \int_{[a,b]} \omega,$$

which was to be proved. □

Exercise 4.7. If $\varphi\colon [c,d] \to [a,b]$ is a decreasing diffeomorphism, show that $\int_{[c,d]} \varphi^*\omega = -\int_{[a,b]} \omega$.

Now let M be a smooth manifold. By a *curve segment* in M we mean a continuous curve $\gamma\colon [a,b] \to M$ whose domain is a compact interval. It is a *smooth curve segment* if it is has a smooth extension to an open set containing $[a,b]$. A *piecewise smooth curve segment* is a curve segment $\gamma\colon [a,b] \to M$ with the property that there exists a finite subdivision $a = a_0 < a_1 < \cdots < a_k = b$ of $[a,b]$ such that $\gamma|_{[a_{i-1},a_i]}$ is smooth for each i. Continuity of γ means that $\gamma(t)$ approaches the same value as t approaches any of the points a_i (other than a_0 or a_k) from the left or the right. Smoothness of γ on each subinterval means that γ has one-sided tangent vectors at each such a_i when approaching from the left or the right, but these one-sided tangent vectors need not be equal.

Lemma 4.16. *If M is a connected smooth manifold, any two points of M can be joined by a piecewise smooth curve segment.*

Proof. Let p be an arbitrary point of M, and define a subset $\mathcal{C} \subset M$ by $\mathcal{C} = \{q \in M : \text{there is a piecewise smooth curve in } M \text{ from } p \text{ to } q\}$. Clearly $p \in \mathcal{C}$, so $\mathcal{C}$ is nonempty. To show $\mathcal{C} = M$, we need to show it is open and closed.

Let $q \in \mathcal{C}$ be arbitrary, which means that there is a piecewise smooth curve segment γ going from p to q. Let U be a coordinate ball centered at q. If q' is any point in U, then it is easy to construct a piecewise smooth curve segment from p to q' by first following γ from p to q, and then following a straight-line path in coordinates from q to q'. Thus $U \subset \mathcal{C}$, which shows that $\mathcal{C}$ is open. On the other hand, if $q \in \partial\mathcal{C}$, let U be a coordinate ball around q as above. The fact that q is a boundary point of $\mathcal{C}$ means that there is some point $q' \in \mathcal{C} \cap U$. In this case, we can construct a piecewise smooth curve from p to q by first following one from p to q' and then following a straight-line path in coordinates from q' to q. This shows that $q \in \mathcal{C}$, so $\mathcal{C}$ is also closed. □

If γ is a smooth curve segment in M and ω is a smooth covector field on M, we define the *line integral* of ω over γ to be the real number

$$\int_\gamma \omega = \int_{[a,b]} \gamma^*\omega.$$

Because $\gamma^*\omega$ is a smooth covector field on $[a,b]$, this definition makes sense. More generally, if γ is piecewise smooth, we define

$$\int_\gamma \omega = \sum_{i=1}^{k} \int_{[a_{i-1},a_i]} \gamma^*\omega,$$

where $[a_{i-1}, a_i]$, $i = 1, \ldots, k$, are the intervals on which γ is smooth.

This definition gives a rigorous meaning to classical line integrals such as $\int_\gamma P\,dx + Q\,dy$ in the plane or $\int_\gamma P\,dx + Q\,dy + R\,dz$ in $\mathbb{R}^3$.

Proposition 4.17 (Properties of Line Integrals). *Let M be a smooth manifold. Suppose $\gamma\colon [a,b] \to M$ is a piecewise smooth curve segment and $\omega, \omega_1, \omega_2 \in \mathcal{T}^*(M)$.*

(*a*) *For any $c_1, c_2 \in \mathbb{R}$,*

$$\int_\gamma (c_1\omega_1 + c_2\omega_2) = c_1 \int_\gamma \omega_1 + c_2 \int_\gamma \omega_2.$$

(*b*) *If γ is a constant map, then $\int_\gamma \omega = 0$.*

(*c*) *If $a < c < b$, then*

$$\int_\gamma \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega,$$

where $\gamma_1 = \gamma|_{[a,c]}$ and $\gamma_2 = \gamma|_{[c,b]}$.

(*d*) *The line integral of ω over γ can also be expressed as the ordinary integral*

$$\int_\gamma \omega = \int_a^b \omega_{\gamma(t)}(\gamma'(t))\,dt.$$

Exercise 4.8. Prove Proposition 4.17.

Example 4.18. Let $M = \mathbb{R}^2 \smallsetminus \{0\}$, let ω be the covector field on M given by

$$\omega = \frac{x\,dy - y\,dx}{x^2 + y^2},$$

and let $\gamma\colon [0, 2\pi] \to M$ be the curve segment defined by

$$\gamma(t) = (\cos t, \sin t).$$

Since $\gamma^*\omega$ can be computed by substituting $x = \cos t$ and $y = \sin t$ everywhere in the formula for ω, we find that

$$\int_\gamma \omega = \int_{[0,2\pi]} \frac{\cos t(\cos t\,dt) - \sin t(-\sin t\,dt)}{\sin^2 t + \cos^2 t} = \int_0^{2\pi} dt = 2\pi.$$

One of the most significant features of line integrals is that they are independent of parametrization, in a sense we now define. If $\gamma\colon [a,b] \to M$ and $\widetilde{\gamma}\colon [c,d] \to M$ are smooth curve segments, we say that $\widetilde{\gamma}$ is a *reparametrization* of γ if $\widetilde{\gamma} = \gamma \circ \varphi$ for some diffeomorphism $\varphi\colon [c,d] \to [a,b]$. If φ is an increasing function, we say $\widetilde{\gamma}$ is a *forward reparametrization*, and if φ is decreasing, it is a *backward reparametrization.* (More generally, one can allow φ to be piecewise smooth, but we will have no need for this extra generalization.)

Proposition 4.19 (Parameter Independence of Line Integrals). *Suppose M is a smooth manifold, ω is a smooth covector field on M, and γ is a piecewise smooth curve segment in M. For any reparametrization $\widetilde{\gamma}$ of γ, we have*

$$\int_{\widetilde{\gamma}} \omega = \begin{cases} \displaystyle\int_\gamma \omega & \text{if } \widetilde{\gamma} \text{ is a forward reparametrization,} \\[2ex] \displaystyle -\int_\gamma \omega & \text{if } \widetilde{\gamma} \text{ is a backward reparametrization.} \end{cases}$$

Proof. First assume that $\gamma\colon [a,b] \to M$ is smooth, and suppose $\varphi\colon [c,d] \to [a,b]$ is an increasing diffeomorphism. Then Proposition 4.15 implies

$$\begin{aligned} \int_{\widetilde{\gamma}} \omega &= \int_{[c,d]} (\gamma \circ \varphi)^* \omega \\ &= \int_{[c,d]} \varphi^* \gamma^* \omega \\ &= \int_{[a,b]} \gamma^* \omega \\ &= \int_\gamma \omega. \end{aligned}$$

When φ is decreasing, the analogous result follows from Exercise 4.7. If γ is only piecewise smooth, the result follows simply by applying the preceding argument on each subinterval where γ is smooth. □

Exercise 4.9. Suppose $F\colon M \to N$ is any smooth map, $\omega \in \mathcal{T}^*(N)$, and γ is a piecewise smooth curve segment in M. Show that

$$\int_\gamma F^* \omega = \int_{F \circ \gamma} \omega.$$

There is one special case in which a line integral is trivial to compute: the line integral of a differential.

Theorem 4.20 (Fundamental Theorem for Line Integrals). *Let M be a smooth manifold. Suppose f is a smooth function on M and*

$\gamma\colon [a,b] \to M$ is a piecewise smooth curve segment in M. Then

$$\int_\gamma df = f(\gamma(b)) - f(\gamma(a)).$$

Proof. Suppose first that γ is smooth. By Proposition 4.10 and Proposition 4.17(d),

$$\int_\gamma df = \int_a^b df_{\gamma(t)}(\gamma'(t))\,dt = \int_a^b (f\circ\gamma)'(t)\,dt.$$

By the one-variable version of the fundamental theorem of calculus, this is equal to $f\circ\gamma(b) - f\circ\gamma(a)$.

If γ is merely piecewise smooth, let $a = a_0 < \cdots < a_k = b$ be the endpoints of the subintervals on which γ is smooth. Applying the above argument on each subinterval and summing, we find that

$$\int_\gamma df = \sum_{i=1}^k \big(f(\gamma(a_i)) - f(\gamma(a_{i-1}))\big) = f(\gamma(b)) - f(\gamma(a)),$$

because the contributions from all the interior points cancel. $\square$

Conservative Covector Fields

Theorem 4.20 shows that the line integral of any covector field ω that can be written as the differential of a smooth function can be computed extremely easily once the smooth function is known. For this reason, there is a special term for covector fields with this property. We say a smooth covector field ω on a manifold M is *exact* on M if there is a function $f \in C^\infty(M)$ such that $\omega = df$. In this case, the function f is called a *potential* for ω. The potential is not uniquely determined, but by Lemma 4.9, the difference between any two potentials for ω must be constant on each component of M.

Because exact differentials are so easy to integrate, it is important to develop criteria for deciding whether a covector field is exact. Theorem 4.20 provides an important clue. It shows that the line integral of an exact covector field depends only on the endpoints $p = \gamma(a)$ and $q = \gamma(b)$: Any other curve segment from p to q would give the same value for the line integral. In particular, if γ is a *closed* curve segment, meaning that $\gamma(a) = \gamma(b)$, then the integral of df over γ is zero.

We say a smooth covector field ω is *conservative* if the line integral of ω over any closed piecewise smooth curve segment is zero. This terminology comes from physics, where a force field is called conservative if the change in energy caused by the force acting along any closed path is zero ("energy is conserved"). (In elementary physics, force fields are usually thought of as vector fields rather than covector fields; see Problem 4-5 for the connection.)

The following lemma gives a useful alternative characterization of conservative covector fields.

Lemma 4.21. *A smooth covector field ω is conservative if and only if the line integral of ω depends only on the endpoints of the curve, i.e., $\int_\gamma \omega = \int_{\widetilde{\gamma}} \omega$ whenever γ and $\widetilde{\gamma}$ are piecewise smooth curve segments with the same starting and ending points.*

> **Exercise 4.10.** Prove Lemma 4.21. [Observe that this would be much harder to prove if we defined conservative fields in terms of smooth curves instead of piecewise smooth ones.]

Theorem 4.22. *A smooth covector field is conservative if and only if it is exact.*

Proof. If $\omega \in \mathcal{T}^*(M)$ is exact, Theorem 4.20 shows that it is conservative, so we need only prove the converse. Suppose therefore that ω is conservative, and assume for the moment that M is connected. Because the line integrals of ω are path independent, we can adopt the following notation: For any points $p, q \in M$, we will use the notation $\int_p^q \omega$ to denote the line integral $\int_\gamma \omega$, where γ is any piecewise smooth curve segment from p to q. Observe that Proposition 4.17(c) implies that

$$\int_{p_1}^{p_2} \omega + \int_{p_2}^{p_3} \omega = \int_{p_1}^{p_3} \omega \tag{4.13}$$

for any three points $p_1, p_2, p_3 \in M$.

Now choose any base point $p_0 \in M$, and define a function $f\colon M \to \mathbb{R}$ by

$$f(q) = \int_{p_0}^{q} \omega.$$

We will show that $df = \omega$. To accomplish this, let $q_0 \in M$ be an arbitrary point, let $(U, (x^i))$ be a coordinate chart centered at q_0, and write the coordinate representation of ω as $\omega = \omega_i dx^i$. We will show that

$$\frac{\partial f}{\partial x^j}(q_0) = \omega_j(q_0)$$

for $j = 1, \ldots, n$, which implies that $df_{q_0} = \omega_{q_0}$.

Fix j, and let $\gamma\colon [-\varepsilon, \varepsilon] \to U$ be the smooth curve segment defined in coordinates by $\gamma(t) = (0, \ldots, t, \ldots, 0)$, with t in the jth place, and with ε chosen small enough that $\gamma[-\varepsilon, \varepsilon] \subset U$. Let $p_1 = \gamma(-\varepsilon)$, and define a new function $\widetilde{f}\colon M \to \mathbb{R}$ by $\widetilde{f}(q) = \int_{p_1}^{q} \omega$. Note that (4.13) implies

$$\widetilde{f}(q) - f(q) = \int_{p_1}^{q} \omega - \int_{p_0}^{q} \omega = \int_{p_0}^{p_1} \omega,$$

which does not depend on q. Thus $\widetilde{f}$ and f differ by a constant, so it suffices to show that $\partial\widetilde{f}/\partial x^j(q_0) = \omega_j(q_0)$.

Now $\gamma'(t) = \partial/\partial x^j|_{\gamma(t)}$ by construction, so

$$\omega_{\gamma(t)}(\gamma'(t)) = \omega_{\gamma(t)}\left(\left.\frac{\partial}{\partial x^j}\right|_{\gamma(t)}\right) = \omega_j(\gamma(t)).$$

Since the restriction of γ to $[-\varepsilon, t]$ is a smooth curve from p_1 to $\gamma(t)$, we have

$$\begin{aligned}\widetilde{f}\circ\gamma(t) &= \int_{p_1}^{\gamma(t)} \omega \\ &= \int_{-\varepsilon}^{t} \omega_{\gamma(s)}(\gamma'(s))\, ds \\ &= \int_{-\varepsilon}^{t} \omega_j(\gamma(s))\, ds.\end{aligned}$$

Thus by the fundamental theorem of calculus,

$$\begin{aligned}\frac{\partial\widetilde{f}}{\partial x^j}(q_0) &= \gamma'(0)\widetilde{f} \\ &= \left.\frac{d}{dt}\right|_{t=0} \widetilde{f}\circ\gamma(t) \\ &= \left.\frac{d}{dt}\right|_{t=0} \int_{-\varepsilon}^{t} \omega_j(\gamma(s))\, ds \\ &= \omega_j(\gamma(0)) = \omega_j(q_0).\end{aligned}$$

This completes the proof that $df = \omega$.

Finally, if M is not connected, let $\{M_i\}$ be the components of M. The argument above shows that for each i there is a smooth function $f_i \in C^\infty(M_i)$ such that $df_i = \omega$ on M_i. Letting $f\colon M \to \mathbb{R}$ be the function that is equal to f_i on M_i, we have $df = \omega$, thus completing the proof. □

It would be nice if every smooth covector field were exact, for then the evaluation of any line integral would just be a matter of finding a potential function and evaluating it at the endpoints, a process analogous to evaluating an ordinary integral by finding an indefinite integral or primitive. However, this is too much to hope for.

Example 4.23. The covector field ω of Example 4.18 cannot be exact on $\mathbb{R}^2 \smallsetminus \{0\}$, because it is not conservative: The computation in that example showed that $\int_\gamma \omega = 2\pi \neq 0$, where γ is the unit circle traversed counterclockwise.

Because exactness has such important consequences for the evaluation of line integrals, we would like to have an easy way to check whether a given covector field is exact. Fortunately, there is a very simple necessary condition, which follows from the fact that partial derivatives of smooth functions can be taken in any order.

To see what this condition is, suppose that ω is exact. Let f be any potential function for ω, and let $(U,(x^i))$ be any coordinate chart on M. Because f is smooth, it satisfies the following identity on U:

$$\frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i}. \tag{4.14}$$

Writing $\omega = \omega_i dx^i$ in coordinates, the fact that $\omega = df$ is equivalent to $\omega_i = \partial f/\partial x^i$. Substituting this into (4.14), we find that the component functions of ω satisfy

$$\frac{\partial \omega_j}{\partial x^i} = \frac{\partial \omega_i}{\partial x^j}. \tag{4.15}$$

We say that a smooth covector field ω is *closed* if its components in every coordinate chart satisfy (4.15). The following lemma summarizes the computation above.

Lemma 4.24. *Every exact covector field is closed.*

The significance of this result is that the property of being closed is one that can be easily checked. First we need the following result, which says that it is not necessary to check the closedness condition in *every* coordinate chart, just in a collection of charts that cover the manifold. The proof of this lemma is a tedious computation; later you will be able to give a somewhat more conceptual proof (see Problem 4-4 and also Chapter 9), so you are free to skip the proof of this lemma if you wish.

Lemma 4.25. *A smooth covector field is closed if and only if it satisfies* (4.15) *in some coordinate chart around every point.*

Proof. If ω is closed, then by definition it satisfies (4.15) in *every* coordinate chart. Conversely, suppose (4.15) holds in some chart around every point, and let $(U,(x^i))$ be an arbitrary coordinate chart. For each $p \in U$, the hypothesis guarantees that there are some coordinates $(\widetilde{x}^j)$ defined near p in which the analogue of (4.15) holds. Using formula (4.4) for the transformation of the components of ω together with the chain rule, we find

$$\begin{aligned}
\frac{\partial\omega_i}{\partial x^j} - \frac{\partial\omega_j}{\partial x^i} &= \frac{\partial}{\partial x^j}\left(\frac{\partial\widetilde{x}^k}{\partial x^i}\widetilde{\omega}_k\right) - \frac{\partial}{\partial x^i}\left(\frac{\partial\widetilde{x}^k}{\partial x^j}\widetilde{\omega}_k\right) \\
&= \left(\frac{\partial^2\widetilde{x}^k}{\partial x^j\partial x^i}\widetilde{\omega}_k + \frac{\partial\widetilde{x}^k}{\partial x^i}\frac{\partial\widetilde{\omega}_k}{\partial x^j}\right) - \left(\frac{\partial^2\widetilde{x}^k}{\partial x^i\partial x^j}\widetilde{\omega}_k + \frac{\partial\widetilde{x}^k}{\partial x^j}\frac{\partial\widetilde{\omega}_k}{\partial x^i}\right) \\
&= \frac{\partial^2\widetilde{x}^k}{\partial x^j\partial x^i}\widetilde{\omega}_k + \frac{\partial\widetilde{x}^k}{\partial x^i}\frac{\partial\widetilde{x}^l}{\partial x^j}\frac{\partial\widetilde{\omega}_k}{\partial\widetilde{x}^l} - \frac{\partial^2\widetilde{x}^k}{\partial x^i\partial x^j}\widetilde{\omega}_k - \frac{\partial\widetilde{x}^k}{\partial x^j}\frac{\partial\widetilde{x}^l}{\partial x^i}\frac{\partial\widetilde{\omega}_k}{\partial\widetilde{x}^l} \\
&= \left(\frac{\partial^2\widetilde{x}^k}{\partial x^j\partial x^i} - \frac{\partial^2\widetilde{x}^k}{\partial x^i\partial x^j}\right)\widetilde{\omega}_k + \frac{\partial\widetilde{x}^k}{\partial x^i}\frac{\partial\widetilde{x}^l}{\partial x^j}\left(\frac{\partial\widetilde{\omega}_k}{\partial\widetilde{x}^l} - \frac{\partial\widetilde{\omega}_l}{\partial\widetilde{x}^k}\right) \\
&= 0 + 0,
\end{aligned}$$

where the fourth equation follows from the third by interchanging the roles of k and l in the last term. □

For example, consider the following covector field on $\mathbb{R}^2$:

$$\omega = y\cos xy\,dx + x\cos xy\,dy.$$

It is easy to check that

$$\frac{\partial(y\cos xy)}{\partial y} = \frac{\partial(x\cos xy)}{\partial x} = \cos xy - xy\sin xy,$$

so ω is closed. In fact, you might guess that $\omega = d(\sin xy)$.

The question then naturally arises whether the converse of Lemma 4.24 is true: is every closed covector field exact? The answer is *almost* yes, but there is an important restriction. It turns out that the answer to the question depends in a subtle way on the shape of the domain, as the next example illustrates.

Example 4.26. Look once again at the covector field ω of Example 4.7. A straightforward computation shows that ω is closed; but as we observed above, it is not exact on $\mathbb{R}^2 \smallsetminus \{0\}$. On the other hand, if we restrict the domain to the right half-plane $U = \{(x, y) : x > 0\}$, a computation shows that $\omega = d(\tan^{-1} y/x)$ there. This can be seen more clearly in polar coordinates, where $\omega = d\theta$. The problem, of course, is that there is no smooth (or even continuous) angle function on all of $\mathbb{R}^2 \smallsetminus \{0\}$, which is a consequence of the "hole" in the center.

This last example illustrates a key fact: The question of whether a particular covector field is exact is a global one, depending on the shape of the domain in question. This observation is the starting point for *de Rham cohomology*, which expresses a deep relationship between smooth structures and topology. We will pursue this relationship in more depth in Chapter 11, but for now we can prove the following result. A subset $V \subset \mathbb{R}^n$ is said to be *star-shaped* with respect to a point $c \in V$ if for every $x \in V$, the line segment from c to x is entirely contained in V. For example, a convex subset is star-shaped with respect to each of its points.

Proposition 4.27. *If M is diffeomorphic to a star-shaped open subset of $\mathbb{R}^n$, then every closed covector field on M is exact.*

Proof. It is easy to check that a diffeomorphism pulls back closed covector fields to closed covector fields and exact covector fields to exact ones; thus it suffices to prove the proposition when M actually *is* a star-shaped open subset of $\mathbb{R}^n$. So suppose $M \subset \mathbb{R}^n$ is star-shaped with respect to $c \in M$, and let $\omega = \omega_i dx^i$ be a closed covector field on M.

As in the proof of Theorem 4.22, we will construct a potential function for ω by integrating along smooth curve segments from c. However, in this case we do not know a priori that the line integrals are path-independent, so we must integrate along specific paths.

For any point $x \in M$, let $\gamma_x \colon [0,1] \to M$ denote the line segment from c to x, parametrized as follows:

$$\gamma_x(t) = c + t(x - c).$$

The hypothesis guarantees that the image of γ_x lies entirely in M for each $x \in M$. Define a function $f \colon M \to \mathbb{R}$ by

$$f(x) = \int_{\gamma_x} \omega.$$

We will show that f is a potential for ω, or equivalently that $\partial f / \partial x^i = \omega_i$ for $i = 1, \ldots, n$. To begin, we compute

$$\begin{aligned} f(x) &= \int_0^1 \omega_{\gamma_x(t)}(\gamma_x'(t))\, dt \\ &= \int_0^1 \omega_i(c + t(x-c))(x^i - c^i)\, dt. \end{aligned}$$

To compute the partial derivatives of f, we note that the integrand is smooth in all variables, so it is permissible to differentiate under the integral sign to obtain

$$\frac{\partial f}{\partial x^j}(x) = \int_0^1 \left(t \frac{\partial \omega_i}{\partial x^j}(c + t(x-c))(x^i - c^i) + \omega_j(c + t(x-c)) \right) dt.$$

Because ω is closed, this reduces to

$$\begin{aligned}
\frac{\partial f}{\partial x^j}(x) &= \int_0^1 \left(t \frac{\partial \omega_j}{\partial x^i}(c + t(x-c))(x^i - c^i) + \omega_j(c + t(x-c)) \right) dt \\
&= \int_0^1 \frac{d}{dt}\bigl(t\omega_j(c + t(x-c))\bigr)\, dt \\
&= \Bigl[t\omega_j(c + t(x-c)) \Bigr]_{t=0}^{t=1} \\
&= \omega_j(x),
\end{aligned}$$

which was to be proved. □

The key to the above construction is that we can reach every point $x \in M$ by a definite path γ_x from c to x, chosen in such a way that γ_x varies smoothly as x varies. That is what fails in the case of the closed covector field ω on the punctured plane (Example 4.23): Because of the hole, it is impossible to choose a smoothly-varying family of paths starting at a fixed base point and reaching every point of the domain. In Chapter 11, we will generalize Proposition 4.27 to show that every closed covector field is exact on any simply connected manifold.

When you actually have to *compute* a potential function for a given covector field that is known to be exact, there is a much simpler procedure that almost always works. Rather than describe it in complete generality, we illustrate it with an example.

Example 4.28. Let ω be a smooth covector field on $\mathbb{R}^3$, say

$$\omega = e^{y^2}\, dx + 2xye^{y^2}\, dy - 2z\, dz.$$

You can check that ω is closed. If f is a potential for ω, we must have

$$\begin{aligned}
\frac{\partial f}{\partial x} &= e^{y^2}, \\
\frac{\partial f}{\partial y} &= 2xye^{y^2}, \\
\frac{\partial f}{\partial z} &= -2z.
\end{aligned}$$

Holding y and z fixed and integrating the first equation with respect to x, we obtain

$$f(x,y,z) = \int e^{y^2}\, dx = xe^{y^2} + C_1(y,z),$$

where the "constant" of integration $C_1(y,z)$ may depend on the choice of (y,z). Now the second equation implies

$$\begin{aligned} 2xye^{y^2} &= \frac{\partial}{\partial y}\big(xe^{y^2} + C_1(y,z)\big) \\ &= 2xye^{y^2} + \frac{\partial C_1}{\partial y}, \end{aligned}$$

which forces $\partial C_1/\partial y = 0$, so C_1 is actually a function of z only. Finally, the third equation implies

$$\begin{aligned} -2z &= \frac{\partial}{\partial z}\big(xe^{y^2} + C_1(z)\big) \\ &= \frac{\partial C_1}{\partial z}, \end{aligned}$$

from which we conclude that $C_1(z) = -2z^2 + C$, where C is an arbitrary constant. Thus a potential function for ω is given by $f(x,y,z) = xe^{y^2} - 2z^2$. Any other potential differs from this one by a constant.

You should convince yourself that the formal procedure we followed in this example is equivalent to choosing an arbitrary base point $c \in \mathbb{R}^3$, and defining $f(x,y,z)$ by integrating ω along a path from c to (x,y,z) consisting of three straight line segments parallel to the axes. This works for any closed covector field defined on an open rectangle in $\mathbb{R}^n$ (which we know must be exact, because a rectangle is convex). In practice, once a formula is found for f on some open rectangle, the same formula typically works for the entire domain. (This is because most of the covector fields for which one can explicitly compute the integrals as we did above are real-analytic, and real-analytic functions are determined by their behavior in any open set.)

Problems

4-1. In each of the cases below, M is a smooth manifold and $f\colon M \to \mathbb{R}$ is a smooth function. Compute the coordinate representation for df, and determine the set of all points $p \in M$ at which $df_p = 0$.

(a) $M = \{(x, y) \in \mathbb{R}^2 : x > 0\}$; $f(x, y) = x/(x^2 + y^2)$. Use standard coordinates (x, y).

(b) M and f are as in part (a); this time use polar coordinates (r, θ).

(c) $M = S^2 \subset \mathbb{R}^3$; $f(p) = z(p)$ (the z-coordinate of p, thought of as a point in $\mathbb{R}^3$). Use stereographic coordinates.

(d) $M = \mathbb{R}^n$; $f(x) = |x|^2$. Use standard coordinates.

4-2. Let M be a smooth manifold.

(a) Given a smooth covector field σ on M, show that the map $\widetilde{\sigma}\colon \mathcal{T}(M) \to C^\infty(M)$ defined by

$$\widetilde{\sigma}(X)(p) = \sigma_p(X_p)$$

is linear over $C^\infty(M)$, in the sense that for any smooth functions $f, f' \in C^\infty(M)$ and smooth vector fields X, X',

$$\widetilde{\sigma}(fX + f'X') = f\widetilde{\sigma}(X) + f'\widetilde{\sigma}(X').$$

(b) Show that a map

$$\widetilde{\sigma}\colon \mathcal{T}(M) \to C^\infty(M)$$

is induced by a smooth covector field as above if and only if it is linear over $C^\infty(M)$.

4-3. The *length* of a smooth curve $\gamma\colon [a, b] \to \mathbb{R}^n$ is defined to be the value of the (ordinary) integral

$$L(\gamma) = \int_a^b |\gamma'(t)|\, dt.$$

Show that there is no smooth covector field $\omega \in \mathcal{T}^*(\mathbb{R}^n)$ with the property that $\int_\gamma \omega = L(\gamma)$ for every smooth curve γ.

4-4. Use Proposition 4.27 to give a simpler proof of Lemma 4.25.

4-5. Line Integrals of Vector Fields: Suppose X is a smooth vector field on an open set $U \subset \mathbb{R}^n$, thought of as a smooth function from

$\mathbb{R}^n$ to $\mathbb{R}^n$. For any piecewise smooth curve segment $\gamma\colon [a,b] \to U$, define the line integral of X over γ by

$$\int_\gamma X \cdot ds = \int_a^b X(\gamma(t)) \cdot \gamma'(t)\, dt,$$

and say X is conservative if its line integral around any closed curve is zero.

(a) Show that X is conservative if and only if there exists a smooth function $f \in C^\infty(U)$ such that $X = \operatorname{grad} f$. [Hint: Consider the covector field $\omega_x(Y) = X(x) \cdot Y$, where the dot denotes the Euclidean dot product.]

(b) If $n = 3$ and X is conservative, show $\operatorname{curl} X = 0$, where

$$\operatorname{curl} X = \left(\frac{\partial X^3}{\partial x^2} - \frac{\partial X^2}{\partial x^3}\right)\frac{\partial}{\partial x^1} + \left(\frac{\partial X^1}{\partial x^3} - \frac{\partial X^3}{\partial x^1}\right)\frac{\partial}{\partial x^2} + \left(\frac{\partial X^2}{\partial x^1} - \frac{\partial X^1}{\partial x^2}\right)\frac{\partial}{\partial x^3}.$$

(c) If $U \subset \mathbb{R}^3$ is star-shaped, show that X is conservative on U if and only if $\operatorname{curl} X = 0$.

4-6. If M is a compact manifold and $f \in C^\infty(M)$, show that df vanishes somewhere on M.

4-7. Is there a smooth covector field on $\mathbb{S}^2$ that vanishes at exactly one point? If so, can it be chosen to be exact?

4-8. Let $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ denote the n-torus. For each $i = 1, \dots, n$, let $\gamma_i\colon [0,1] \to \mathbb{T}^n$ be the curve segment

$$\gamma_i(t) = (1, \dots, e^{2\pi i t}, \dots, 1) \qquad \text{(with } e^{2\pi i t} \text{ in the } i\text{th place)},$$

where we think of $\mathbb{T}^n$ as a subset of $\mathbb{C}^n \cong \mathbb{R}^{2n}$. Show that a closed covector field ω on $\mathbb{T}^n$ is exact if and only if $\int_{\gamma_i} \omega = 0$ for $i = 1, \dots, n$. [Hint: Consider first $E^*\omega$, where $E\colon \mathbb{R}^n \to \mathbb{T}^n$ is the covering map $E(x^1, \dots, x^n) = (e^{2\pi i x^1}, \dots, e^{2\pi i x^n})$.]

4-9. If $F\colon M \to N$ is a smooth map, show that $F^*\colon T^*N \to T^*M$ is smooth.

4-10. Consider the smooth function $\det\colon \mathrm{GL}(n,\mathbb{R}) \to \mathbb{R}$.

(a) Using matrix entries (A_i^j) as global coordinates on $\mathrm{GL}(n,\mathbb{R})$, show that the partial derivatives of the determinant map $\det\colon \mathrm{GL}(n,\mathbb{R}) \to \mathbb{R}$ are given by

$$\frac{\partial}{\partial A_i^j} \det(A) = (\det A)(A^{-1})_j^i.$$

[Hint: Expand $\det A$ by minors along the ith column and use Cramer's rule.]

(b) Conclude that the differential of the determinant function is

$$d(\det)_A(B) = (\det A)\operatorname{tr}(A^{-1}B)$$

for $A \in \mathrm{GL}(n,\mathbb{R})$ and $B \in T_A\,\mathrm{GL}(n,\mathbb{R}) \cong \mathrm{M}(n,\mathbb{R})$, where $\operatorname{tr} A = \sum_i A_i^i$ is the trace of A.

5
Submanifolds

Many of the most familiar examples of manifolds arise naturally as subsets of other manifolds—for example, the n-sphere is a subset of $\mathbb{R}^{n+1}$ and the n-torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ is a subset of $\mathbb{C} \times \cdots \times \mathbb{C} = \mathbb{C}^n$. In this chapter, we will explore conditions under which a subset of a smooth manifold can be considered as a smooth manifold in its own right. As you will soon discover, the situation is quite a bit more subtle than the analogous theory of topological subspaces.

Because submanifolds are typically presented as images or level sets of smooth maps, a good portion of the chapter is devoted to analyzing the conditions under which such sets are smooth manifolds. We begin by introducing three special types of maps whose level sets and images are well behaved: submersions, immersions, and embeddings. Then we define the most important type of smooth submanifolds, called embedded submanifolds. These are modeled locally on linear subspaces of Euclidean space.

Next, in order to show how submersions, immersions, and embeddings can be used to define submanifolds, we will prove an analytic result that will prove indispensable in the theory of smooth manifolds: the inverse function theorem. This theorem and its corollaries show that, under certain hypotheses on the rank of its push-forward, a smooth map behaves locally like its push-forward.

The remainder of the chapter consists of various applications of the inverse function theorem to the study of submanifolds. We show that level sets of submersions, level sets of constant-rank smooth maps, and images of embeddings are embedded submanifolds. We also observe that the image of an injective immersion looks locally like an embedded submanifold, but may not be one globally; this leads to the definition of a more general kind of submanifold, called an immersed submanifold.

At the end of the chapter, we apply the theory of submanifolds to study conditions under which an algebraic subgroup of a Lie group is itself a Lie group.

Submersions, Immersions, and Embeddings

Because the push-forward of a smooth map F at a point p represents the "best linear approximation" to F near p, we can learn something about F itself by studying linear-algebraic properties of its push-forward at each point. The most important such property is its rank (the dimension of its image).

If $F\colon M \to N$ is a smooth map, we define the *rank* of F at $p \in M$ to be the rank of the linear map $F_*\colon T_pM \to T_{F(p)}N$; it is of course just the rank of the matrix of partial derivatives of F in any coordinate chart, or the dimension of $\operatorname{Im} F_* \subset T_{F(p)}N$. If F has the same rank k at every point, we say it has *constant rank*, and write $\operatorname{rank} F = k$.

An *immersion* is a smooth map $F\colon M \to N$ with the property that F_* is injective at each point (or equivalently $\operatorname{rank} F = \dim M$). Similarly, a *submersion* is a smooth map $F\colon M \to N$ such that F_* is surjective at each point (equivalently, $\operatorname{rank} F = \dim N$). As we will see in this chapter, immersions and submersions behave locally like injective and surjective linear maps, respectively.

One special kind of immersion is particularly important. A (*smooth*) *embedding* is an injective immersion $F\colon M \to N$ that is also a topological embedding, i.e., a homeomorphism onto its image $F(M) \subset N$ in the subspace topology. Since this is the primary kind of embedding we will be concerned with in this book, the term "embedding" will always mean smooth embedding unless otherwise specified.

Example 5.1 (Submersions, Immersions, and Embeddings).

(a) If $M_1, \dots, M_k$ are smooth manifolds, each of the projections $\pi_i\colon M_1 \times \cdots \times M_k \to M_i$ is a submersion. In particular, the projection $\pi\colon \mathbb{R}^{n+k} \to \mathbb{R}^n$ onto the first n coordinates is a submersion.

(b) Similarly, with $M_1, \dots, M_k$ as above, if $p_i \in M_i$ are arbitrarily chosen points, each of the maps $\iota_j\colon M_j \to M_1 \times \cdots \times M_k$ given by

$$\iota_j(q) = (p_1, \dots, p_{j-1}, q, p_{j+1}, \dots, p_k)$$

is an embedding. In particular, the inclusion map $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+k}$ given by sending $(x^1, \dots, x^n)$ to $(x^1, \dots, x^n, 0, \dots, 0)$ is an embedding.

(c) If $\gamma\colon J \to M$ is a smooth curve in a smooth manifold M, then γ is an immersion if and only if $\gamma'(t) \neq 0$ for all $t \in J$.

(d) Any smooth covering map $\pi\colon \widetilde{M} \to M$ is both an immersion and a submersion.

(e) If E is a smooth vector bundle over a smooth manifold M, the projection map $\pi\colon E \to M$ is a submersion.

(f) Let $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ denote the torus. The smooth map $F\colon \mathbb{T}^2 \to \mathbb{R}^3$ given by

$$F(e^{i\varphi}, e^{i\theta}) = ((2 + \cos\varphi)\cos\theta, (2 + cos\varphi)\sin\theta, \sin\varphi)$$

is a smooth embedding of $\mathbb{T}^2$ into $\mathbb{R}^3$ whose image is the doughnut-shaped surface obtained by revolving the circle $(y-2)^2 + z^2 + 1$ about the z-axis.

Exercise 5.1. Verify the claims in the preceding example.

To understand more fully what it means to be an embedding, it is useful to bear in mind some examples of injective immersions that are *not* embeddings. The next two examples illustrate two rather different ways in which an injective immersion can fail to be an embedding.

Example 5.2. Consider the map $\gamma\colon (-\pi/2, 3\pi/2) \to \mathbb{R}^2$ given by

$$\gamma(t) = (\sin 2t, \cos t).$$

Its image is a curve that looks like a figure eight in the plane (Figure 5.1). (It is the locus of points (x, y) where $x^2 = 4y^2(1 - y^2)$, as you can check.) It is easy to check that it is an injective immersion because $\gamma'(t)$ never vanishes; but it is not a topological embedding, because its image is compact in the subspace topology while its domain is not.

Example 5.3. Let $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{C}^2$ denote the torus, and let c by any irrational number. The map $\gamma\colon \mathbb{R} \to \mathbb{T}^2$ given by

$$\gamma(t) = (e^{2\pi i t}, e^{2\pi i c t})$$

is an immersion because $\gamma'(t)$ never vanishes. It is also injective, because $\gamma(t_1) = \gamma(t_2)$ implies that both $t_1 - t_2$ and $ct_1 - ct_2$ are integers, which is impossible unless $t_1 = t_2$.

Consider the set $\gamma(\mathbb{Z}) = \{\gamma(n) : n \in \mathbb{Z}\}$. If γ were a homeomorphism onto its image, this set would have no limit point in $\gamma(\mathbb{R})$, because $\mathbb{Z}$ has no limit point in $\mathbb{R}$. However, we will show that $\gamma(0)$ is a limit point of $\gamma(\mathbb{Z})$. To prove this claim, we need to show that given any $\varepsilon > 0$, there is a nonzero integer k such that $|\gamma(k) - \gamma(0)| < \varepsilon$.

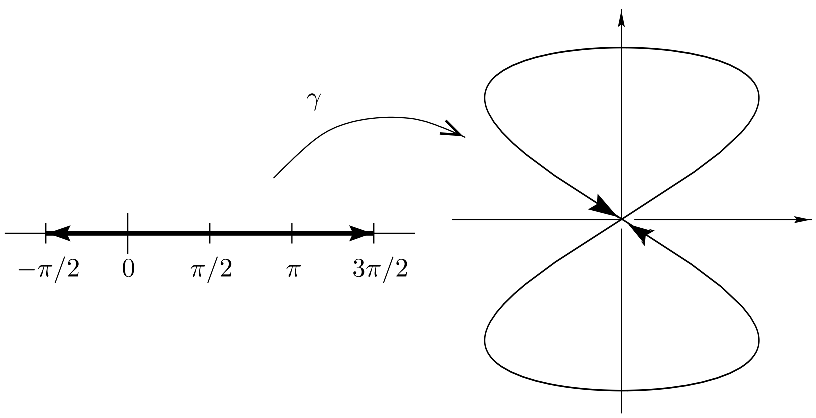

FIGURE 5.1. The figure eight curve of Example 5.2.

Since $\mathbb{S}^1$ is compact, the infinite set $\{e^{2\pi icn} : n \in \mathbb{Z}\}$ has a limit point, say $z_0 \in \mathbb{S}^1$. Given $\varepsilon > 0$, we can choose distinct integers n_1 and n_2 such that $|e^{2\pi icn_j} - z_0| < \varepsilon/2$, and therefore $|e^{2\pi icn_1} - e^{2\pi icn_2}| < \varepsilon$. Taking $k = n_1 - n_2$, this implies that

$$|e^{2\pi ick} - 1| = |e^{-2\pi in_2}(e^{2\pi icn_1} - e^{2\pi icn_2})| = |e^{2\pi icn_1} - e^{2\pi icn_2}| < \varepsilon,$$

and so

$$|\gamma(k) - \gamma(0)| = |(1, e^{2\pi ick}) - (1,1)| < \varepsilon.$$

In fact, it is not hard to show that the image set $\gamma(\mathbb{R})$ is actually dense in $\mathbb{T}^2$ (see Problem 5-4).

As the next lemma shows, one simple criterion that rules out such cases is to require that F be a closed map (i.e., V closed in M implies $F(V)$ closed in N), for then it follows easily that it is a homeomorphism onto its image. Another is that F be a *proper map*, which means that for any compact set $K \subset N$, the inverse image $F^{-1}(K)$ is compact.

Proposition 5.4. *Suppose $F\colon M \to N$ is an injective immersion. If any one of the following conditions holds, then F is an embedding with closed image.*

(*a*) *F is a closed map.*

(*b*) *F is a proper map.*

(*c*) *M is compact.*

Proof. For set-theoretic reasons, there exists an inverse map $F^{-1}\colon F(M) \to M$, and F is an embedding if and only if F^{-1} is continuous. If F is closed, then for every closed set $V \subset M$, $(F^{-1})^{-1}(V) = F(V)$ is closed in N and therefore also in $F(M)$. This implies F^{-1} is continuous, and proves (a). Every proper map between manifolds is closed (see [Lee00, Prop. 4.32]), so (a) implies (b). Finally, a simple topological argument (see [Lee00, Lemma 4.25] shows that every continuous map from a compact space to a Hausdorff space is closed, so (a) implies (c) as well. □

Exercise 5.2. Show that a composition of submersions is a submersion, a composition of immersions is an immersion, and a composition of embeddings is an embedding.

Embedded Submanifolds

Smooth submanifolds are modeled locally on the standard embedding of $\mathbb{R}^k$ into $\mathbb{R}^n$, identifying $\mathbb{R}^k$ with the subspace

$$\{(x^1, \dots, x^k, x^{k+1}, \dots, x^n) : x^{k+1} = \dots = x^n = 0\}$$

of $\mathbb{R}^n$. Somewhat more generally, if U is an open subset of $\mathbb{R}^n$, a *k-slice* of U is any subset of the form

$$S = \{(x^1, \dots, x^k, x^{k+1}, \dots, x^n) \in U : x^{k+1} = c^{k+1}, \dots, x^n = c^n\}$$

for some constants $c^{k+1}, \dots, c^n$. Clearly any k-slice is homeomorphic to an open subset of $\mathbb{R}^k$. (Sometimes it is convenient to consider slices defined by setting some other subset of the coordinates equal to constants instead of the last ones. The meaning should be clear from the context.)

Let M be a smooth n-manifold, and let (U, φ) be a smooth chart on M. We say a subset $S \subset U$ is a k-slice of U if $\varphi(S)$ is a k-slice of $\varphi(U)$. A subset $N \subset M$ is called an *embedded submanifold of dimension k* (or an *embedded k-submanifold* or a *regular submanifold*) of M if for each point $p \in N$ there exists a chart (U, φ) for M such that $p \in U$ and $U \cap N$ is a k-slice of U. In this situation, we call the chart (U, φ) a *slice chart for N in M*, and the corresponding coordinates $(x^1, \dots, x^n)$ are called *slice coordinates*. The difference $n - k$ is called the *codimension* of N in M. By convention, we consider an open submanifold to be an embedded submanifold of codimension zero.

The definition of an embedded submanifold is a local one. It is useful to express this formally as a lemma.

Lemma 5.5. *Let M be a smooth manifold and N a subset of M. Suppose every point $p \in N$ has a neighborhood $U \subset M$ such that $U \cap N$ is an embedded submanifold of U. Then N is an embedded submanifold of M.*

Exercise 5.3. Prove Lemma 5.5.

The next proposition explains the reason for the name "embedded submanifold."

Proposition 5.6. *Let $N \subset M$ be an embedded k-dimensional submanifold of M. With the subspace topology, N is a topological manifold of dimension k, and it has a unique smooth structure such that the inclusion map $N \hookrightarrow M$ is a smooth embedding.*

Proof. N is automatically Hausdorff and second countable because M is, and both properties are inherited by subspaces. To see that it is locally Euclidean, we will construct an atlas. For this proof, let $\pi\colon \mathbb{R}^n \to \mathbb{R}^k$ denote the projection onto the first k coordinates. For any slice chart (U, φ), let

$$\begin{aligned} V &= U \cap N, \\ \widetilde{V} &= \pi \circ \varphi(V), \\ \psi &= \pi \circ \varphi|_V \colon V \to \widetilde{V}. \end{aligned}$$

Then ψ is easily seen to be a homeomorphism, because it has a continuous inverse given by $\varphi^{-1} \circ j|_{\widetilde{V}}$, where $j\colon \mathbb{R}^k \to \mathbb{R}^n$ is the smooth map

$$j(x^1, \dots, x^k) = (x^1, \dots, x^k, c^{k+1}, \dots, c^n).$$

Thus N is a topological k-manifold, and the inclusion map $\iota\colon N \hookrightarrow M$ is a topological embedding (i.e., a homeomorphism onto its image).

To see that N is a smooth manifold, we need to check that the charts constructed above are smoothly compatible. Suppose (U, φ) and (U', φ') are two slice charts for N in M, and let (V, ψ), (V', ψ') be the corresponding charts for N. The transition map is given by $\psi' \circ \psi^{-1} = \pi \circ \varphi' \circ \varphi^{-1} \circ j$, which is a composition of the smooth maps π, $\varphi' \circ \varphi^{-1}$, and j. Thus the atlas we have constructed is in fact a smooth atlas, and defines a smooth structure on N. In any such chart, the inclusion map $N \hookrightarrow M$ has the coordinate representation

$$(x^1, \dots, x^k) \mapsto (x^1, \dots, x^k, c^{k+1}, \dots, c^n),$$

which is obviously an immersion. Since the inclusion is an injective immersion and a topological embedding, it is a smooth embedding as claimed.

The last thing we have to prove is that this is the unique smooth structure making the inclusion map a smooth embedding. Suppose that $\mathcal{A}$ is a (possibly different) smooth structure on N with the property that $N \hookrightarrow M$ is a smooth embedding. To show that the given smooth structure is the same as the one we have constructed, it suffices to show that each of the charts we constructed above is compatible with every chart in $\mathcal{A}$. Thus let (U, φ) be a slice chart for N in M, let (V, ψ) be the corresponding chart for N constructed above, and let (W, θ) be an arbitrary chart in $\mathcal{A}$. We need to show that $\psi \circ \theta^{-1}\colon \theta(W \cap V) \to \psi(W \cap V)$ is a diffeomorphism.

Observe first that $\psi \circ \theta^{-1}$ is a homeomorphism, and is smooth because it can be written as the following composition of smooth maps:

$$\theta(W \cap V) \xrightarrow{\theta^{-1}} W \cap V \overset{\iota}{\hookrightarrow} U \xrightarrow{\varphi} \mathbb{R}^n \xrightarrow{\pi} \mathbb{R}^k,$$

where we think of $W \cap V$ as an open subset of N (with the smooth structure $\mathcal{A}$) and U as an open subset of M. To prove that it is a diffeomorphism, we will show that $\psi \circ \theta^{-1}$ is an immersion and appeal to Proposition 5.7 below, which says that such a map is automatically a diffeomorphism. To show it is an immersion, we must show that its push-forward is injective. By the argument above, $(\psi \circ \theta^{-1})_* = \pi_* \circ \varphi_* \circ \iota_* \circ (\theta^{-1})_*$. Each of the linear maps φ_*, ι_*, and $(\theta^{-1})_*$ is injective—in fact φ_* and $(\theta^{-1})_*$ are bijective—and thus their composition is injective. Although π_* is not injective, the composition will be injective provided $\operatorname{Im}(\varphi \circ \iota \circ \theta^{-1})_* \cap \operatorname{Ker} \pi_* = \varnothing$ (see Exercise A.10(b) in the Appendix). Since ι takes its values in N, $\varphi \circ \iota \circ \theta^{-1}$ takes its values in the slice where the coordinates $x^{k+1}, \dots, x^n$ are constant:

$$\varphi \circ \iota \circ \theta^{-1}(y^1, \dots, y^k) = (x^1(y), \dots, x^k(y), c^{k+1}, \dots, c^n).$$

It follows easily that the push-forward of this map at any point takes its values in the span of $(e_1, \dots, e_k)$, which has trivial intersection with $\operatorname{Ker} \pi_* = \operatorname{span}(e_{k+1}, \dots, e_n)$. □

In general, a smooth homeomorphism need not have a smooth inverse. A simple counterexample is the map $F\colon \mathbb{R} \to \mathbb{R}$ given by $F(x) = x^3$, whose inverse map is not differentiable at the origin. The problem in this example is that the derivative of F vanishes at the origin, which forces the derivative of the inverse map to blow up. As the next proposition shows, if a smooth homeomorphism is also an immersion, the inverse map will be smooth.

Proposition 5.7 (Smoothness of Inverse Maps). *Suppose M and N are smooth manifolds of the same dimension, and $F\colon M \to N$ is a homeomorphism that is also a smooth immersion. Then F^{-1} is smooth, so F is a diffeomorphism.*

Proof. The only thing that needs to be proved is that F^{-1} is smooth, which is a local property, so by restricting to coordinate domains and replacing F with its coordinate representation, we may as well assume that M and N are open subsets of $\mathbb{R}^n$.

The assumption that F is an immersion means that the total derivative $DF(a)$ is injective for each $a \in M$, and therefore is invertible for dimensional reasons. If F^{-1} were differentiable at $b \in N$, the chain rule would imply

$$\begin{aligned} \mathrm{Id} &= D(F \circ F^{-1})(b) \\ &= DF(F^{-1}(b)) \circ DF^{-1}(b), \end{aligned}$$

from which it would follow that $DF^{-1}(b) = DF(F^{-1}(b))^{-1}$. We will begin by showing that F^{-1} is differentiable at each point of N, with total derivative given by this formula.

Let $b \in N$ and set $a = F^{-1}(b) \in M$. For $v, w \in \mathbb{R}^n$ small enough that $a + v \in M$ and $b + w \in N$, define $R(v)$ and $S(w)$ by

$$\begin{aligned} R(v) &= F(a+v) - F(a) - DF(a)v, \\ S(w) &= F^{-1}(b+w) - F^{-1}(b) - DF(a)^{-1}w. \end{aligned}$$

Because F is smooth, it is differentiable at a, which means that $\lim_{v \to 0} R(v)/|v| = 0$. We need to show that $\lim_{w \to 0} S(w)/|w| = 0$.

For sufficiently small $w \in \mathbb{R}^n$, define

$$v(w) = F^{-1}(b+w) - F^{-1}(b) = F^{-1}(b+w) - a.$$

It follows that

$$\begin{aligned} F^{-1}(b+w) = F^{-1}(b) + v(w) &= a + v(w), \\ w = (b+w) - b &= F(a + v(w)) - F(a), \end{aligned} \tag{5.1}$$

and therefore

$$\begin{aligned} S(w) &= F^{-1}(b+w) - F^{-1}(b) - DF(a)^{-1}w \\ &= v(w) - DF(a)^{-1}w \\ &= DF(a)^{-1}(DF(a)v(w) - w) \\ &= DF(a)^{-1}(DF(a)v(w) + F(a) - F(a+v(w))) \\ &= -DF(a)^{-1}R(v(w)). \end{aligned}$$

We will show below that there are positive constants c and C such that

$$c|w| \le |v(w)| \le C|w| \tag{5.2}$$

for all sufficiently small w. In particular, this implies that $v(w) \neq 0$ when w is sufficiently small and nonzero. From this together with the result of Exercise A.24 in the Appendix, we conclude that

$$\begin{aligned} \frac{|S(w)|}{|w|} &\le |DF(a)^{-1}| \frac{|R(v(w))|}{|w|} \\ &= |DF(a)^{-1}| \frac{|R(v(w))|}{|v(w)|} \frac{|v(w)|}{|w|} \\ &\le C|DF(a)^{-1}| \frac{|R(v(w))|}{|v(w)|}, \end{aligned}$$

which approaches zero as $w \to 0$ because $v(w) \to 0$ and F is differentiable.

To complete the proof that F^{-1} is differentiable, it remains only to prove (5.2). From the definition of $R(v)$ and (5.1),

$$\begin{aligned} v(w) &= DF(a)^{-1}DF(a)v(w) \\ &= DF(a)^{-1}(F(a+v(w)) - F(a) - R(v(w))) \\ &= DF(a)^{-1}(w - R(v(w))), \end{aligned}$$

which implies

$$|v(w)| \le |DF(a)^{-1}|\,|w| + |DF(a)^{-1}|\,|R(v(w))|.$$

Because $|R(v)|/|v| \to 0$ as $v \to 0$, there exists $\delta_1 > 0$ such that $|v| < \delta_1$ implies $|R(v)| \le |v|/(2|DF(a)^{-1}|)$. By continuity of F^{-1}, there exists $\delta_2 > 0$ such that $|w| < \delta_2$ implies $|v(w)| < \delta_1$, and therefore

$$|v(w)| \le |DF(a)^{-1}|\,|w| + (1/2)|v(w)|.$$

Subtracting $(1/2)|v(w)|$ from both sides, we obtain

$$|v(w)| \le 2|DF(a)^{-1}|\,|w|$$

whenever $|w| < \delta_2$. This is the second inequality of (5.2). To prove the first, we use (5.1) again to get

$$w = F(a+v(w)) - F(a) = DF(a)v(w) + R(v(w)).$$

Therefore, when $|w| < \delta_2$,

$$|w| \le |DF(a)|\,|v(w)| + |R(v(w))| \le \left(|DF(a)| + \frac{1}{2|DF(a)^{-1}|}\right)|v(w)|.$$

This completes the proof that F^{-1} is differentiable.

By Exercise A.25, the partial derivatives of F^{-1} are defined at each point $y \in N$. Observe that the formula $DF^{-1}(y) = DF(F^{-1}(y))^{-1}$ implies that the map $DF^{-1}\colon N \to \mathrm{GL}(n,\mathbb{R})$ can be written as the composition

$$N \xrightarrow{F^{-1}} M \xrightarrow{DF} GL(n,\mathbb{R}) \xrightarrow{i} GL(n,\mathbb{R}), \tag{5.3}$$

where $i(A) = A^{-1}$. Matrix inversion is a smooth map, because $\mathrm{GL}(n,\mathbb{R})$ is a Lie group. Also, DF is a smooth map because its component functions are the partial derivatives of F, which are assumed to be smooth. Because DF^{-1} is a composition of continuous maps, it is continuous, and therefore the partial derivatives of F^{-1} are continuous, which means that F^{-1} is of class C^1.

Now assume by induction that we have shown F^{-1} is of class C^k. This means that each of the maps in (5.3) is of class C^k. Because DF^{-1} is a composition of C^k functions, it is itself C^k; this implies that the partial derivatives of F^{-1} are of class C^k, so F^{-1} itself is of class C^{k+1}. Continuing by induction, we conclude that F^{-1} is smooth. □

The Tangent Space to an Embedded Submanifold

If N is an embedded submanifold of $\mathbb{R}^n$, we intuitively think of the tangent space T_pN at a point of N as a subspace of the tangent space $T_p\mathbb{R}^n$. Similarly, the tangent space to a submanifold of an abstract smooth manifold can be viewed as a subspace of the tangent space to the ambient manifold, once we make appropriate identifications.

Let M be a smooth manifold, and let $N \subset M$ be an embedded submanifold. Since the inclusion map $\iota\colon N \hookrightarrow M$ is an immersion, at each point $p \in N$ we have an injective linear map $\iota_*\colon T_pN \to T_pM$. We will adopt the convention of identifying T_pN with its image under this map, thereby thinking of T_pN as a certain linear subspace of T_pM. Thought of as a derivation, a vector $X \in T_pN$, identified with $\iota_*X \in T_pM$, acts on smooth functions on M in the following way:

$$Xf = (\iota_*X)f = X(f \circ \iota) = X\left(f|_N\right).$$

The next proposition gives a useful way to characterize T_pN as a subspace of T_pM.

Proposition 5.8. *Suppose $N \subset M$ is an embedded submanifold and $p \in N$. As a subspace of T_pM, the tangent space T_pN is given by*

$$T_pN = \{X \in T_pM : Xf = 0 \textit{ whenever } f \in C^\infty(M) \textit{ and } f|_N \equiv 0\}.$$

Proof. First suppose $X \in T_pN \subset T_pM$. This means, more precisely, that $X = \iota_*Y$ for some $Y \in T_pN$. If f is any smooth function on M that vanishes on N, then $f \circ \iota \equiv 0$, so

$$Xf = (\iota_*Y)f = Y(f \circ \iota) \equiv 0.$$

Conversely, if $X \in T_pM$ satisfies $Xf = 0$ whenever f vanishes on N, we need to show that there is a vector $Y \in T_pN$ such that $X = \iota_*Y$. Let $(x^1, \dots, x^n)$ be slice coordinates for N in some neighborhood U of p, so that $U \cap N$ is the subset of U where $x^{k+1} = \cdots = x^n = 0$, and $(x^1, \dots, x^k)$ are coordinates for $U \cap N$. Because the inclusion map $\iota\colon N \cap U \hookrightarrow M$ has the coordinate representation

$$\iota(x^1, \dots, x^k) = (x^1, \dots, x^k, 0, \dots, 0)$$

in these coordinates, it follows that T_pN (that is, ι_*T_pN) is exactly the subspace of T_pM spanned by $\partial/\partial x^1|_p, \dots, \partial/\partial x^k|_p$. If we write the coordinate representation of X as

$$X = \sum_{i=1}^n X^i \left.\frac{\partial}{\partial x^i}\right|_p,$$

we see that $X \in T_pN$ if and only if $X^i = 0$ for $i > k$.

Let φ be a bump function supported in U that is equal to 1 in a neighborhood of p. Choose an index $j > k$, and consider the function $f(x) = \varphi(x)x^j$, extended to be zero on $M \smallsetminus U$. Then f vanishes identically on N, so

$$0 = Xf = \sum_{i=1}^{n} X^i \frac{\partial(\varphi(x)x^j)}{\partial x^i}(p) = X^j.$$

Thus $X \in T_pN$ as desired. □

Examples of Embedded Submanifolds

One straightforward way to construct embedded submanifolds is by using the graphs of smooth functions. Let $U \subset \mathbb{R}^n$ be an open set, and let $F\colon U \to \mathbb{R}^k$ be a smooth function. The *graph* of F is the subset of $\mathbb{R}^n \times \mathbb{R}^k$ defined by

$$\Gamma(F) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^k : x \in U \text{ and } y = F(x)\}.$$

Lemma 5.9 (Graphs as Submanifolds). *If $U \subset \mathbb{R}^n$ is open and $F\colon U \to \mathbb{R}^k$ is smooth, then the graph of F is an embedded n-dimensional submanifold of $\mathbb{R}^{n+k}$.*

Proof. Define a map $\varphi\colon U \times \mathbb{R}^k \to U \times \mathbb{R}^k$ by

$$\varphi(x, y) = (x, y - F(x)).$$

It is clearly smooth, and in fact it is a diffeomorphism because its inverse can be written explicitly:

$$\varphi^{-1}(u, v) = (u, v + F(u)).$$

Because $\varphi(\Gamma(F))$ is the slice $\{(u, v) : v = 0\}$ of $U \times \mathbb{R}^k$, this shows that $\Gamma(F)$ is an embedded submanifold. □

Example 5.10 (Spheres). To show that $\mathbb{S}^n$ is an embedded submanifold of $\mathbb{R}^{n+1}$, we use the preceding lemma together with Lemma 5.5. Let $\mathbb{B}^n$ be the open unit ball in $\mathbb{R}^n$, and define functions $F^{\pm}\colon \mathbb{B}^n \to \mathbb{R}$ by

$$F^{\pm}(u) = \pm\sqrt{1 - |u|^2}.$$

For any $i \in \{1, \ldots, n\}$, the intersection of $\mathbb{S}^n$ with the open set where $x^i > 0$ is the graph of the smooth function

$$x^i = F^+(x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^{n+1}).$$

Similarly, the intersection of $\mathbb{S}^n$ with $\{x : x^i < 0\}$ is the graph of F^-. Since every point in $\mathbb{S}^n$ is in one of these sets, Lemma 5.5 shows that $\mathbb{S}^n$ is an embedded submanifold of $\mathbb{R}^{n+1}$. The smooth structure thus induced on $\mathbb{S}^n$ is the same as the one we defined in Chapter 1: In fact, the coordinates for $\mathbb{S}^n$ defined by these slice charts are exactly the graph coordinates we defined in Example 1.11.

Exercise 5.4. Let $U = \{(x, y, z) : x, y, z > 0\} \subset \mathbb{R}^3$. and let $\Phi\colon U \to \mathbb{R}^3$ be the *spherical coordinate map*

$$\begin{aligned}\Phi(x, y, z) &= (\rho, \varphi, \theta)\\ &= \left(\sqrt{x^2 + y^2 + z^2},\ \cos^{-1}\frac{z}{\sqrt{x^2 + y^2 + z^2}},\ \cos^{-1}\frac{x}{\sqrt{x^2 + y^2}}\right).\end{aligned}$$

Show that (U, Φ) is a slice chart for $\mathbb{S}^2$ in $\mathbb{R}^3$.

As Example 5.10 illustrates, showing directly from the definition that a subset of a manifold is an embedded submanifold can be somewhat cumbersome. In practice, submanifolds are usually presented to us in one of the following two ways:

- Level set of a smooth map: Many submanifolds are most naturally defined as the set of points where some smooth map takes on a fixed value, called a *level set* of the map. For example, the n-sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ is defined as the level set $f^{-1}(1)$, where $f\colon \mathbb{R}^{n+1} \to \mathbb{R}$ is the function $f(x) = |x|^2$.

- Image of a smooth map: In some cases, it is more natural to describe a submanifold as the image of a smooth map. For example, the map $F\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $F(\theta, \varphi) = ((2 + \cos\varphi)\cos\theta, (2 + \cos\varphi)\sin\theta, \sin\varphi)$ has as its image a doughnut-shaped torus of revolution.

Thus two important questions we will need to address are:

- When is a level set of a smooth map an embedded submanifold?

- When is the image of a smooth map an embedded submanifold?

It is easy to construct examples of both cases that are not embedded submanifolds. For example, the smooth map $F\colon \mathbb{R} \to \mathbb{R}^2$ defined by $F(t) = (t^2, t^3)$ has as its image a curve that has a "cusp" or "kink" at the origin. Similarly, the map $\Phi\colon \mathbb{R}^2 \to \mathbb{R}$ given by $\Phi(x, y) = xy$ has the union of the x and y axes as its zero level set. As Problem 5-5 shows, neither of these sets is an embedded submanifold of $\mathbb{R}^2$.

The rest of this chapter is devoted to the study of these two questions, and in particular to developing sufficient conditions under which both kinds of sets *are* embedded submanifolds.

The Inverse Function Theorem and Its Friends

The key to answering the questions at the end of the previous section is understanding how the local behavior of a smooth map is modeled by the behavior of its push-forward. To set the stage, we will consider a linear version of the problem: Let S be a k-dimensional linear subspace of $\mathbb{R}^n$, and let us examine how we might use linear maps to define it.

First, every subspace S is the kernel of some linear map. (Such a linear map is easily constructed by choosing a basis for S and extending it to a basis for $\mathbb{R}^n$.) By the rank-nullity law, if $S = \operatorname{Ker} L$, then $\operatorname{Im} L$ must have dimension $n - k$. Therefore, a natural way to define a k-dimensional subspace $S \subset \mathbb{R}^n$ is to give a surjective linear map $L\colon \mathbb{R}^n \to \mathbb{R}^{n-k}$ whose kernel is S. The vector equation $Lx = 0$ is equivalent to $n - k$ scalar equations, each of which can be thought of as "cutting out" one more dimension of S.

On the other hand, every subspace is also the image of some linear map. A choice of basis for S can be used to define an injective linear map $E\colon \mathbb{R}^k \to \mathbb{R}^n$ whose image is S. Such a map can be thought of as a "parametrization" of S.

In the context of smooth manifolds, the analogue of a surjective linear map is a submersion, and the analogue of an injective linear map is an immersion. Let M be an n-manifold. By analogy with the linear situation, we might expect that a level set of a submersion from M to an $(n - k)$-manifold is an embedded k-dimensional submanifold of M. We will see below that this is the case. Analogously, we might expect that the image of a smooth embedding from a k-manifold to M is an embedded k-dimensional submanifold. This is also the case.

The basis for all these results is the following analytic theorem. It is the simplest of several results we will develop in this section that show how the local behavior of a smooth map is modeled by the behavior of its push-forward.

Theorem 5.11 (Inverse Function Theorem). *Suppose M and N are smooth manifolds and $F\colon M \to N$ is a smooth map. If F_* is invertible at a point $p \in M$, then there exist connected neighborhoods U_0 of p and V_0 of $F(p)$ such that $F\colon U_0 \to V_0$ is a diffeomorphism.*

The proof of this theorem is based on the following elementary result about metric spaces. If X is a metric space, a map $G\colon X \to X$ is said to be a *contraction* if there is a constant $\lambda < 1$ such that $d(G(x), G(y)) \le \lambda d(x, y)$ for all $x, y \in X$. Clearly any contraction is continuous.

Lemma 5.12 (Contraction Lemma). *Let X be a complete metric space. Every contraction $G\colon X \to X$ has a unique fixed point, i.e., a point $x \in X$ such that $G(x) = x$.*

Proof. Uniqueness is immediate, for if x and x' are both fixed points of G, the contraction property implies $d(x,x') = d(G(x),G(x')) \le \lambda d(x,x')$, which is possible only if $x = x'$.

To prove the existence of a fixed point, let x_0 be an arbitrary point in X, and define a sequence $\{x_n\}$ inductively by $x_{n+1} = G(x_n)$. For any $i \ge 1$ we have $d(x_i,x_{i+1}) = d(G(x_{i-1}),G(x_i)) \le \lambda d(x_{i-1},x_i)$, and therefore by induction

$$d(x_i,x_{i+1}) \le \lambda^i d(x_0,x_1).$$

If $j \ge i \ge N$,

$$\begin{aligned} d(x_i,x_j) &\le d(x_i,x_{i+1}) + d(x_{i+1},x_{i+2}) + \cdots + d(x_{j-1},x_j) \\ &\le (\lambda^i + \cdots + \lambda^{j-1}) d(x_0,x_1) \\ &\le \lambda^i \left(\sum_{n=0}^{\infty} \lambda^n \right) d(x_0,x_1) \\ &\le \frac{\lambda^N}{1-\lambda} d(x_0,x_1). \end{aligned}$$

Since this last expression can be made as small as desired by choosing N large, the sequence $\{x_n\}$ is Cauchy and therefore converges to a limit $x \in X$. Because G is continuous,

$$G(x) = G\left(\lim_{n\to\infty} x_n \right) = \lim_{n\to\infty} G(x_n) = \lim_{n\to\infty} x_{n+1} = x,$$

so x is the desired fixed point. □

Proof of the inverse function theorem. The fact that $F_*\colon T_pM \to T_{F(p)}N$ is an isomorphism implies that M and N have the same dimension n. Choose coordinate domains U centered at p and V centered at $F(p)$; considering the coordinate maps as identifications as usual, we may as well assume that U and V are actually open subsets of $\mathbb{R}^n$ and $F(0) = 0$. The map $F_1 = DF(0)^{-1} \circ F$ satisfies $F_1(0) = 0$ and $DF_1(0) = \mathrm{Id}$. If the theorem is true for F_1, then it is true for $F = DF(0) \circ F_1$. Henceforth, replacing F by F_1, we will assume that F is defined in a neighborhood of 0, $F(0) = 0$, and $DF(0) = \mathrm{Id}$.

Let $H(x) = x - F(x)$. Then $DH(0) = \mathrm{Id} - \mathrm{Id} = 0$. Because the matrix entries of $DH(x)$ are continuous functions of x, there is a number $\varepsilon > 0$ such that $|DH(x)| \le 1/2$ for all $x \in \overline{B_\varepsilon(0)}$. If $x, x' \in B_\varepsilon(0)$, the Lipschitz estimate for smooth functions (Proposition A.28 in the Appendix) implies

$$|H(x') - H(x)| \le \tfrac{1}{2}|x' - x|. \tag{5.4}$$

Since $x' - x = F(x') - F(x) + H(x') - H(x)$, it follows that

$$|x' - x| \le |F(x') - F(x)| + |H(x') - H(x)| \le |F(x') - F(x)| + \tfrac{1}{2}|x' - x|.$$

Subtracting $\frac{1}{2}|x' - x|$ from both sides, we conclude that

$$|x' - x| \leq 2|F(x') - F(x)| \tag{5.5}$$

for all $x, x' \in \overline{B_\varepsilon(0)}$. In particular, this shows that F is injective on $B_\varepsilon(0)$.

Now let $y \in B_{\varepsilon/2}(0)$ be arbitrary. We will show that there exists $x \in B_\varepsilon(0)$ such that $F(x) = y$. Let $G(x) = H(x) + y = x - F(x) + y$, so that $G(x) = x$ if and only if $F(x) = y$. If $|x| \leq \varepsilon$, (5.4) implies

$$|G(x)| \leq |H(x)| + |y| < \frac{1}{2}|x| + \frac{\varepsilon}{2} \leq \varepsilon, \tag{5.6}$$

so G maps $\overline{B_\varepsilon(0)}$ to itself. It follows from (5.4) that $|G(x) - G(x')| = |H(x) - H(x')| \leq \frac{1}{2}|x' - x|$, so G is a contraction. Since $\overline{B_\varepsilon(0)}$ is compact and therefore complete, the contraction lemma implies that G has a unique fixed point $x \in \overline{B_\varepsilon(0)}$. From (5.6), $|x| = |G(x)| < \varepsilon$, so in fact $x \in B_\varepsilon(0)$, thus proving the claim.

Let $U = B_\varepsilon(0) \cap F^{-1}(B_{\varepsilon/2}(0))$. Then U is open, and the argument above shows that $F\colon U \to B_{\varepsilon/2}(0)$ is bijective, so $F^{-1}\colon B_{\varepsilon/2}(0) \to U$ exists. Substituting $x = F^{-1}(y)$ and $x' = F^{-1}(y')$ into (5.5) shows that F^{-1} is continuous. Let U_0 be the connected component of U containing 0, and $V_0 = F(U_0)$. Then $F\colon U_0 \to V_0$ is a homeomorphism, and Proposition 5.7 shows that it is a diffeomorphism. □

For our purposes, the most important consequence of the inverse function theorem is the following, which says that a constant-rank smooth map can be placed locally into a particularly simple canonical form by a change of coordinates. It is a nonlinear version of the canonical form theorem for linear maps (Theorem A.4).

Theorem 5.13 (Rank Theorem). *Suppose M and N are smooth manifolds of dimensions m and n, respectively, and $F\colon M \to N$ is a smooth map with constant rank k. For each $p \in M$ there exist coordinates $(x^1, \dots, x^m)$ centered at p and $(v^1, \dots, v^n)$ centered at $F(p)$ in which F has the following coordinate representation:*

$$F(x^1, \dots, x^m) = (x^1, \dots, x^k, 0, \dots, 0). \tag{5.7}$$

Proof. Replacing M and N by coordinate domains $U \subset M$ near p and $V \subset N$ near $F(p)$ and replacing F by its coordinate representation, we may as well assume that $F\colon U \to V$ where U is an open subset of $\mathbb{R}^m$ and V is an open subset of $\mathbb{R}^n$. The fact that $DF(p)$ has rank k implies that its matrix has some $k \times k$ minor with nonzero determinant. By reordering the coordinates, we may assume that it is the upper left minor, $(\partial F^i/\partial x^j)$ for $i, j = 1, \dots, k$. Let us relabel the standard coordinates as $(x, y) = (x^1, \dots, x^k, y^1, \dots, y^{m-k})$ in $\mathbb{R}^m$ and $(v, w) = (v^1, \dots, v^k, w^1, \dots, w^{n-k})$ in $\mathbb{R}^n$. If we write $F(x, y) = (Q(x, y), R(x, y))$, then our hypothesis is that $(\partial Q^i/\partial x^j)$ is nonsingular at p.

Define $\varphi\colon U \to \mathbb{R}^m$ by

$$\varphi(x, y) = (Q(x, y), y).$$

Its total derivative at p is

$$D\varphi(p) = \begin{pmatrix} \dfrac{\partial Q^i}{\partial x^j}(p) & \dfrac{\partial Q^i}{\partial y^j}(p) \\ 0 & I_{m-k} \end{pmatrix},$$

which is nonsingular because its columns are independent. Therefore, by the inverse function theorem, there are connected neighborhoods U_0 of p and V_0 of $\varphi(p)$ such that $\varphi\colon U_0 \to V_0$ is a diffeomorphism. Writing the inverse map as $\varphi^{-1}(x, y) = (A(x, y), B(x, y))$ for some smooth functions $A\colon V_0 \to \mathbb{R}^k$ and $B\colon V_0 \to \mathbb{R}^{m-k}$, we compute

$$\begin{aligned}(x, y) &= \varphi(\varphi^{-1}(x, y)) \\ &= \varphi(A(x, y), B(x, y)) \\ &= (Q(A(x, y), B(x, y)), B(x, y)).\end{aligned} \tag{5.8}$$

Comparing y components, it follows that $B(x, y) = y$, and therefore φ^{-1} has the form

$$\varphi^{-1}(x, y) = (A(x, y), y).$$

We will take $\varphi\colon U_0 \to \mathbb{R}^m$ as our coordinate chart near p. Observe that $\varphi \circ \varphi^{-1} = \mathrm{Id}$ implies $Q(A(x, y), y) = x$, and therefore $F \circ \varphi^{-1}$ has the form

$$F \circ \varphi^{-1}(x, y) = (x, \widetilde{R}(x, y)),$$

where we have put $\widetilde{R}(x, y) = R(A(x, y), y)$.

The Jacobian matrix of this map at an arbitrary point $(x, y) \in \varphi(U_0)$ is

$$D(F \circ \varphi^{-1})(x, y) = \begin{pmatrix} I_k & 0 \\ \dfrac{\partial \widetilde{R}^i}{\partial x^j} & \dfrac{\partial \widetilde{R}^i}{\partial y^j} \end{pmatrix}.$$

Since composing with a diffeomorphism does not change the rank of a map, this matrix has rank exactly equal to k everywhere in U_0. Since the first k columns are obviously independent, the rank can be k only if the partial derivatives $\partial \widetilde{R}^i / \partial y^j$ vanish identically on U_0, which implies that $\widetilde{R}$ is actually independent of $(y^1, \dots, y^{m-k})$. Thus in fact

$$F \circ \varphi^{-1}(x, y) = (x, \widetilde{R}(x)).$$

To complete the proof, we need to choose coordinates for $\mathbb{R}^n$ near $F(p)$. Define $\psi\colon V_0 \to \mathbb{R}^n$ by

$$\psi(v,w) = (v, w - \widetilde{R}(v)).$$

This is a diffeomorphism onto its image, because its inverse is given explicitly by $\psi^{-1}(v,w) = (v, w + \widetilde{R}(v))$; thus ψ is a coordinate chart on V_0. In terms of the coordinate maps φ for the domain and ψ for the range, F has the coordinate representation

$$\psi \circ F \circ \varphi^{-1}(x,y) = \psi(x, \widetilde{R}(x)) = (x, \widetilde{R}(x) - \widetilde{R}(x)) = (x, 0),$$

which was to be proved. □

The following corollary can be viewed as a somewhat more invariant statement of the rank theorem.

Corollary 5.14. *Let $F\colon M \to N$ be a smooth map, and suppose M is connected. Then the following are equivalent:*

(*a*) *For each $p \in M$ there exist coordinates near p and $F(p)$ in which the coordinate representation of F is linear.*

(*b*) *F has constant rank.*

Proof. First suppose F has a linear coordinate representation in a neighborhood of each point. Since any linear map has constant rank, it follows that the rank of F is constant in a neighborhood of each point, and thus by connectedness it is constant on all of M. Conversely, if F has constant rank, the rank theorem shows that it has the linear coordinate representation (5.7) in a neighborhood of each point. □

Another useful consequence of the inverse function theorem is the implicit function theorem, which gives conditions under which a level set of a smooth map is locally the graph of a smooth function.

Theorem 5.15 (Implicit Function Theorem). *Let $U \subset \mathbb{R}^n \times \mathbb{R}^k$ be an open set, and let $(x,y) = (x^1,\dots,x^n,y^1,\dots,y^k)$ denote the standard coordinates on U. Suppose $\Phi\colon U \to \mathbb{R}^k$ is a smooth map, $(a,b) \in U$, and $c = \Phi(a,b)$. If the $k \times k$ matrix*

$$\left(\frac{\partial \Phi^i}{\partial y^j}(a,b)\right)$$

is nonsingular, then there exist neighborhoods V_0 of a and W_0 of b and a smooth map $F\colon V_0 \to W_0$ such that $\Phi^{-1}(c) \cap V_0 \times W_0$ is the graph of F, i.e., $\Phi(x,y) = c$ for $(x,y) \in V_0 \times W_0$ if and only if $y = F(x)$.

Proof. Under the hypotheses of the theorem, consider the smooth map $\Psi\colon U \to \mathbb{R}^n \times \mathbb{R}^k$ defined by $\Psi(x,y) = (x, \Phi(x,y))$. Its total derivative at (a,b) is

$$D\Psi(a,b) = \begin{pmatrix} I_n & 0 \\ \dfrac{\partial \Phi^i}{\partial x^j}(a,b) & \dfrac{\partial \Phi^i}{\partial y^j}(a,b) \end{pmatrix},$$

which is nonsingular by hypothesis. Thus by the inverse function theorem there exist connected neighborhoods U_0 of (a,b) and Y_0 of (a,c) such that $\Psi\colon U_0 \to Y_0$ is a diffeomorphism. Shrinking U_0 and Y_0 if necessary, we may assume $U_0 = V \times W$ is a product neighborhood. Arguing exactly as in the proof of the rank theorem (but with the roles of x and y reversed), the inverse map has the form

$$\Psi^{-1}(x,y) = (x, B(x,y))$$

for some smooth map $B\colon Y_0 \to W$.

Now let $V_0 = \{x \in V : (x,c) \in Y_0\}$ and $W_0 = W$, and define $F\colon V_0 \to W_0$ by $F(x) = B(x,c)$. Comparing y components in the relation $(x,c) = \Psi \circ \Psi^{-1}(x,c)$ yields

$$c = \Phi(x, B(x,c)) = \Phi(x, F(x))$$

whenever $x \in V_0$, so the graph of F is contained in $\Phi^{-1}(c)$. Conversely, suppose $(x,y) \in V_0 \times W_0$ and $\Phi(x,y) = c$. Then $\Psi(x,y) = (x, \Phi(x,y)) = (x,c)$, so

$$(x,y) = \Psi^{-1}(x,c) = (x, B(x,c)) = (x, F(x)),$$

which implies that $y = F(x)$. This completes the proof. □

First Consequences

When we apply the inverse function theorem and its consequences to maps between manifolds, we obtain a wealth of important results. The first one is a significant strenghthening of Proposition 5.7.

Proposition 5.16. *Suppose M and N are smooth manifolds of the same dimension, and $F\colon M \to N$ is a smooth immersion. Then F is a local diffeomorphism. If F is bijective, it is a diffeomorphism.*

Proof. The fact that F is an immersion means that F_* is bijective at each point for dimensional reasons. Then the fact that F is a local diffeomorphism is an immediate consequence of the inverse function theorem. If F is bijective, then it is a diffeomorphism by Exercise 2.5. □

The next proposition is a powerful consequence of the rank theorem. (In Chapter 6, we will generalize this proposition to characterize surjective and bijective maps of constant rank.)

Proposition 5.17. *Let $F\colon M \to N$ be a smooth map of constant rank. If F is injective, then it is an immersion.*

Proof. Let $m = \dim M$, $n = \dim N$, and suppose F has constant rank k. If F is not an immersion, then $k < m$. By the rank theorem, in a neighborhood of any point there is a chart in which F has the coordinate representation

$$F(x^1, \ldots, x^m) = (x^1, \ldots, x^k, 0, \ldots, 0). \tag{5.9}$$

It follows that $F(0, \ldots, 0, \varepsilon) = F(0, \ldots, 0, 0)$ for any sufficiently small ε, so F is not injective. $\square$

Another important application of the rank theorem is to vastly expand our understanding of the properties of submersions. As the next few results show, surjective submersions play a role in smooth manifold theory closely analogous to the role played by quotient maps in topology.

Proposition 5.18 (Properties of Submersions). *Let $\pi\colon M \to N$ be a submersion.*

(*a*) *π is an open map.*

(*b*) *Every point of M is in the image of a smooth local section of π.*

(*c*) *If π is surjective, it is a quotient map.*

Proof. Given $p \in M$, let $q = \pi(p) \in N$. Because a submersion has constant rank, by the rank theorem we can choose coordinates $(x^1, \ldots, x^m)$ centered at p and $(y^1, \ldots, y^k)$ centered at q in which π has the coordinate representation $\pi(x^1, \ldots, x^m) = (x^1, \ldots, x^k)$. If ε is a sufficiently small positive number, the coordinate cube

$$\mathcal{C}_\varepsilon = \{x : |x^i| < \varepsilon \text{ for } i = 1, \ldots, m\}$$

is a neighborhood of p whose image under π is the cube

$$\mathcal{C}'_\varepsilon = \{y : |y^i| < \varepsilon \text{ for } i = 1, \ldots, k\}.$$

The map $\sigma\colon \mathcal{C}'_\varepsilon \to \mathcal{C}_\varepsilon$ whose coordinate representation is $\sigma(x^1, \ldots, x^k) = (x^1, \ldots, x^k, 0, \ldots, 0)$ is a smooth local section of π satisfying $\sigma(q) = p$. This proves (b).

Suppose W is any open subset of M and $q \in \pi(W)$. For any $p \in W$ such that $\pi(p) = q$, W contains an open coordinate cube $\mathcal{C}_\varepsilon$ centered at p as above, and thus $\pi(W)$ contains an open coordinate cube centered at $\pi(p)$. This proves that W is open, so (a) holds. Because a surjective open map is automatically a quotient map, (c) follows from (a). $\square$

The next three propositions provide important tools that we will use frequently when studying submersions. The general philosophy is one that we will see repeatedly in this book: to "push" a smooth object (such as a smooth map) down via a submersion, pull it back via local sections. It is convenient to introduce the following terminology: If $\pi\colon M \to N$ is any surjective map, the *fiber* of π over $q \in N$ is the set $\pi^{-1}(q) \subset M$.

Proposition 5.19. *Suppose M, N, and P are smooth manifolds, $\pi\colon M \to N$ is a surjective submersion, and $F\colon N \to P$ is any map. Then F is smooth if and only if $F \circ \pi$ is smooth:*

$$\begin{tikzcd} M \arrow[d, "\pi"'] \arrow[dr, "F \circ \pi"] & \\ N \arrow[r, "F"'] & P. \end{tikzcd}$$

Proof. If F is smooth, then $F \circ \pi$ is smooth by composition. Conversely, suppose that $F \circ \pi$ is smooth. For any $q \in N$, let $\sigma\colon U \to M$ be a smooth section of π defined on a neighborhood U of q. Then $\pi \circ \sigma = \mathrm{Id}_U$ implies

$$F|_U = F \circ \mathrm{Id}_U = (F \circ \pi) \circ \sigma,$$

which is a composition of smooth maps. This shows that F is smooth in a neighborhood of each point, so it is smooth. □

The next proposition gives a very general sufficient condition under which a smooth map can be "pushed down" by a submersion.

Proposition 5.20 (Passing Smoothly to the Quotient). *Suppose $\pi\colon M \to N$ is a surjective submersion. If $F\colon M \to P$ is a smooth map that is constant on the fibers of π, then there is a unique smooth map $\widetilde{F}\colon N \to P$ such that $\widetilde{F} \circ \pi = F$:*

$$\begin{tikzcd} M \arrow[d, "\pi"'] \arrow[dr, "F"] & \\ N \arrow[r, "\widetilde{F}"'] & P. \end{tikzcd}$$

Proof. Clearly, if $\widetilde{F}$ exists, it will have to satisfy $\widetilde{F}(q) = F(p)$ whenever $p \in \pi^{-1}(q)$. We use this to *define* $\widetilde{F}$: Given $q \in M$, let $\widetilde{F}(q) = F(p)$, where $p \in M$ is any point in the fiber over q. (Such a point exists because we are assuming that π is surjective.) This is well-defined because F is constant on the fibers of π, and it satisfies $\widetilde{F} \circ \pi = F$ by construction. Thus $\widetilde{F}$ is smooth by Proposition 5.19. □

Our third proposition can be interpreted as a uniqueness result for smooth manifolds defined as quotients of other smooth manifolds by submersions.

Proposition 5.21. *Suppose* $\pi_1 \colon M \to N_1$ *and* $\pi_2 \colon M \to N_2$ *are surjective submersions that are constant on each other's fibers (i.e.,* $\pi_1(p) = \pi_1(p')$ *if and only if* $\pi_2(p) = \pi_2(p')$*). Then there exists a unique diffeomorphism* $F \colon N_1 \to N_2$ *such that* $F \circ \pi_1 = \pi_2$*:*

$$\begin{tikzcd} & M \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ N_1 \arrow[rr, "F"'] & & N_2. \end{tikzcd}$$

Exercise 5.5. Prove Proposition 5.21.

Level Sets

Using the tools we have developed so far, we can now give some very general criteria for level sets to be submanifolds.

Theorem 5.22 (Constant Rank Level Set Theorem). *Let M and N be smooth manifolds, and let $F \colon M \to N$ be a smooth map with constant rank k. Each level set of F is a closed embedded submanifold of codimension k in M.*

Proof. Let $c \in N$ be arbitrary, and let S denote the level set $F^{-1}(c) \subset M$. Clearly S is closed in M by continuity. To show it is an embedded submanifold, we need to show that for each $p \in S$ there is a slice chart for S in M near p. From the rank theorem, there are coordinate charts (U, φ) centered at p and (V, ψ) centered at $c = F(p)$ in which F has the coordinate representation (5.7), and therefore $S \cap U$ is the slice $\{(x^1, \ldots, x^n) \in U : x^1 = \cdots = x^k = 0\}$. □

Corollary 5.23 (Submersion Theorem). *If $F \colon M \to N$ is a submersion, then each level set of F is a closed embedded submanifold whose codimension is equal to the dimension of N.*

Proof. A submersion has constant rank equal to the dimension of N. □

This result is extremely useful, because many submanifolds are most naturally presented as level sets of submersions. In fact, it can be strengthened considerably, because we need only check the rank condition on the level set we are interested in. If $F \colon M \to N$ is a smooth map, a point $p \in M$ is said to be a *regular point* of F if $F_* \colon T_pM \to T_{F(p)}N$ is surjective; it is a *critical point* otherwise. (This means, in particular, that every point is critical if $\dim M < \dim N$.) A point $c \in N$ is said to be a *regular value* of F if every point of the level set $F^{-1}(c)$ is a regular point, and a *critical value* otherwise. In particular, if $F^{-1}(c) = \varnothing$, c is regular. Finally, a level set $F^{-1}(c)$ is called a *regular level set* if c is a regular value; in other words, a regular level set is a level set consisting entirely of regular points.

Corollary 5.24 (Regular Level Set Theorem). *Every regular level set of a smooth map is a closed embedded submanifold whose codimension is equal to the dimension of the range.*

Proof. Let $F\colon M \to N$ be a smooth map and let $c \in N$ be a regular value such that $F^{-1}(c) \neq \varnothing$. The fact that c is a regular value means that F_* has rank equal to the dimension of N at every point of $F^{-1}(c)$. To prove the corollary, it suffices to show that the set U of points where $\operatorname{rank} F_* = \dim N$ is open in M, for then we can apply the preceding corollary with M replaced by U.

To see that U is open, let $m = \dim M$, $n = \dim N$, and suppose $p \in U$. Choosing coordinates near p and $F(p)$, the assumption that $\operatorname{rank} F_* = n$ at p means that the $n \times m$ matrix representing F_* in coordinates has an $n \times n$ minor whose determinant is nonzero. By continuity, this determinant will be nonzero in some neighborhood of p, which means that F has rank n in this whole neighborhood. □

Exercise 5.6. If $f\colon M \to \mathbb{R}$ is a smooth real-valued function, show that $p \in M$ is a regular point of f if and only if $df_p \neq 0$.

Example 5.25 (Spheres). Using Corollary 5.24, we can give a much easier proof that $\mathbb{S}^n$ is an embedded submanifold of $\mathbb{R}^{n+1}$. The sphere is easily seen to be a regular level set of the function $f\colon \mathbb{R}^{n+1} \to \mathbb{R}$ given by $f(x) = |x|^2$, since $df = 2\sum_i x^i\, dx^i$ vanishes only at the origin, and thus it is an embedded n-dimensional submanifold of $\mathbb{R}^{n+1}$.

Example 5.26 (The Orthogonal Group). A real $n \times n$ matrix A is said to be *orthogonal* if the linear map $A\colon \mathbb{R}^n \to \mathbb{R}^n$ preserves the Euclidean inner product:

$$(Ax) \cdot (Ay) = x \cdot y.$$

The set $\mathrm{O}(n)$ of all orthogonal $n \times n$ matrices is clearly a subgroup of $\mathrm{GL}(n, \mathbb{R})$, called the n-dimensional *orthogonal group.*

It is easy to see that a matrix A is orthogonal if and only if it takes the standard basis of $\mathbb{R}^n$ to an orthonormal basis, which is equivalent to the columns of A being orthonormal. Since the (i, j)-entry of the matrix $A^T A$ is the dot product of the ith and jth columns of A, this condition is also equivalent to the requirement that $A^T A = I_n$. We will show that $\mathrm{O}(n)$ is an embedded submanifold of $\mathrm{GL}(n, \mathbb{R})$.

Let $\mathrm{S}(n,\mathbb{R})$ denote the set of symmetric $n \times n$ matrices, which is easily seen to be a linear subspace of $\mathrm{M}(n,\mathbb{R})$ of dimension $n(n+1)/2$ because each symmetric matrix is uniquely determined by its values on and above the main diagonal. Define $\Phi\colon \mathrm{GL}(n,\mathbb{R}) \to \mathrm{S}(n,\mathbb{R})$ by

$$\Phi(A) = A^T A.$$

We will show that the identity matrix I_n is a regular value of Φ, from which it follows that $\mathrm{O}(n) = \Phi^{-1}(I_n)$ is an embedded submanifold of codimension $n(n+1)/2$, that is, of dimension $n^2 - n(n+1)/2 = n(n-1)/2$.

Let $A \in \mathrm{O}(n)$ be arbitrary, and let us compute the push-forward $\Phi_*\colon T_A \mathrm{GL}(n,\mathbb{R}) \to T_{\Phi(A)} \mathrm{S}(n,\mathbb{R})$. We can identify the tangent spaces $T_A \mathrm{GL}(n,\mathbb{R})$ and $T_{\Phi(A)} \mathrm{S}(n,\mathbb{R})$ with $\mathrm{M}(n,\mathbb{R})$ and $\mathrm{S}(n,\mathbb{R})$, respectively, because $\mathrm{GL}(n,\mathbb{R})$ is an open subset of the vector space $\mathrm{M}(n,\mathbb{R})$ and $\mathrm{S}(n,\mathbb{R})$ is itself a vector space. For any $B \in \mathrm{M}(n,\mathbb{R})$, the curve $\gamma(t) = A + tB$ satisfies $\gamma(0) = A$ and $\gamma'(0) = B$. We compute

$$\begin{aligned} \Phi_* B = \Phi_* \gamma'(0) = (\Phi \circ \gamma)'(0) &= \left.\frac{d}{dt}\right|_{t=0} \Phi(A + tB) \\ &= \left.\frac{d}{dt}\right|_{t=0} (A+tB)^T(A+tB) = B^T A + A^T B. \end{aligned}$$

If $C \in \mathrm{S}(n,\mathbb{R})$ is arbitrary,

$$\Phi_*(\tfrac{1}{2}AC) = \tfrac{1}{2}C^T A^T A + \tfrac{1}{2}A^T A C = \tfrac{1}{2}C + \tfrac{1}{2}C = C.$$

Thus Φ_* is surjective, which proves the claim.

Example 5.27 (The Special Linear Group). The *special linear group* $\mathrm{SL}(n,\mathbb{R})$ is the set of $n \times n$ matrices with determinant equal to 1. Because the determinant function satisfies $\det(AB) = \det A \det B$, $\mathrm{SL}(n,\mathbb{R})$ is a subgroup of $\mathrm{GL}(n,\mathbb{R})$. We will show that $\det\colon \mathrm{GL}(n,\mathbb{R}) \to \mathbb{R}$ is a smooth submersion, from which it follows that $\mathrm{SL}(n,\mathbb{R}) = \det^{-1}(1)$ is an embedded submanifold. Let $A \in \mathrm{GL}(n,\mathbb{R})$ be arbitrary. Problem 4-10 shows that the differential of det is given by $d(\det)_A(B) = (\det A)\operatorname{tr}(A^{-1}B)$ for $B \in T_A \mathrm{GL}(n,\mathbb{R}) \cong \mathrm{M}(n,\mathbb{R})$. Choosing $B = A$ yields

$$d(\det)_A(A) = (\det A)\operatorname{tr}(A^{-1}A) = (\det A)\operatorname{tr}(I_n) = n \det A \neq 0.$$

This shows that $d(\det)$ never vanishes on $\mathrm{GL}(n,\mathbb{R})$, so det is a submersion and $\mathrm{SL}(n,\mathbb{R})$ is an embedded submanifold of dimension $n^2 - 1$ in $\mathrm{GL}(n,\mathbb{R})$.

Not all embedded submanifolds are naturally given as level sets of submersions as in these examples. However, the next proposition shows that every embedded submanifold is at least locally of this form.

Proposition 5.28. *Let S be a subset of a smooth n-manifold M. Then S is an embedded k-submanifold of M if and only if every point $p \in S$ has a neighborhood U in M such that $U \cap S$ is a level set of a submersion $F\colon U \to \mathbb{R}^{n-k}$.*

Proof. First suppose S is an embedded k-submanifold. If $(x^1, \dots, x^n)$ are slice coordinates for S on an open set $U \subset M$, the map $F\colon U \to \mathbb{R}^{n-k}$ given in coordinates by $F(x) = (x^{k+1}, \dots, x^n)$ is easily seen to be a submersion whose zero level set is $S \cap U$. Conversely, suppose that around every point $p \in S$ there is a neighborhood U and a submersion $F\colon U \to \mathbb{R}^{n-k}$ such that $S \cap U = F^{-1}(c)$ for some $c \in \mathbb{R}^{n-k}$. By the submersion theorem, $S \cap U$ is an embedded submanifold of U, and so by Lemma 5.5, S is itself an embedded submanifold. □

If $S \subset M$ is an embedded submanifold, a smooth map $F\colon M \to N$ such that S is a regular level set of F is called a *defining map* for S. In the special case $N = \mathbb{R}^{n-k}$ (so that F is a real-valued or vector-valued function), it is usually called a *defining function*. Examples 5.25 through 5.27 show that $f(x) = |x|^2$ is a defining function for the sphere, $\Phi(A) = A^T A$ is a defining function for $\mathrm{O}(n)$, and the determinant is a defining function for $\mathrm{SL}(n, \mathbb{R})$. More generally, if U is an open subset of M and $F\colon U \to N$ is a smooth map such that $S \cap U$ is a regular level set of F, then F is called a *local defining map* (or *local defining function* if $N = \mathbb{R}^{n-k}$) for S. Proposition 5.28 says that every embedded submanifold admits a local defining function in a neighborhood of each point.

The next lemma shows that defining maps give a concise characterization of the tangent space to an embedded submanifold.

Lemma 5.29. *Suppose $S \subset M$ is an embedded submanifold. If $F\colon U \to N$ is any local defining map for S, then $T_pS = \operatorname{Ker} F_*\colon T_pM \to T_{F(p)}N$ for each $p \in U$.*

Proof. Recall that we identify T_pS with the subspace $\iota_*(T_pS) \subset T_pM$, where $\iota\colon S \hookrightarrow M$ is the inclusion map. Because $F \circ \iota$ is constant on $S \cap U$, it follows that $F_* \circ \iota_*$ is the zero map from T_pS to $T_{F(p)}N$, and therefore $\operatorname{Im} \iota_* \subset \operatorname{Ker} F_*$. On the other hand, F_* is surjective by the definition of a defining map, so the rank-nullity law implies that

$$\dim \operatorname{Ker} F_* = \dim T_pM - \dim T_{F(p)}N = \dim T_pS = \dim \operatorname{Im} \iota_*,$$

which implies that $\operatorname{Im} \iota_* = \operatorname{Ker} F_*$. □

Our next example is a bit more complicated. It will be of use to us in the next chapter.

Example 5.30 (Matrices of fixed rank). As in Chapter 1, let $\mathrm{M}(m \times n, \mathbb{R})$ denote the mn-dimensional vector space of $m \times n$ real matrices. For any k, let $\mathrm{M}_k(m \times n, \mathbb{R})$ denote the subset of $\mathrm{M}(m \times n, \mathbb{R})$ consisting of matrices of rank k. We showed in Example 1.10 that $\mathrm{M}_k(m \times n, \mathbb{R})$ is an open submanifold of $\mathrm{M}(m \times n, \mathbb{R})$ when $k = \min(m, n)$. Now we will show that when $0 \le k \le \min(m, n)$, $\mathrm{M}_k(m \times n, \mathbb{R})$ is an embedded submanifold of codimension $(m-k)(n-k)$ in $\mathrm{M}(m \times n, \mathbb{R})$.

Let E_0 be an arbitrary $m \times n$ matrix of rank k. This implies that E_0 has some $k \times k$ minor with nonzero determinant. For the time being, let us assume that it is the upper left minor. Writing E_0 in block form as

$$E_0 = \begin{pmatrix} A_0 & B_0 \\ C_0 & D_0 \end{pmatrix},$$

where A_0 is a $k \times k$ matrix and D_0 is of size $(m-k) \times (n-k)$, our assumption is that A_0 is nonsingular.

Let U be the set

$$U = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathrm{M}(m \times n, \mathbb{R}) : \det A \neq 0 \right\}.$$

By continuity of the determinant function, U is an open subset of $\mathrm{M}(m \times n, \mathbb{R})$ containing E_0. Given $E = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in U$, consider the invertible $n \times n$ matrix

$$P = \begin{pmatrix} A^{-1} & -A^{-1}B \\ 0 & I_{n-k} \end{pmatrix}.$$

Since multiplication by an invertible matrix does not change the rank of a matrix, the rank of E is the same as that of

$$EP = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A^{-1} & -A^{-1}B \\ 0 & I_{n-k} \end{pmatrix} = \begin{pmatrix} I_k & 0 \\ CA^{-1} & D - CA^{-1}B \end{pmatrix}. \tag{5.10}$$

Clearly EP has rank k if and only if $D - CA^{-1}B$ is the zero matrix. (To understand where P came from, observe that E has rank k if and only if it can be reduced by elementary column operations to a matrix whose last $n-k$ columns are zero. Since elementary column operations correspond to right multiplication by invertible matrices, it is natural to look for a matrix P satisfying (5.10).)

Thus we are led to define $F\colon U \to \mathrm{M}(m-k \times n-k, \mathbb{R})$ by

$$F \begin{pmatrix} A & B \\ C & D \end{pmatrix} = D - CA^{-1}B.$$

Clearly F is smooth. To show that it is a submersion, we need to show that $DF(E)$ is surjective for each $E \in U$. Since $\mathrm{M}(m-k \times n-k, \mathbb{R})$ is a vector space, tangent vectors at $F(E)$ can be naturally identified with

$(m-k)\times(n-k)$ matrices. Given $E = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ and any matrix $X \in \mathrm{M}(m-k\times n-k,\mathbb{R})$, define a curve $\gamma\colon(-\varepsilon,\varepsilon)\to U$ by

$$\gamma(t) = \begin{pmatrix} A & B \\ C & D+tX \end{pmatrix}.$$

Then

$$F_*\gamma'(0) = (F\circ\gamma)'(t) = \left.\frac{d}{dt}\right|_{t=0}(D+tX-CA^{-1}B) = X.$$

Thus F is a submersion and so $\mathrm{M}_k(m\times n,\mathbb{R})\cap U$ is an embedded submanifold of U.

Now if E_0' is an arbitrary matrix of rank k, just note that it can be transformed to one in U by a rearrangement of its rows and columns. Such a rearrangement is a linear isomorphism $R\colon \mathrm{M}(m\times n,\mathbb{R})\to\mathrm{M}(m\times n,\mathbb{R})$ that preserves rank, so $U' = R^{-1}(U)$ is a neighborhood of E_0' and $F\circ R\colon U'\to\mathrm{M}(m-k\times n-k,\mathbb{R})$ is a submersion whose zero level set is $\mathrm{M}_k(m\times n,\mathbb{R})\cap U'$. Thus every point in $\mathrm{M}_k(m\times n,\mathbb{R})$ has a neighborhood U' in $\mathrm{M}(m\times n,\mathbb{R})$ such that $U'\cap\mathrm{M}_k(m\times n,\mathbb{R})$ is an embedded submanifold of U', so $\mathrm{M}_k(m\times n,\mathbb{R})$ is an embedded submanifold by Lemma 5.5.

Images of Embeddings and Immersions

The next question we want to address is when the image of a smooth map is an embedded submanifold. The most important case is that of an embedding, for which the answer is provided by the following theorem.

Theorem 5.31. *The image of a smooth embedding is an embedded submanifold.*

Proof. Let $F\colon N\to M$ be an embedding. We need to show that each point of $F(N)$ has a coordinate neighborhood $U\subset M$ in which $F(N)\cap U$ is a slice.

Let $p\in N$ be arbitrary, and let (U,φ), (V,ψ) be coordinate charts centered at p and $F(p)$ in which $F|_U\colon U\to V$ has the coordinate representation (5.7). In particular (shrinking U and V if necessary), this implies that $F(U)$ is a slice in V. Because an embedding is a homeomorphism onto its image with the subspace topology, the fact that $F(U)$ is open in $F(N)$ means that there is an open set $W\subset M$ such that $F(U) = W\cap F(N)$. Replacing V by $\widetilde{V} = V\cap W$, we obtain a slice chart $(\widetilde{V},\psi|_{\widetilde{V}})$ containing $F(p)$ such that $\widetilde{V}\cap F(N) = \widetilde{V}\cap F(U)$ is a slice of $\widetilde{V}$. □

The preceding theorem combined with Proposition 5.6 can be summarized by the following corollary.

Corollary 5.32. *Embedded submanifolds are precisely the images of embeddings.*

Immersed Submanifolds

Although embedded submanifolds are the most natural and common submanifolds and suffice for most purposes, it is sometimes important to consider a more general notion of submanifold. In particular, when we study Lie groups later in this chapter and foliations in chapter 14, we will encounter subsets of smooth manifolds that are images of injective immersions, but not necessarily of embeddings. To see what kinds of phenomena occur, look back again at the two examples we introduced earlier of injective immersions that are *not* embeddings. The "figure eight curve" of Example 5.2 and the dense curve on the torus of Example 5.3 are both injective immersions but not embeddings. In fact, their image sets are not embedded submanifolds (see Problems 5-3 and 5-4).

So as to have a convenient language for talking about examples like these, we make the following definition. Let M be a smooth manifold. An *immersed submanifold of dimension k* (or *immersed k-submanifold*) of M is a subset $N \subset M$ endowed with a k-manifold topology (not necessarily the subspace topology) together with a smooth structure such that the inclusion map $\iota\colon N \hookrightarrow M$ is a smooth immersion.

Immersed submanifolds usually arise in the following way. Given an injective immersion $F\colon N \to M$, we can give the image set $F(N) \subset M$ a topology simply by declaring a set $U \subset F(N)$ to be open if and only if $F^{-1}(U) \subset N$ is open. With this topology, $F(N)$ is clearly a topological k-manifold homeomorphic to N, and there is a unique smooth structure on it such that $F\colon N \to F(N)$ is a diffeomorphism. (The smooth coordinate maps are just the maps of the form $\varphi \circ F^{-1}$, where φ is a smooth coordinate map for N.) With this topology and smooth structure, $\iota\colon F(N) \hookrightarrow M$ is clearly a smooth immersion, because it is equal to the composition of a diffeomorphism followed by an immersion:

$$F(N) \xrightarrow{F^{-1}} N \xrightarrow{F} M.$$

Example 5.33 (Immersed Submanifolds). Because the figure eight of Example 5.2 and the dense curve of Example 5.3 are images of injective immersions, they are immersed submanifolds when given appropriate topologies and smooth structures. As smooth manifolds, they are diffeomorphic to $\mathbb{R}$. They are *not* embedded, because neither one has the subspace topology.

Since the inclusion map of an immersed submanifold is by definition an injective immersion, this discussion shows that immersed submanifolds are precisely the images of injective immersions. Clearly every embedded submanifold is also an immersed submanifold. The converse is not true: An immersed submanifold is embedded precisely when it has the subspace topology, or equivalently when the immersion is an embedding.

The following lemma shows that the *local* structure of an immersed submanifold is the same as that of an embedded one.

Lemma 5.34. *Let $F\colon N \to M$ be an immersion. Then F is locally an embedding: For any $p \in N$, there exists a neighborhood U of p in N such that $F|_U\colon U \to M$ is an embedding.*

Exercise 5.7. Prove Lemma 5.34.

It is important to be clear about what this lemma does and does not say. Given an immersed submanifold $N \subset M$ and a point $p \in N$, it is possible to find a neighborhood U of p (in N) such that U is embedded; but it may not be possible to find a neighborhood V of p *in M* such that $V \cap N$ is embedded.

Because immersed submanifolds are the more general of the two types of submanifolds, the term "submanifold" without further qualification means an immersed submanifold, which includes an embedded submanifold as a special case. If there is room for confusion, it is usually better to specify explicitly which type of submanifold is meant, particularly because some authors do not follow this convention, but instead reserve the unqualified term "submanifold" to mean what we call an embedded submanifold.

Even though an immersed submanifold $N \subset M$ is not a topological subspace of M, its tangent space at any point $p \in N$ can nonetheless be viewed as a linear subspace of T_pM, as for an embedded submanifold. If $\iota\colon N \to M$ is the inclusion map, then ι is a smooth immersion, so $\iota_*\colon T_pN \to T_pM$ is injective. Just as in the case of embedded submanifolds, we will routinely identify T_pN with the subspace $\iota_* T_pN \subset T_pM$.

The Case of Manifolds with Boundary

The definitions of this chapter extend easily to manifolds with boundary. First, if M and N are manifolds with boundary, a smooth map $F\colon M \to N$ is said to be a immersion if F_* is injective at each point, a submersion if F_* is surjective at each point, and an embedding if it is an immersion and a homeomorphism onto its image (with the subspace topology. A subset $S \subset M$ is said to be an *immersed submanifold* of M if S is endowed with a smooth manifold structure such that the inclusion map is an immersion, and an *embedded submanifold* if in addition S has the subspace topology. (We do not necessarily require the existence of slice coordinates for embedded submanifolds, because such coordinates can be problematic if S contains boundary points of M.)

More generally, an immersed or embedded submanifold with boundary in M is defined in exactly the same way, except that now S itself is allowed to have a boundary.

Exercise 5.8. If M is a smooth manifold with boundary, show that ∂M is a smooth submanifold of M.

Restricting Maps to Submanifolds

Given a smooth map $F\colon M \to N$, it is important to know whether F is still smooth when its domain or range is restricted to a submanifold. In the case of restricting the domain, the answer is easy.

Proposition 5.35 (Restricting the Domain of a Smooth Map). *If $F\colon M \to N$ is a smooth map and $S \subset M$ is an (immersed or embedded) submanifold, then $F|_S\colon S \to N$ is smooth.*

Proof. The inclusion map $\iota\colon S \hookrightarrow M$ is smooth by definition of an immersed submanifold. Since $F|_S = F \circ \iota$, the result follows. □

When the range is restricted, however, the resulting map may not be smooth, as the following example shows.

Example 5.36. Let $N \subset \mathbb{R}^2$ be the figure eight submanifold, with the topology and smooth structure induced by the immersion γ of Example 5.2. Define a smooth map $G\colon \mathbb{R} \to \mathbb{R}^2$ by

$$G(t) = (\sin 2t, \cos t).$$

(This is the same formula that we used to define γ, but now the domain is the whole real line instead of just a subinterval.) It is easy to check that the image of G lies in N. However, as a map from $\mathbb{R}$ to N, G is not even continuous, because $\gamma^{-1} \circ G$ is not continuous at $t = -\pi/2$.

The next proposition gives sufficient conditions for a map to be smooth when its range is restricted to an immersed submanifold. It shows that the failure of continuity is the only thing that can go wrong.

Proposition 5.37 (Restricting the Range of a Smooth Map). *Let $S \subset N$ be an immersed submanifold, and let $F\colon M \to N$ be a smooth map whose image is contained in S. If F is continuous as a map from M to S, then $F\colon M \to S$ is smooth.*

Proof. Let $p \in M$ be arbitrary and let $q = F(p) \in S$. Because the inclusion map $\iota\colon S \hookrightarrow N$ is an immersion, Lemma 5.34 guarantees that there is a neighborhood V of q in S such that $\iota|_V\colon V \hookrightarrow N$ is an embedding. Thus there exists a slice chart (W, ψ) for V in N centered at q. (Of course, it might not be a slice chart fo S in N.) The fact that (W, ψ) is a slice chart means that $(V_0, \widetilde{\psi})$ is a chart for V, where $V_0 = W \cap V$ and $\widetilde{\psi} = \pi \circ \psi$, with $\pi\colon \mathbb{R}^n \to \mathbb{R}^k$ the projection onto the first $k = \dim S$ coordinates. Since $V_0 = (\iota|_V)^{-1}(W)$ is open in V, it is open in N in its given topology, and so $(V_0, \widetilde{\psi})$ is also a chart for N.

Let $U = F^{-1}(V_0) \subset M$, which is an open set containing p. (Here is where we use the hypothesis that F is continuous.) Choose a coordinate chart

(U_0, φ) for M such that $p \in U_0 \subset U$. Then the coordinate representation of $F\colon M \to S$ with respect to the charts (U_0, φ) and $(V_0, \widetilde{\psi})$ is

$$\widetilde{\psi} \circ F \circ \varphi^{-1} = \pi \circ (\psi \circ F \circ \varphi^{-1}),$$

which is smooth because $F\colon M \to N$ is smooth. □

In the special case in which the submanifold S is embedded, the continuity hypothesis is always satisfied.

Corollary 5.38 (Embedded Case). *Let $S \subset N$ be an embedded submanifold. Then any smooth map $F\colon M \to N$ whose image is contained in S is also smooth as a map from M to S.*

Proof. When $S \subset N$ has the subpace topology, a continuous map $F\colon M \to N$ whose image is contained in S is automatically continuous into S. (This is the characteristic property of the subspace topology—see [Lee00].) □

Vector Fields and Covector Fields on Submanifolds

If $N \subset M$ is an immersed or embedded submanifold, a vector field X on M does not necessarily restrict to a vector field on N, because X_p may not lie in the subspace $T_pN \subset T_pM$ at a point $p \in N$. A vector field X on M is said to be *tangent to* N if $X_p \in T_pN \subset T_pM$ for each $p \in N$.

Lemma 5.39. *Let $N \subset M$ be an immersed or embedded submanifold, and let $\iota\colon N \hookrightarrow M$ denote the inclusion map. If X is a smooth vector field on M that is tangent to N, then there is a unique smooth vector field on N, denoted by $X|_N$, that is ι-related to X.*

Proof. For a vector field $Z \in \mathfrak{T}(N)$ to be ι-related to X means that $\iota_* Z_p = X_p$ for each $p \in N$. Because the injective linear map $\iota_*\colon T_pN \to T_pM$ is just inclusion (under our identification of T_pN with a subspace of T_pM), this means that the uniqe vector field Z that is ι-related to X, if one exists, must be given by $Z_p = X_p$ for all $p \in N$. So we need only check that this defines a smooth vector field on N.

Let p be any point in N. Since an immersed submanifold is locally embedded, there is a neighborhood V of p in N that is embedded in M. Let $(y^1, \dots, y^m)$ be slice coordinates for V in a neighborhood U of p in M, such that $V \cap U$ is the set where $y^{n+1} = \cdots = y^m = 0$. We can write X locally on U as

$$X_q = \sum_{i=1}^{m} X^i(q) \left.\frac{\partial}{\partial y^i}\right|_q, \quad q \in U.$$

At points $q \in V \cap U$, $T_qN = T_qV$ is spanned by $\partial/\partial y^1|_q, \ldots, \partial/\partial y^n|_q$, so the condition that X be tangent to N means that $X^{n+1}(q) = \cdots = X^m(q) = 0$ at any point $q \in V \cap U$. Thus our vector field Z has the local coordinate expression

$$Z_q = \sum_{i=1}^{n} X^i(q) \left.\frac{\partial}{\partial y^i}\right|_q, \quad q \in V \cap U.$$

Since the component functions X^i are restrictions of smooth functions on U, they are smooth, and thus Z is smooth. □

The restriction of covector fields to submanifolds is much simpler. Suppose $N \subset M$ is an immersed submanifold, and let $\iota\colon N \hookrightarrow M$ denote the inclusion map. If σ is any smooth covector field on M, the pullback by ι yields a smooth covector field $\iota^*\sigma$ on N. To see what this means, let $X_p \in T_pN$ be arbitrary, and compute

$$\begin{aligned}(\iota^*\sigma)_p(X_p) &= \sigma_p(\iota_* X_p)\\ &= \sigma_p(X_p),\end{aligned}$$

since $\iota_*\colon T_pN \to T_pM$ is just the inclusion map, under our usual identification of T_pN with a subspace of T_pM. Thus $\iota^*\sigma$ is just the restriction of σ to vectors tangent to N. For this reason we often write $\sigma|_N$ in place of $\iota^*\sigma$, and call it the *restriction* of σ to N. Be warned, however, that $\sigma|_N$ might equal zero at a given point of N, even though *considered as a covector field on M*, σ might not vanish there. An example will help to clarify this distinction.

Example 5.40. Let $\sigma = dy$ on $\mathbb{R}^2$, and let N be the x-axis, considered as a submanifold of $\mathbb{R}^2$. As a covector field on $\mathbb{R}^2$, σ does not vanish at any point, because one of its components is always 1. However, the restriction $\sigma|_N$ is identically zero:

$$\sigma|_N = \iota^* dy = d(y \circ \iota) = 0,$$

because y vanishes identically on N.

To distinguish the two ways in which we might interpret the statement "σ vanishes on N," we will say that σ *vanishes along* N or *vanishes at points of* N if $\sigma_p = 0$ for every point $p \in N$. The weaker condition that $\sigma|_N = 0$ will be expressed by saying that the restriction of σ to N vanishes.

Exercise 5.9. Suppose M is a smooth manifold and $N \subset M$ is an immersed submanifold. If $f \in C^\infty(M)$, show that $d(f|_N) = (df)|_N$. Conclude that if f is constant on N, then the restriction of df to N is zero.

Lie Subgroups

A *Lie subgroup* of a Lie group G is a subgroup of G endowed with a topology and smooth structure making it into a Lie group and an immersed submanifold of G. The following proposition shows that embedded subgroups are automatically Lie subgroups.

Proposition 5.41. *Let G be a Lie group, and suppose $H \subset G$ is a subgroup that is also an embedded submanifold. Then H is a closed Lie subgroup of G.*

Proof. We need only check that multiplication $H \times H \to H$ and inversion $H \to H$ are smooth maps. Because multiplication is a smooth map from $G \times G$ into G, its restriction is clearly smooth from $H \times H$ into G (this is true even if H is merely immersed). Because H is a subgroup, multiplication takes $H \times H$ into H, and since H is embedded, this is a smooth map into H by Corollary 5.38. A similar argument applies to inversion. This proves that H is a Lie subgroup.

To prove that H is closed, suppose that $\{h_i\}$ is a sequence of points in H converging to a point $g \in G$. Let U be the domain of a slice chart for H containing the identity, and let W be a smaller neighborhood of e such that $\overline{W} \subset U$. Since the map $\mu\colon G \times G \to G$ given by $\mu(g_1, g_2) = g_1^{-1} g_2$ is continuous, there is a neighborhood V of the identity with the property that $V \times V \subset \mu^{-1}(W)$, which means that $g_1^{-1} g_2 \in W$ whenever $g_1, g_2 \in V$.

Because $g^{-1} h_i \to e$, by discarding finitely many terms of the sequence we may assume that $g^{-1} h_i \in V$ for all i. This implies that

$$h_j^{-1} h_i = (g^{-1} h_j)^{-1} (g^{-1} h_i) \in W$$

for all i and j. Fixing j and letting $i \to \infty$, we find $h_j^{-1} h_i \to h_j^{-1} g \in \overline{W} \subset U$. Since $H \cap U$ is a slice, it is closed in U, and therefore $h_j^{-1} g \in H$, which implies $g \in H$. Thus H is closed. □

We will see in Chapter 15 that this proposition has an important converse, called the closed subgroup theorem.

Example 5.42 (Lie Subgroups).

(a) The subset $\mathrm{GL}^+(n, \mathbb{R}) \subset \mathrm{GL}(n, R)$ consisting of real $n \times n$ matrices with positive determinant is a subgroup because $\det(AB) = (\det A)(\det B)$. It is an open subset of $\mathrm{GL}(n, \mathbb{R})$ by continuity of the determinant function, and therefore it is an embedded Lie subgroup of dimension n^2.

(b) The circle group $\mathbb{S}^1$ is a Lie subgroup of $\mathbb{C}^*$ because it is a subgroup and an embedded submanifold.

(c) The orthogonal group $\mathrm{O}(n)$ (the group of $n \times n$ orthogonal matrices) is an embedded Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$ (see Example 5.26). It is a compact group because it is a closed and bounded subset of $\mathrm{M}(n, \mathbb{R})$: closed because it is a level set of the continuous map $\Phi(A) = A^T A$, and bounded because each column of an orthogonal matrix has norm 1, which implies that the Euclidean norm of $A \in \mathrm{O}(n)$ is $(\sum_{ij} (A_i^j)^2)^{1/2} = \sqrt{n}$.

(d) The special linear group $\mathrm{SL}(n, \mathbb{R})$ (the set of $n \times n$ real matrices of determinant 1) is a subgroup and an embedded submanifold of codimension 1 in $\mathrm{GL}(n, \mathbb{R})$ by Example 5.27. Therefore it is a Lie subgroup.

(e) The *special orthogonal group* is defined as $\mathrm{SO}(n) = \mathrm{O}(n) \cap \mathrm{SL}(n, \mathbb{R}) \subset \mathrm{GL}(n, \mathbb{R})$. Because every matrix $A \in \mathrm{O}(n)$ satisfies

$$1 = \det I_n = \det(A^T A) = (\det A)(\det A^T) = (\det A)^2,$$

it follows that $\det A = \pm 1$ for all $A \in \mathrm{O}(n)$. Therefore, $\mathrm{SO}(n)$ is the open subgroup of $\mathrm{O}(n)$ consisting of matrices of positive determinant, and is therefore also an embedded Lie subgroup of dimension $n(n-1)/2$ in $\mathrm{GL}(n, \mathbb{R})$. It is a compact group because it is a closed subset of $\mathrm{O}(n)$.

(f) Let $H \subset \mathbb{T}^2$ be the dense immersed submanifold of the torus that is the image of the immersion $\gamma\colon \mathbb{R} \to \mathbb{T}^2$ defined in Example 5.3. It is easy to check that γ is a group homomorphism and therefore H is a subgroup of $\mathbb{T}^2$. Because the smooth structure on H is defined so that $\gamma\colon \mathbb{R} \to H$ is a diffeomorphism, H is a Lie group (in fact, Lie isomorphic to $\mathbb{R}$) and is therefore a Lie subgroup of $\mathbb{T}^2$.

Problems

5-1. Let $F\colon \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$F(x,y) = x^3 + xy + y^3 + 1.$$

Which level sets of F are embedded submanifolds of $\mathbb{R}^2$?

5-2. Define a map $F\colon \mathbb{P}^2 \to \mathbb{R}^4$ by

$$F[x,y,z] = \frac{(x^2 - y^2, xy, xz, yz)}{x^2 + y^2 + z^2}.$$

Show that F is a smooth embedding.

5-3. Show that the image of the curve $\gamma\colon (-\pi/2, 3\pi/2) \to \mathbb{R}^2$ of Example 5.2 is not an embedded submanifold of $\mathbb{R}^2$.

5-4. Let $\gamma\colon \mathbb{R} \to \mathbb{T}^2$ be the curve of Example 5.3.

(a) Show that the image set $\gamma(\mathbb{R})$ is dense in $\mathbb{T}^2$.

(b) Show that $\gamma(\mathbb{R})$ is not an embedded submanifold of the torus.

5-5. Let $F\colon \mathbb{R} \to \mathbb{R}^2$ and $\Phi\colon \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$F(t) = (t^2, t^3),$$
$$\Phi(x,y) = xy.$$

(a) Show that neither $F(\mathbb{R})$ nor $\Phi^{-1}(0)$ is an embedded submanifold of $\mathbb{R}^2$.

(b) Can either set be given a topology and smooth structure making it into an immersed submanifold of $\mathbb{R}^2$? [Hint: Consider tangent vectors to the submanifold at the origin.]

5-6. Show that an embedded submanifold is closed if and only if the inclusion map is proper.

5-7. Suppose M is a smooth manifold, $p \in M$, and $y^1, \dots, y^n$ are smooth real-valued functions defined on a neighborhood of p in M.

(a) If $dy^1_p, \dots, dy^n_p$ form a basis for T_pM, show that $(y^1, \dots, y^n)$ are coordinates for M in some neighborhood of p.

(b) If $dy^1_p, \dots, dy^n_p$ are independent, show that there are functions $y^{n+1}, \dots, y^m$ such that $(y^1, \dots, y^m)$ are coordinates for M in some neighborhood of p.

(c) If $dy^1_p, \dots, dy^n_p$ span T^*_pM, show that there are indices $i_1, \dots, i_k$ such that $(y^{i_1}, \dots, y^{i_k})$ are coordinates for M in some neighborhood of p.

5-8. Let M be a smooth compact manifold. Show that there is no smooth submersion $F\colon M \to \mathbb{R}^k$ for any $k > 0$.

5-9. Suppose $\pi\colon M \to N$ is a smooth map such that every point of M is in the image of a smooth local section of π. Show that π is a submersion.

5-10. Consider the map $F\colon \mathbb{R}^4 \to \mathbb{R}^2$ defined by

$$F(x, y, s, t) = (x^2 + y, x^2 + y^2 + s^2 + t^2 + y).$$

Show that $(0, 1)$ is a regular value of F, and that the level set $F^{-1}(0, 1)$ is diffeomorphic to $\mathbb{S}^2$.

5-11. Let $S \subset \mathbb{R}^2$ be the square of side 2 centered at the origin:

$$S = \{(x, y) : \max(|x|, |y|) = 1\}.$$

If $F\colon \mathbb{S}^1 \to S$ is any homeomorphism, show that either F or F^{-1} is not smooth.

5-12. Show that every bijective bundle map is a bundle isomorphism. More precisely, if E and E' are vector bundles over a smooth manifold M, and $F\colon E \to E'$ is a bijective bundle map, show that F^{-1} is also a bundle map.

5-13. Let $F\colon M \to N$ be a smooth map of constant rank k, and let $S = F(M)$. Show that S can be given a topology and smooth structure such that it is an immersed k-dimensional submanifold of N and $F\colon M \to S$ is smooth. Are the topology and smooth structure uniquely determined by these conditions?

5-14. Decide whether each of the following statements is true or false, and discuss why.

(a) If $F\colon M \to N$ is a smooth map, $c \in N$, and $F^{-1}(c)$ is an embedded submanifold of M whose codimension is equal to the dimension of N, then c is a regular value of F.

(b) If $S \subset M$ is a closed embedded submanifold, there is a smooth map $F\colon M \to P$ such that S is a regular level set of F.

5-15. Let $M \subset N$ be a closed embedded submanifold.

(a) Suppose $f \in C^\infty(M)$. (This means that f is smooth when considered as a function on M, not as a function on a closed subset of N.) Show that f is the restriction of a smooth function on N.

(b) If $X \in \mathfrak{T}(M)$, show that there is a smooth vector field Y on N such that $X = Y|_M$.

(c) Find counterexamples to both results if the hypothesis that M is closed is omitted.

5-16. Let $N \subset M$ be a connected immersed submanifold. Show that a function $f \in C^\infty(M)$ is constant on N if and only if $(df)|_N = 0$.

5-17. If $N \subset M$ is an embedded submanifold and $\gamma\colon J \to M$ is a smooth curve whose image happens to lie in N, show that $\gamma'(t)$ is in the subspace $T_{\gamma(t)}N$ of $T_{\gamma(t)}M$ for all $t \in J$. Give a counterexample if N is not embedded.

5-18. Let M be a smooth manifold. Two embedded submanifolds $N_1, N_2 \subset M$ are said to be *transverse* (or to *intersect transversely*) if for each $p \in N_1 \cap N_2$, the tangent spaces T_pN_1 and T_pN_2 together span T_pM. If N_1 and N_2 are transverse, show that $N_1 \cap N_2$ is either empty or an embedded submanifold of M. Give a counterexample when N_1 and N_2 are not transverse.

5-19. Let M be a smooth n-manifold with boundary. Recall from Chapter 1 that a point $p \in M$ called a *boundary point* of M if $\varphi(p) \in \partial\mathbb{H}^n$ for some generalized chart (U, φ), and an *interior point* if $\varphi(p) \in \operatorname{Int}\mathbb{H}^n$ for some generalized chart. Show that the set of boundary points and the set of interior points are disjoint. [Hint: If $\varphi(p) \in \partial\mathbb{H}^n$ and $\psi(p) \in \operatorname{Int}\mathbb{H}^n$, show that $\varphi \circ \psi^{-1}$ is an open map into $\mathbb{R}^n$ and derive a contradiction.]

5-20. Let M_1, M_2 be connected smooth manifolds of dimension n. For $i = 1, 2$, let (W_i, φ_i) be a coordinate domain centered at some point $p_i \in M_i$ such that $\varphi_i(W_i) = B_2(0) \subset \mathbb{R}^n$. Define $U_i = \varphi_i^{-1}(B_1(0)) \subset W_i$ and $M_i' = M_i \smallsetminus U_i$. The *connected sum* of M_1 and M_2, denoted by $M_1 \# M_2$, is the quotient space of $M_1' \amalg M_2'$ obtained by identifying each $q \in \partial U_1$ with $\varphi_2^{-1} \circ \varphi_1(q) \in \partial U_2$. Show that $M_1 \# M_2$ is connected, and has a unique smooth n-manifold structure such that the restriction of the quotient map to each M_i' is an embedding (where M_i' is thought of as a smooth manifold with boundary). Show that there are open subsets $\widetilde{M}_1, \widetilde{M}_2 \subset M_1 \# M_2$ that are diffeomorphic to $M_1 \smallsetminus \{p_1\}$ and $M_2 \smallsetminus \{p_2\}$, respectively, and such that $\widetilde{M}_1 \cap \widetilde{M}_2$ is diffeomorphic to $B_2(0) \smallsetminus \{0\}$.

6

Embedding and Approximation Theorems

The purpose of this chapter is to address two fundamental questions about smooth manifolds. The questions may seem unrelated at first, but their solutions are closely related.

The first question is "Which smooth manifolds can be smoothly embedded in Euclidean spaces?" The answer, as we will see, is that they all can. This justifies our habit of visualizing manifolds as subsets of $\mathbb{R}^n$.

The second question is "To what extent can continuous maps between manifolds be approximated by smooth ones?" We will give two different answers, both of which are useful in different contexts. Stated simply, we will show that any continuous map from a smooth manifold into $\mathbb{R}^n$ can be uniformly approximated by a smooth map, and that any continuous map from one smooth manifold to another is homotopic to a smooth map.

The essential strategy for answering both questions is the same: first use analysis in $\mathbb{R}^n$ to construct a "local" solution in a fixed coordinate chart; then use partitions of unity to piece together the local solutions into a global one.

Before we begin, we need extend the notion of sets of measure zero to manifolds. These are sets that are "small" in the sense that is closely related to having zero volume (even though we do not yet have a way to measure volume quantitatively on manifolds), and include things like countable sets and submanifolds of lower dimension.

Sets of Measure Zero in Manifolds

Recall what it means for a set $A \subset \mathbb{R}^n$ to have measure zero (see the Appendix): for any $\delta > 0$, A can be covered by a countable collection of open cubes whose total volume is less than δ. The next lemma shows that cubes can be replaced by balls in the definition.

Lemma 6.1. *A subset $A \subset \mathbb{R}^n$ has measure zero if and only if, for every $\delta > 0$, A can be covered by a countable collection of open balls whose total volume is less than δ.*

Proof. This is based on the easily-verified geometric fact that every open cube of volume v is contained in an open ball of volume $c_n v$, and every open ball of volume v is contained in an open cube of volume $c'_n v$, where c_n and c'_n are constants depending only on n. Thus if A has measure zero, there is a countable cover of A by open cubes with total volume less than δ. Enclosing each cube in a ball whose volume is c_n times that of the cube, we obtain an open cover of A by open balls of total volume less than $c_n\delta$, which can be made as small as desired by taking δ sufficiently small. The converse is similar. $\square$

We wish to extend the notion of measure zero in a diffeomorphism-invariant fashion to subsets of manifolds. Because a manifold does not come with a metric, volumes of cubes or balls do not make sense, so we cannot simply use the same definition. However, the key is provided by the next lemma, which implies that the condition of having measure zero is diffeomorphism-invariant for subsets of $\mathbb{R}^n$.

Lemma 6.2. *Suppose $A \subset \mathbb{R}^n$ has measure zero and $F\colon A \to \mathbb{R}^n$ is a smooth map. Then $F(A)$ has measure zero.*

Proof. By definition, F has an extension, still called F, to a smooth function on a neighborhood W of A in $\mathbb{R}^n$. Let $\overline{B}$ be any closed ball contained in W. Since $\overline{B}$ is compact, there is a constant C such that $|DF(x)| \le C$ for all $x \in \overline{B}$. Using the Lipschitz estimate for smooth functions (Proposition A.28), we have

$$|F(x) - F(x')| \le C|x - x'| \tag{6.1}$$

for all $x, x' \in \overline{B}$.

Given $\delta > 0$, we can choose a countable cover $\{B_j\}$ of $A \cap \overline{B}$ by open balls satisfying

$$\sum_j \operatorname{Vol}(B_j) < \delta.$$

Then by (6.1), $F(B_j)$ is contained in a ball $\widetilde{B}_j$ whose radius is no more than C times that of B_j. Since the volume of a ball in $\mathbb{R}^n$ is proportional to the nth power of its radius, we conclude that $F(A \cap \overline{B})$ is contained in the collection of balls $\{\widetilde{B}_j\}$, whose total volume is no greater than

$$\sum_j \operatorname{Vol}(\widetilde{B}_j) < C^n \delta.$$

Since this can be made as small as desired, it follows that $F(A \cap \overline{B})$ has measure zero. Since $F(A)$ is the union of countably many such sets, it too has measure zero. □

Lemma 6.3. *Suppose $F\colon U \to \mathbb{R}^n$ is a smooth map, where U is an open subset of $\mathbb{R}^m$ and $m < n$. Then $F(U)$ has measure zero in $\mathbb{R}^n$.*

Proof. Let $\pi\colon \mathbb{R}^n \to \mathbb{R}^m$ denote the projection onto the first m coordinates, and let $\widetilde{U} = \pi^{-1}(U)$. The result follows by applying the preceding lemma to $\widetilde{F} = F \circ \pi\colon \widetilde{U} \to \mathbb{R}^n$, because $F(U) = \widetilde{F}(\widetilde{U} \cap \mathbb{R}^m)$, which is the image of a set of measure zero. □

We say a subset A of a smooth n-manifold M has measure zero if for every smooth chart (U, φ) for M, the set $\varphi(A \cap U)$ has measure zero in $\mathbb{R}^n$. It follows immediately from Lemma A.30(c), that any set of measure zero has dense complement, because if $M \smallsetminus A$ is not dense then A contains an open set, which would imply $\psi(A \cap V)$ would contain an open set for some coordinate chart (V, ψ).

The following lemma shows that we need only check this condition for a single collection of charts whose domains cover A.

Lemma 6.4. *Suppose A is a subset of a smooth n-manifold M, and for some collection $\{(U_\alpha, \varphi_\alpha)\}$ of charts whose domains cover A, $\varphi_\alpha(A \cap U_\alpha)$ has measure zero in $\mathbb{R}^n$ for each α. Then A has measure zero in M.*

Proof. Let (V, ψ) be an arbitrary coordinate chart. We need to show that $\psi(A \cap V)$ has measure zero. Some countable collection of the U_α's covers $A \cap V$. For each such U_α, we have

$$\psi(A \cap V \cap U_\alpha) = (\psi \circ \varphi_\alpha^{-1}) \circ \varphi_\alpha(A \cap V \cap U_\alpha).$$

Now $\varphi_\alpha(A \cap V \cap U_\alpha)$ is a subset of $\varphi_\alpha(A \cap U_\alpha)$, which has measure zero by hypothesis. By Lemma 6.2 applied to $\psi \circ \varphi_\alpha^{-1}$, therefore, $\psi(A \cap V \cap U_\alpha)$ has measure zero. Since $\psi(A \cap V)$ is the union of countably many such sets, it too has measure zero. □

As our first application of sets of measure zero in manifolds, we prove the following proposition, which is an analogue of Proposition 5.17.

Proposition 6.5. *Let $F\colon M \to N$ be a smooth map of constant rank.*

(*a*) *If F is surjective, then it is a submersion.*

(*b*) *If F is bijective, then it is a diffeomorphism.*

Proof. As in the proof of Proposition 5.17, let $m = \dim M$, $n = \dim N$, and $k = \operatorname{rank} F$. If F is not a submersion, then $k < n$. By the rank theorem, each point has a coordinate neighborhood in which F has the coordinate representation

$$F(x^1, \dots, x^m) = (x^1, \dots, x^k, 0, \dots, 0). \tag{6.2}$$

Since any open cover of a manifold has a countable subcover, we can choose countably many charts $\{(U_i, \varphi_i)\}$ for M and corresponding charts $\{(V_i, \psi_i)\}$ for N such that the sets $\{U_i\}$ cover M, F maps U_i into V_i, and the coordinate representation of $F\colon U_i \to V_i$ is as in (6.2). Since $F(U_i)$ is contained in a k-dimensional slice of V_i, it has measure zero in N. Because $F(M)$ is equal to the countable union of sets $F(U_i)$ of measure zero, $F(M)$ itself has measure zero in N, which implies that F cannot be surjective. This proves (a).

To prove (b), note that a bijective map of constant rank is a submersion by part (a) and an immersion by Proposition 5.17, so M and N have the same dimension. Then Proposition 5.16 implies that F is a diffeomorphism. □

The next theorem is the main result of this section.

Theorem 6.6. *Suppose M and N are smooth manifolds with* $\dim M < \dim N$*, and $F\colon M \to N$ is a smooth map. Then $F(M)$ has measure zero in N. In particular, $N \smallsetminus F(M)$ is dense in N.*

Proof. Write $m = \dim M$ and $n = \dim N$, and let $\{(U_i, \varphi_i)\}$ be a countable covering of M by coordinate charts. Given any coordinate chart (V, ψ) for N, we need to show that $\psi(F(M) \cap V)$ has measure zero in $\mathbb{R}^n$. Observe that this set is the countable union of sets of the form $\psi \circ F \circ \varphi_i^{-1}(\varphi_i(F^{-1}(V) \cap U_i))$, each of which has measure zero by Lemma 6.3. □

Corollary 6.7. *If M is a smooth manifold and $N \subset M$ is an immersed submanifold of positive codimension, then N has measure zero in M.*

Theorem 6.6 can be considered as a special case of the following deeper (and somewhat harder to prove) theorem due to Arthur Sard.

Theorem 6.8 (Sard's Theorem). *If $F\colon M \to N$ is any smooth map, the set of critical values of F has measure zero in N.*

We will neither use nor prove this theorem in this book. For a proof, see [Mil65], [Ste64], or [Bre93].

The Whitney Embedding Theorem

Our first major task in this chapter is to show that every smooth n-manifold can be embedded in $\mathbb{R}^{2n+1}$. We will begin by proving that if $m \geq 2n$, any smooth map into $\mathbb{R}^m$ can be perturbed slightly to be an immersion.

Theorem 6.9. *Let $F\colon M \to \mathbb{R}^m$ be any smooth map, where M is a smooth n-manifold and $m \geq 2n$. For any $\varepsilon > 0$, there is a smooth immersion $\widetilde{F}\colon M \to \mathbb{R}^m$ such that $\sup_M |\widetilde{F} - F| \leq \varepsilon$.*

Proof. Let $\{W_i\}$ be any regular open cover of M as defined in Chapter 2 (for example, a regular refinement of the trivial cover consisting of M alone). Then each W_i is the domain of a chart $\psi_i\colon W_i \to B_3(0)$, and the precompact sets $U_i = \psi_i^{-1}(B_1(0))$ still cover M. For each $k \in \mathbb{N}$, let $M_k = \bigcup_{i=1}^k U_i$. We interpret M_0 to be the empty set. We will modify F inductively on one set W_i at a time.

Let $\{\varphi_i\}$ be a partition of unity subordinate to $\{W_i\}$. Let $F_0 = F$, and suppose by induction we have defined smooth maps $F_j\colon M \to \mathbb{R}^m$ for $j = 1, \dots, k-1$ satisfying

(i) $\sup_M |F_j - F| < \varepsilon$;

(ii) $F_j(x) = F_{j-1}(x)$ unless $x \in W_j$;

(iii) $(F_j)_*$ is injective at each point of $\overline{M}_j$.

For any $m \times n$ matrix A, define a new map $F_A\colon M \to \mathbb{R}^m$ as follows: On $M \smallsetminus \operatorname{supp}\varphi_k$, $F_A = F_{k-1}$; and on W_k, F_A is the map given in coordinates by

$$F_A(x) = F_{k-1}(x) + \varphi_k(x)Ax,$$

where $A\colon \mathbb{R}^n \to \mathbb{R}^m$ is thought of as a linear map. (When computing in W_k, we simplify the notation by identifying maps with their coordinate representations as usual.) Since both definitions agree on the set $W_k \smallsetminus \operatorname{supp}\varphi_k$ where they overlap, this defines a smooth map. We will eventually set $F_k = F_A$ for a suitable choice of A.

Because (i) holds for $j = k-1$, there is a constant $\varepsilon_0 < \varepsilon$ such that $|F_{k-1}(x) - F(x)| \leq \varepsilon_0$ for x in the compact set $\operatorname{supp}\varphi_k$. By continuity, therefore, there is some $\delta > 0$ such that $|A| < \delta$ implies

$$\sup_M |F_A - F_{k-1}| = \sup_{x \in \operatorname{supp}\varphi_k} |\varphi_k(x)Ax| < \varepsilon - \varepsilon_0,$$

and therefore

$$\sup_M |F_A - F| \leq \sup_M |F_A - F_{k-1}| + \sup_M |F_{k-1} - F| < (\varepsilon - \varepsilon_0) + \varepsilon_0 = \varepsilon.$$

Let $P\colon W_k \times \mathrm{M}(m \times n, \mathbb{R}) \to \mathrm{M}(m \times n, \mathbb{R})$ be the matrix-valued function

$$P(x, A) = DF_A(x).$$

By the inductive hypothesis, $P(x, A)$ has rank n when (x, A) is in the compact set $(\operatorname{supp}\varphi_k \cap \overline{M}_{k-1}) \times \{0\}$. By choosing δ even smaller if necessary, we may also ensure that $\operatorname{rank} P(x, A) = n$ whenever $x \in \operatorname{supp}\varphi_k \cap \overline{M}_{k-1}$ and $|A| < \delta$.

The last condition we need to ensure is that $\operatorname{rank}(F_A)_* = n$ on $\overline{U}_k$ and therefore on $\overline{M}_k = \overline{M}_{k-1} \cup \overline{U}_k$. Notice that $DF_A(x) = DF_{k-1}(x) + A$ for $x \in \overline{U}_k$ because $\varphi_k \equiv 1$ there, and therefore $DF_A(x)$ has rank n in $\overline{U}_k$ if and only if A is not of the form $B - DF_{k-1}(x)$ for any $x \in \overline{U}_k$ and any matrix B of rank less than n. To ensure this, let $Q\colon W_k \times \mathrm{M}(m \times n, \mathbb{R}) \to \mathrm{M}(m \times n, \mathbb{R})$ be the smooth map

$$Q(x, B) = B - DF_{k-1}(x).$$

We need to show that there is some matrix A with $|A| < \delta$ that is not of the form $Q(x, B)$ for any $x \in \overline{U}_k$ and any matrix B of rank less than n. For each $j = 0, \ldots, n-1$, the set $\mathrm{M}_j(m \times n, \mathbb{R})$ of $m \times n$ matrices of rank j is an embedded submanifold of $\mathrm{M}(m \times n, \mathbb{R})$ of codimension $(m-j)(n-j)$ by Example 5.30. By Theorem 6.6, therefore, $Q(W_k \times \mathrm{M}_j(m \times n, \mathbb{R}))$ has measure zero in $\mathrm{M}(m \times n, \mathbb{R})$ provided the dimension of $W_k \times \mathrm{M}_j(m \times n, \mathbb{R})$ is strictly less than the dimension of $\mathrm{M}(m \times n, \mathbb{R})$, which is to say

$$n + mn - (m-j)(n-j) < mn$$

or equivalently

$$n - (m-j)(n-j) < 0. \tag{6.3}$$

When $j = n-1$, $n - (m-j)(n-j) = 2n - m - 1$, which is negative because we are assuming $m \geq 2n$. For $j \leq n-1$, $n - (m-j)(n-j)$ is increasing in j because its derivative with respect to j is positive there. Thus (6.3) holds whenever $0 \leq j \leq n-1$. This implies that for each $j = 0, \ldots, n-1$, the image under Q of $W_k \times \mathrm{M}_j(m \times n, \mathbb{R})$ has measure zero in $\mathrm{M}(m \times n, \mathbb{R})$. Choosing A such that $|A| < \delta$ and A is not in the union of these image sets, and setting $F_k = F_A$, we obtain a map satisfying the three conditions of the inductive hypothesis for $j = k$.

Now let $\widetilde{F}(x) = \lim_{k\to\infty} F_k(x)$. By local finiteness of the cover $\{W_j\}$, for each k there is some $N(k) > k$ such that $W_k \cap W_j = \varnothing$ for all $j \geq N(k)$, and then condition (ii) implies that $F_{N(k)} = F_{N(k)+1} = \cdots = F_i$ on W_k for all $i \geq N(k)$. Thus the sequence $\{F_k(x)\}$ is eventually constant for x in a neighborhood of any point, and so $\widetilde{F}\colon M \to \mathbb{R}^m$ is a smooth map. It is an immersion because $\widetilde{F} = F_{N(k)}$ on W_k, which has rank n by (iii). □

Corollary 6.10 (Whitney Immersion Theorem). *Every smooth n-manifold admits an immersion into $\mathbb{R}^{2n}$.*

Proof. Just apply the preceding theorem to any smooth map $F\colon M \to \mathbb{R}^{2n}$, for example a constant map. □

Next we show how to perturb our immersion to be injective. The intuition behind this theorem is that, due to the rank theorem, the image of an immersion looks locally like an n-dimensional affine subspace (after a suitable change of coordinates), so if $F(M) \subset \mathbb{R}^m$ has self-intersections, they will look locally like the intersection between two n-dimensional affine subspaces. If m is at least $2n+1$, such affine subspaces of $\mathbb{R}^m$ can be translated slightly so as to be disjoint, so we might hope to remove the self-intersections by perturbing F a little. The details of the proof are a bit more involved, but the idea is the same.

Theorem 6.11. *Let M be a smooth n-manifold, and suppose $m \geq 2n+1$ and $F\colon M \to \mathbb{R}^m$ is an immersion. Then for any $\varepsilon > 0$ there is an injective immersion $\widetilde{F}\colon M \to \mathbb{R}^m$ such that $\sup_M |\widetilde{F} - F| \leq \varepsilon$.*

Proof. Because an immersion is locally an embedding, there is an open cover $\{W_i\}$ of M such that the restriction of F to each W_i is injective. Passing to a refinement, we may assume that it is a regular cover. As in the proof of the previous theorem, let $\psi_i\colon W_i \to B_3(0)$ be the associated charts, $U_i = \psi_i^{-1}(B_1(0))$, and let $\{\varphi_i\}$ be a partition of unity subordinate to $\{W_i\}$. Let $M_k = \bigcup_{i=1}^k U_k$.

As before, we will modify F inductively to make it injective on successively larger sets. Let $F_0 = F$, and suppose by induction we have defined smooth maps $F_j\colon M \to \mathbb{R}^m$ for $j = 1, \ldots, k-1$ satisfying

(i) F_j is an immersion;

(ii) $\sup_M |F_j - F| < \varepsilon$;

(iii) $F_j(x) = F_{j-1}(x)$ unless $x \in W_j$;

(iv) F_j is injective on $\overline{M}_j$;

(v) F_j is injective on W_i for each i.

Define the next map $F_k\colon M \to \mathbb{R}^m$ by

$$F_k(x) = F_{k-1}(x) + \varphi_k(x)b,$$

where $b \in \mathbb{R}^m$ is to be determined.

We wish to choose b such that $F_k(x) \neq F_k(y)$ when x and y are distinct points of $\overline{M}_k$. To begin, by an argument analogous to that of Theorem 6.9, there exists δ such that $|b| < \delta$ implies

$$\sup_M |F_k - F| \leq \sup_{\operatorname{supp}\varphi_k} |F_k - F_{k-1}| + \sup_M |F_{k-1} - F| < \varepsilon.$$

Choosing δ smaller if necessary, we may also ensure that $(F_k)_*$ is injective at each point of the compact set $\operatorname{supp}\varphi_k$; since $(F_k)_* = (F_{k-1})_*$ is already injective on the rest of M, this implies that F_k is an immersion.

Next, observe that if $F_k(x) = F_k(y)$, then exactly one of the following two cases must hold:

Case I: $\varphi_k(x) \neq \varphi_k(y)$ and

$$b = -\frac{F_{k-1}(x) - F_{k-1}(y)}{\varphi_k(x) - \varphi_k(y)}. \tag{6.4}$$

Case II: $\varphi_k(x) = \varphi_k(y)$ and therefore also $F_{k-1}(x) = F_{k-1}(y)$.

Define an open subset $U \subset M \times M$ by

$$U = \{(x, y) : \varphi_k(x) \neq \varphi_k(y)\},$$

and let $R\colon U \to \mathbb{R}^m$ be the smooth map

$$R(x, y) = -\frac{F_{k-1}(x) - F_{k-1}(y)}{\varphi_k(x) - \varphi_k(y)}.$$

Because $\dim U = \dim(M \times M) = 2n < m$, Theorem 6.6 implies that $R(U)$ has measure zero in $\mathbb{R}^m$. Therefore there exists $b \in \mathbb{R}^m$ with $|b| < \delta$ such that (6.4) does not hold for any $(x, y) \in U$. With this b, (i)–(iii) hold with $j = k$. We need to show that (iv) and (v) hold as well.

If $F_k(x) = F_k(y)$ for some $x, y \in \overline{M}_k$, case I above cannot hold by our choice of b. Therefore we are in case II: $\varphi_k(x) = \varphi_k(y)$ and $F_{k-1}(x) = F_{k-1}(y)$. If $\varphi_k(x) = \varphi_k(y) = 0$, then $x, y \in \overline{M}_k \smallsetminus \overline{U}_k \subset \overline{M}_{k-1}$, contradicting the fact that F_{k-1} is injective on $\overline{M}_{k-1}$ by the inductive hypothesis. On the other hand, if $\varphi_k(x)$ and $\varphi_k(y)$ are nonzero, then $x, y \in \operatorname{supp}\varphi_k \subset W_k$, which contradicts the fact that F_{k-1} is injective on W_k by (v). Similarly, if $F_k(x) = F_k(y)$ for some $x, y \in W_i$, the same argument shows that $F_{k-1}(x) = F_{k-1}(y)$, contradicting (v).

Now we let $\widetilde{F}(x) = \lim_{j\to\infty} F_j(x)$. As before, for any k, this sequence is constant on W_k for j sufficiently large, so defines a smooth function. If $\widetilde{F}(x) = \widetilde{F}(y)$, choose k such that $x, y \in \overline{M}_k$. For sufficiently large j, $\widetilde{F} = F_j$ on $\overline{M}_k$, so the injectivity of F_j on $\overline{M}_k$ implies that $x = y$. □

We can now prove the main result of this section.

Theorem 6.12 (Whitney Embedding Theorem). *Every smooth n-manifold admits an embedding into $\mathbb{R}^{2n+1}$ as a closed submanifold.*

Proof. Let M be a smooth n-manifold. By Proposition 5.4(b), a proper injective immersion is an embedding with closed image. We will begin by constructing a smooth proper map $F_0\colon M \to \mathbb{R}^{2n+1}$ and using the previous two theorems to perturb it to a proper injective immersion.

To construct a proper map, let $\{V_j\}$ be any countable open cover of M by precompact open sets, and let $\{\varphi_j\}$ be a subordinate partition of unity. Define $f \in C^\infty(M)$ by

$$f(p) = \sum_{j=1}^{\infty} j\varphi_j(p).$$

For any positive integer N, if $p \notin \bigcup_{j=1}^{N} \overline{V}_j$, then $\varphi_j(p) = 0$ for $1 \le j \le N$, so

$$|f(p)| = f(p) = \sum_{j=N+1}^{\infty} j\varphi_j(p) > \sum_{j=N+1}^{\infty} N\varphi_j(p) \ge N \sum_{j=1}^{\infty} \varphi_j(p) = N.$$

Therefore $f^{-1}[-N, N]$ is contained in the compact set $\bigcup_{j=1}^{N} \overline{V}_j$. This implies that f is proper and so is the map $F_0 \colon M \to \mathbb{R}^{2n+1}$ defined by $F_0 = (f, 0, \ldots, 0)$.

Now by Theorem 6.9, there is an immersion $F_1 \colon M \to \mathbb{R}^{2n+1}$ satisfying $\sup_M |F_1 - F_0| \le 1$. And by Theorem 6.11, there is an injective immersion $F_2 \colon M \to \mathbb{R}^{2n+1}$ satisfying $\sup_M |F_2 - F_1| \le 1$. If $K \subset \mathbb{R}^{2n+1}$ is any compact set, it is contained in some ball $B_R(0)$, and thus if $F_2(p) \in K$ we have

$$|F_0(p)| \le |F_0(p) - F_1(p)| + |F_1(p) - F_2(p)| + |F_2(p)| \le 1 + 1 + R,$$

which implies $F_2^{-1}(K)$ is a closed subset of $F_0^{-1}(\overline{B_{2+R}(0)})$, which is compact because F_0 is proper. Thus F_2 is a proper injective immersion and hence an embedding. □

This theorem, first proved by Hassler Whitney in 1936 [Whi36], answered a question that had been nagging mathematicians since the notion of an abstract manifold was first introduced: Are there abstract smooth manifolds that are not diffeomorphic to embedded submanifolds of Euclidean space? Although this version of the theorem will be quite sufficient for our purposes, it is interesting to note that eight years later [Whi44b, Whi44a], using much more sophisticated techniques of algebraic topology, Whitney was able to obtain the following improvements.

Theorem 6.13 (Strong Whitney Immersion Theorem). *If $n > 1$, every smooth n-manifold admits an immersion into $\mathbb{R}^{2n-1}$.*

Theorem 6.14 (Strong Whitney Embedding Theorem). *Every smooth n-manifold admits an embedding into $\mathbb{R}^{2n}$.*

The Whitney Approximation Theorem

In this section we prove the two theorems mentioned at the beginning of the chapter on approximation of continuous maps by smooth ones.

We begin with the case of maps into Euclidean spaces. The following theorem shows, in particular, that any continuous map from a smooth manifold M into $\mathbb{R}^k$ can be uniformly approximated by a smooth map. In fact, for later use, we will prove something stronger. If $\delta\colon M \to \mathbb{R}$ is a positive continuous function, we say two maps $F, \widetilde{F}\colon M \to \mathbb{R}^k$ are *δ-close* if $|F(x) - \widetilde{F}(x)| < \delta(x)$ for all $x \in M$.

Theorem 6.15 (Whitney Approximation Theorem). *Let M be a smooth manifold and let $F\colon M \to \mathbb{R}^k$ be a continuous map. Given a positive continuous function $\delta\colon M \to \mathbb{R}$, there exists a smooth map $\widetilde{F}\colon M \to \mathbb{R}^k$ that is δ-close to F. If F is smooth on a closed subset $A \subset M$, then $\widetilde{F}$ can be chosen to be equal to F on A.*

Proof. If F is smooth on the closed set A, then by definition there is some neighborhood U of A on which $F|_A$ has a smooth extension; call this extension F_0. (If there is no such set, we just take $U = A = \varnothing$.) Let

$$U_0 = \{y \in U : |F_0(y) - F(y)| < \delta(y)\}.$$

It is easy to verify that U_0 is an open set containing A.

We will show that there is a countable open cover $\{U_i\}$ of $M \smallsetminus A$ and points $v_i \in \mathbb{R}^k$ such that

$$|F(y) - v_i| < \delta(y) \text{ for all } y \in U_i. \tag{6.5}$$

To see this, for any $x \in M \smallsetminus A$, let U_x be a neighborhood of x contained in $M \smallsetminus A$ and small enough that

$$\begin{aligned} \delta(y) &> \tfrac{1}{2}\delta(x) \text{ and} \\ |F(y) - F(x)| &< \tfrac{1}{2}\delta(x) \end{aligned}$$

for all $y \in U_x$. Then if $y \in U_x$, we have

$$|F(y) - F(x)| < \tfrac{1}{2}\delta(x) < \delta(y).$$

The collection of all such sets U_x as x ranges over points of $M \smallsetminus A$ is an open cover of $M \smallsetminus A$. Choose a countable subcover $\{U_{x_i}\}_{i=1}^{\infty}$. Setting $U_i = U_{x_i}$ and $v_i = F(x_i)$, we have (6.5).

Let $\{\varphi_0, \varphi_i\}$ be a partition of unity subordinate to the cover $\{U_0, U_i\}$ of M, and define $\widetilde{F}\colon M \to \mathbb{R}^k$ by

$$\widetilde{F}(y) = \varphi_0(y)F_0(y) + \sum_{i \geq 1} \varphi_i(y)v_i.$$

Then clearly $\widetilde{F}$ is smooth, and is equal to F on A. For any $y \in M$, the fact that $\sum_{i\geq 0}\varphi_i \equiv 1$ implies that

$$\begin{aligned}
|\widetilde{F}(y) - F(y)| &= \left|\varphi_0(y)F_0(y) + \sum_{i\geq 1}\varphi_i(y)v_i - \left(\varphi_0(y) + \sum_{i\geq 1}\varphi_i(y)\right)F(y)\right| \\
&\leq \varphi_0(y)|F_0(y) - F(y)| + \sum_{i\geq 1}\varphi_i(y)|v_i - F(y)| \\
&< \varphi_0(y)\delta(y) + \sum_{i\geq 1}\varphi_i(y)\delta(y) \\
&= \delta(y),
\end{aligned}$$

which shows that $\widetilde{F}$ is δ-close to F. □

Next we wish to consider a continuous map $F\colon N \to M$ between smooth manifolds. Using the Whitney embedding theorem, we can consider M as an embedded submanifold of some Euclidean space $\mathbb{R}^m$, and approximate F by a smooth map into $\mathbb{R}^m$. However, in general, the image of this smooth map will not lie in M. To correct for this, we need to know that there is a smooth retraction from some neighborhood of M onto M. For this purpose, we introduce a few more definitions.

Let $M \subset \mathbb{R}^m$ be an embedded n-dimensional submanifold. Identifying the tangent space T_pM at a point $p \in M$ with a subspace of $T_p\mathbb{R}^m \cong \mathbb{R}^m$, we define the *normal space* to M at p to be the subspace $N_pM \subset \mathbb{R}^m$ consisting of all vectors that are orthogonal to T_pM with respect to the Euclidean dot product. The *normal bundle* of M is the subset $NM \subset \mathbb{R}^m \times \mathbb{R}^m$ defined by

$$NM = \coprod_{p\in M} N_pM = \{(p,v) \in \mathbb{R}^m \times \mathbb{R}^m : p \in M \text{ and } v \in N_pM\}.$$

Lemma 6.16. *For any embedded submanifold $M \subset \mathbb{R}^m$, the normal bundle NM is an embedded m-dimensional submanifold of $\mathbb{R}^m \times \mathbb{R}^m$.*

Proof. Let $n = \dim M$. Given any point $p \in M$, there exist slice coordinates $(y^1, \dots, y^m)$ on a neighborhood U of p in $\mathbb{R}^m$ such that $U \cap M$ is defined by $y^{n+1} = \dots = y^m = 0$. Define $\Phi\colon U \times \mathbb{R}^m \to \mathbb{R}^{m-n} \times \mathbb{R}^n$ by

$$\Phi(x,v) = \left(y^{n+1}(x), \dots, y^m(x), v \cdot \left.\frac{\partial}{\partial y^1}\right|_x, \dots, v \cdot \left.\frac{\partial}{\partial y^n}\right|_x\right).$$

Then $NM \cap (U \times \mathbb{R}^m) = \Phi^{-1}(0)$ because the vectors $\partial/\partial y^1|_x, \dots, \partial/\partial y^n|_x$ span T_xM at each point $x \in U$. Using the usual change of basis formula for coordinate derivatives, the dot product $v \cdot \partial/\partial y^i|_x$ can be expanded as

$$v \cdot \left.\frac{\partial}{\partial y^i}\right|_x = \left(v^j \frac{\partial}{\partial x^j}\right) \cdot \left(\frac{\partial x^k}{\partial y^i}(y(x)) \left.\frac{\partial}{\partial x^k}\right|_x\right) = \sum_{j=1}^m v^j \frac{\partial x^j}{\partial y^i}(y(x)).$$

(We write the summation explicitly in the last term because the positions of the indices do not conform to the summation convention, as is usual when dealing with the Euclidean dot product.) Thus the Jacobian of Φ at a point $x \in U$ is the $m \times 2m$ matrix

$$D\Phi(x) = \begin{pmatrix} \dfrac{\partial y^j}{\partial x^i}(x) & 0 \\ * & \dfrac{\partial x^j}{\partial y^i}(y(x)) \end{pmatrix}.$$

The m rows of this matrix are obviously independent, so Φ is a submersion and therefore $NM \cap (U \times \mathbb{R}^m)$ is an embedded submanifold. Since the same is true in a neighborhood of each point of NM, the result follows. $\square$

The subset $M \times \{0\} \subset NM$ is clearly diffeomorphic to M. We will identify this subset with M, and thus consider M itself as a subset of NM. The normal bundle comes with a natural projection map $\pi\colon NM \to M$ defined by $\pi(x, v) = x$; it is clearly smooth because it is the restriction of the projection $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ onto the first factor.

Define a map $E\colon NM \to \mathbb{R}^m$ by

$$E(x, v) = x + v.$$

This just maps each normal space N_xM affinely onto the affine subspace through x and orthogonal to T_xM. Clearly E is smooth because it is the restriction of the addition map $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ to NM. A *tubular neighborhood* of M is a neighborhood U of M in $\mathbb{R}^m$ that is the diffeomorphic image under E of an open subset $V \subset NM$ of the form

$$V = \{(x, v) \in NM : |v| < \delta(x)\}, \tag{6.6}$$

for some positive continuous function $\delta\colon M \to \mathbb{R}$.

Theorem 6.17 (Tubular Neighborhood Theorem). *Every embedded submanifold of $\mathbb{R}^m$ has a tubular neighborhood.*

Proof. We begin by showing that E is a diffeomorphism in a neighborhood of each point of $M \subset NM$. Because NM and $\mathbb{R}^m$ have the same dimension, it suffices to show that E_* is surjective at each point. If $v \in T_xM$, there is a smooth curve $\gamma\colon (-\varepsilon, \varepsilon) \to M$ such that $\gamma(0) = x$ and $\gamma'(0) = v$. Let $\widetilde{\gamma}\colon (-\varepsilon, \varepsilon) \to NM$ be the curve $\widetilde{\gamma}(t) = (\gamma(t), 0)$. Then

$$E_*v = (E \circ \widetilde{\gamma})'(0) = \left.\frac{d}{dt}\right|_{t=0} (\gamma(t) + 0) = v.$$

On the other hand, if $w \in N_xM$, then defining $\sigma\colon (-\varepsilon, \varepsilon) \to NM$ by $\sigma(t) = (x, tw)$, we obtain

$$E_*w = (E \circ \sigma)'(0) = \left.\frac{d}{dt}\right|_{t=0} (x + tw) = w.$$

Since T_xM and N_xM span $\mathbb{R}^m$, this shows E_* is surjective. By the inverse function theorem, E is a diffeomorphism on a neighborhood of x in NM, which we can take to be of the form $V_\delta(x) = \{(x', v') : |x - x'| < \delta, |v'| < \delta\}$ for some $\delta > 0$. (This uses the fact that M is embedded and therefore its topology is induced by the Euclidean metric.)

To complete the proof, we need to show that there is an open set V of the form (6.6) on which E is a global diffeomorphism. For each point $x \in M$, let $r(x)$ be the supremum of all δ such that E is a diffeomorphism on $V_\delta(x)$, or $r(x) = 1$ if this supremum is greater than 1. Then $r\colon M \to \mathbb{R}$ is continuous for the following reason. Given $x, x' \in M$, if $|x - x'| < r(x)$, then by the triangle inequality $V_\delta(x')$ is contained in $V_{r(x)}(x)$ for $\delta = r(x) - |x - x'|$, which implies that $r(x') \geq r(x) - |x - x'|$, or $r(x) - r(x') \leq |x - x'|$. The same is true trivially if $|x - x'| \geq r(x)$. Reversing the roles of x and x' yields the opposite inequality, which shows that $|r(x) - r(x')| \leq |x - x'|$, so r is continuous.

Now let $V = \{(x, v) \in NM : |v| < \frac{1}{2}r(x)\}$. We will show that E is injective on V. Suppose that (x, v) and (x', v') are points in V such that $E(x, v) = E(x', v')$. Assume without loss of generality that $r(x') \leq r(x)$. Then $|v'| < \frac{1}{2}r(x') \leq \frac{1}{2}r(x)$, and it follows from $x + v = x' + v'$ that

$$|x - x'| = |v - v'| \leq |v| + |v'| < \tfrac{1}{2}r(x) + \tfrac{1}{2}r(x') \leq r(x).$$

This implies that both (x, v) and (x', v') are in the set $V_{r(x)}(x)$ on which E is injective, so $(x, v) = (x', v')$. Setting $U = E(V)$, we conclude that $E\colon V \to U$ is a smooth bijection and a local diffeomorphism, hence a diffeomorphism by Proposition 5.7. Thus U is a tubular neighborhood of M. □

The principal reason we are interested in tubular neighborhoods is because of the next proposition. Recall that a *retraction* of a topological space X onto a subspace $M \subset X$ is a continuous map $r\colon X \to M$ such that $r|_M$ is the identity map of M.

Proposition 6.18. *Let $M \subset \mathbb{R}^m$ be an embedded submanifold and let U be a tubular neighborhood of M. Then there exists a smooth retraction of U onto M.*

Proof. By definition, there is an open subset $V \subset NM$ containing M such that $E\colon V \to U$ is a diffeomorphism. Just define $r\colon U \to M$ by $r = \pi \circ E^{-1}$, where $\pi\colon NM \to M$ is the natural projection. Clearly r is smooth. For $x \in M$, note that $E(x,0) = x$, so $r(x) = \pi \circ E^{-1}(x) = \pi(x,0) = x$, which shows that r is a retraction. □

The next theorem gives a form of smooth approximation for continuous maps between manifolds. It will have important applications later when we study de Rham cohomology. If $F, G\colon M \to N$ are continuous maps, recall that a *homotopy* from F to G is a continuous map $H\colon M \times I \to N$ (where $I = [0,1]$ is the unit interval) such that

$$H(x,0) = F(x),$$

$$H(x,1) = G(x)$$

for all $x \in M$. If $H(x,s) = F(x) = G(x)$ for all $s \in I$ and all x in some subset $A \subset M$, the homotopy is said to be *relative to* A. If there exists a homotopy from F to G, we say that F and G are *homotopic* (or *homotopic relative to* A if appropriate).

Theorem 6.19 (Whitney Approximation on Manifolds). *Let N and M be smooth manifolds, and let $F\colon N \to M$ be a continuous map. Then F is homotopic to a smooth map $\widetilde{F}\colon N \to M$. If F is smooth on a closed subset $A \subset N$, then the homotopy can be taken to be relative to A.*

Proof. By the Whitney embedding theorem, we may as well assume that M is an embedded submanifold of $\mathbb{R}^m$. Let U be a tubular neighborhood of M in $\mathbb{R}^m$, and let $r\colon U \to M$ be the smooth retraction given by Lemma 6.18. For any $x \in M$, let

$$\delta(x) = \sup\{\varepsilon \le 1 : B_\varepsilon(x) \subset U\}.$$

By a triangle-inequality argument entirely analogous to the one in the proof of the tubular neighborhood theorem, $\delta\colon M \to \mathbb{R}$ is continuous.

Let $\widetilde{\delta} = \delta \circ F\colon N \to \mathbb{R}$. By the Whitney approximation theorem, there exists a smooth map $\widetilde{F}\colon N \to \mathbb{R}^m$ that is a $\widetilde{\delta}$-approximation to F, and is equal to F on A (which might be the empty set). Define a homotopy $H\colon N \times I \to M$ by

$$H(p,t) = r((1-t)F(p) + t\widetilde{F}(p)).$$

This is well defined, because our condition on $\widetilde{F}$ guarantees that for each p, $|\widetilde{F}(p) - F(p)| < \widetilde{\delta}(p) = \delta(F(p))$, which means that $\widetilde{F}(p)$ is contained in the ball of radius $\delta(F(p))$ around $F(p)$; since this ball is contained in U, so is the entire line segment from $F(p)$ to $\widetilde{F}(p)$.

Thus H is a homotopy between $H_0(p) = H(p,0)$ and $H_1(p) = H(p,1)$. It satisfies $H(p,t) = F(p)$ for all $p \in A$, since $F = \widetilde{F}$ there. Clearly $H_0 = F$ from the definition, and $H_1(p) = r(\widetilde{F}(p))$ is smooth. $\square$

If M and N are smooth manifolds, two smooth maps $F, G\colon M \to N$ are said to be *smoothly homotopic* if there is a smooth map $H\colon M \times I \to N$ that is a homotopy between F and G.

Proposition 6.20. *If $F, G\colon M \to N$ are homotopic smooth maps, then they are smoothly homotopic. If F is homotopic to G relative to some closed subset $A \subset M$, then they are smoothly homotopic relative to A.*

Proof. Let $H\colon M \times I \to M$ be a homotopy from F to G (relative to A, which may be empty). We wish to show that H can be replaced by a smooth homotopy.

Because Theorem 6.19 does not apply directly to manifolds with boundary, we first need to extend H to a manifold without boundary containing $M \times I$. Let $J = (-\varepsilon, 1+\varepsilon)$ for some $\varepsilon > 0$, and define $\overline{H}\colon M \times J \to M$ by

$$\overline{H}(x,t) = \begin{cases} H(x,t) & t \in [0,1] \\ H(x,0) & t \leq 0 \\ H(x,1) & t \geq 1. \end{cases}$$

This is continuous by the gluing lemma [Lee00, Lemma 3.8]. Moreover, the restriction of $\overline{H}$ to $M \times \{0\} \cup M \times \{1\}$ is smooth, because it is equal to $F \circ \pi_1$ on $M \times \{0\}$ and $G \circ \pi_1$ on $M \times \{1\}$ (where $\pi_1\colon M \times I \to M$ is the projection on the first factor). If $F \simeq G$ relative to A, $\overline{H}$ is also smooth on $A \times I$. Therefore, Theorem 6.19 implies that there is a smooth map $\widetilde{H}\colon M \times J \to N$ (homotopic to $\overline{H}$, but we do not need that here) whose restriction to $M \times \{0\} \cup M \times \{1\} \cup A \times I$ equals $\overline{H}$ (and therefore H). Restricting back to $M \times I$ again, we see that $\widetilde{H}|_{M \times I}$ is a smooth homotopy (relative to A) between F and G. $\square$

Problems

6-1. Show that any two points in a connected smooth manifold can be joined by a smooth curve segment.

6-2. Let $M \subset \mathbb{R}^m$ be an embedded submanifold, let U be a tubular neighborhood of M, and let $r\colon U \to M$ be the retraction defined in Proposition 6.18. Show that U can be chosen small enough that for each $x \in U$, $r(x)$ is the point in M closest to x. [Hint: First show that each point $x \in U$ has a closest point $y \in M$, and this point satisfies $(x - y) \perp T_y M$.]

6-3. If $M \subset \mathbb{R}^m$ is an embedded submanifold and $\varepsilon > 0$, let M_ε be the set of points in $\mathbb{R}^m$ whose distance from M is less than ε. If M is compact, show that for sufficiently small ε, ∂M_ε is a compact embedded submanifold of $\mathbb{R}^m$, and $\overline{M}_\varepsilon$ is a smooth manifold with boundary.

6-4. Let $M \subset \mathbb{R}^m$ be an embedded submanifold of dimension n. For each $p \in M$, show that there exist a neighborhood U of p in M and smooth maps $X_1, \dots, X_{m-n}\colon U \to \mathbb{R}^m$ such that $(X_1(q), \dots, X_{m-n}(q))$ form an orthonormal basis for $N_q M$ at each point $q \in U$. [Hint: Let (y^i) be slice coordinates and apply the Gram-Schmidt algorithm to the vectors $\partial/\partial y^i$.]

6-5. Let $M \subset \mathbb{R}^m$ be an embedded submanifold, and let NM be its normal bundle. Show that NM is a vector bundle with projection $\pi\colon NM \to M$. [Hint: Use Problem 6-4.]

7
Lie Group Actions

In this chapter, we continue our study of Lie groups. Because their most important applications involve actions by Lie groups on other manifolds, this chapter concentrates on properties of Lie group actions.

We begin by defining Lie group actions on manifolds and explaining some of their main properties. The main result of the chapter is a theorem describing conditions under which the quotient of a smooth manifold by a group action is again a smooth manifold. At the end of the chapter, we explore two classes of such actions in more detail: actions by discrete groups, which are closely connected with covering spaces, and transitive actions, which give rise to homogeneous spaces.

Group Actions on Manifolds

The importance of Lie groups stems primarily from their actions on manifolds. Let G be a Lie group and M a smooth manifold. A *left action* of G on M is a map $G \times M \to M$, often written as $(g, p) \mapsto g \cdot p$, that satisfies

$$\begin{aligned} g_1 \cdot (g_2 \cdot p) &= (g_1 g_2) \cdot p, \\ e \cdot p &= p. \end{aligned} \tag{7.1}$$

A *right action* is defined analogously as a map $M \times G \to M$ with composition working in the reverse order:

$$(p \cdot g_1) \cdot g_2 = p \cdot (g_1 g_2),$$

$$p \cdot e = p.$$

A manifold M endowed with a specific G-action is called a (left or right) *G-space*.

Sometimes it is useful to give a name to an action, such as $\theta\colon G \times M \to M$, with the action of a group element g on a point p usually written $\theta_g(p)$. In terms of this notation, the conditions (7.1) for a left action read

$$\begin{aligned} \theta_{g_1} \circ \theta_{g_2} &= \theta_{g_1 g_2}, \\ \theta_e &= \mathrm{Id}_M, \end{aligned} \tag{7.2}$$

while for a right action the first equation is replaced by

$$\theta_{g_1} \circ \theta_{g_2} = \theta_{g_2 g_1}.$$

For left actions, we will generally use the notations $g \cdot p$ and $\theta_g(p)$ interchangeably. The latter notation contains a bit more information, and is useful when it is important to specify the specific action under consideration, while the former is often more convenient when the action is understood. For right actions, the notation $p \cdot g$ is generally preferred because of the way composition works.

A right action can always be converted to a left action by the trick of defining $g \cdot p$ to be $p \cdot g^{-1}$; thus any results about left actions can be translated into results about right actions, and vice versa. We will usually focus our attention on left actions, because their group law (7.2) has the property that multiplication of group elements corresponds to composition of functions. However, there are some circumstances in which right actions arise naturally; we will see several such actions later in this chapter.

Let us introduce some basic terminology regarding Lie group actions. Let $\theta\colon G \times M \to M$ be a left action of a Lie group G on a smooth manifold M. (The definitions for right actions are analogous.)

- The action is said to be *smooth* if it is smooth as a map from $G \times M$ into M, that is, if $\theta_g(p)$ depends smoothly on (g, p). If this is the case, then for each $g \in G$, the map $\theta_g\colon M \to M$ is a diffeomorphism, with inverse $\theta_{g^{-1}}$.

- For any $p \in M$, the *orbit* of p under the action is the set

  $$G \cdot p = \{g \cdot p : g \in G\},$$

  the set of all images of p under elements of G.

- The action is *transitive* if for any two points $p, q \in M$, there is a group element g such that $g \cdot p = q$, or equivalently if the orbit of any point is all of M.
- Given $p \in M$, the *isotropy group* of p, denoted by G_p, is the set of elements $g \in G$ that fix p:

$$G_p = \{g \in G : g \cdot p = p\}.$$

- The action is said to be *free* if the only element of G that fixes any element of M is the identity: $g \cdot p = p$ for some $p \in M$ implies $g = e$. This is equivalent to the requirement that $G_p = \{e\}$ for every $p \in M$.
- The action is said to be *proper* if the map $G \times M \to M \times M$ given by $(g, p) \mapsto (g \cdot p, p)$ is a proper map (i.e., the preimage of any compact set is compact). (Note that this is *not* the same as requiring that the map $G \times M \to M$ defining the action be a proper map.)

It is not always obvious how to tell whether a given action is proper. The following alternative characterization of proper actions is often useful.

Lemma 7.1. *Suppose a Lie group G acts smoothly on a smooth manifold M. The action is proper if and only if for every compact subset $K \subset M$, the set $G_K = \{g \in G \colon (g \cdot K) \cap K \neq \varnothing\}$ is compact.*

Proof. Let $\Theta \colon G \times M \to M \times M$ denote the map $\Theta(g, p) = (g \cdot p, p)$. Suppose first that Θ is proper. Then for any compact set $K \subset M$, it is easy to check that

$$\begin{aligned} G_K &= \{g \in G : \text{there exists } p \in K \text{ such that } g \cdot p \in K\} \\ &= \{g \in G : \text{there exists } p \in M \text{ such that } \Theta(g, p) \in K \times K\} \\ &= \pi_G\big(\Theta^{-1}(K \times K)\big), \end{aligned}$$

where $\pi_G \colon G \times M \to G$ is the projection. Thus G_K is compact. Conversely, suppose G_K is compact for every compact set $K \subset M$. If $L \subset M \times M$ is compact, let $K = \pi_1(L) \cup \pi_2(L) \subset M$, where $\pi_1, \pi_2 \colon M \times M \to M$ are the projections on the first and second factors, respectively. Then

$$\Theta^{-1}(L) \subset \Theta^{-1}(K \times K) \subset \{(g, p) : g \cdot p \in K \text{ and } p \in K\} \subset G_K \times K.$$

Since $\Theta^{-1}(L)$ is closed by continuity, it is a closed subset of the compact set $G_K \times K$ and is therefore compact. $\square$

One special case in which this condition is automatic is when the group is compact.

Corollary 7.2. *Any smooth action by a compact Lie group on a smooth manifold is proper.*

Proof. Let G be a compact Lie group acting smoothly on M. For any compact set $K \subset M$, the set G_K is closed in G by continuity, and therefore is compact. □

Example 7.3 (Lie group actions).

(a) The *natural action* of $\mathrm{GL}(n,\mathbb{R})$ on $\mathbb{R}^n$ is the left action given by matrix multiplication: $(A,x) \mapsto Ax$, considering $x \in \mathbb{R}^n$ as a column matrix. This is an action because matrix multiplication is associative: $(AB)x = A(Bx)$. It is smooth because the components of Ax depend polynomially on the matrix entries of A and the components of x. Because any nonzero vector can be taken to any other by a linear transformation, there are exactly two orbits: $\{0\}$ and $\mathbb{R}^n \smallsetminus \{0\}$.

(b) The restriction of the natural action to $\mathrm{O}(n) \times \mathbb{R}^n \to \mathbb{R}^n$ defines a smooth left action of $\mathrm{O}(n)$ on $\mathbb{R}^n$. In this case, the orbits are the origin and the spheres centered at the origin. To see why, note that any orthogonal linear transformation preserves norms, so $\mathrm{O}(n)$ takes the sphere of radius R to itself; on the other hand, any vector of length R can be taken to any other by an orthogonal matrix. (If v and v' are such vectors, complete $v/|v|$ and $v'/|v'|$ to orthonormal bases and let A and A' be the orthogonal matrices whose columns are these orthonormal bases; then it is easy to check that $A'A^{-1}$ takes v to v'.)

(c) Further restricting the natural action to $\mathrm{O}(n) \times \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$, we obtain a transitive action of $\mathrm{O}(n)$ on $\mathbb{S}^{n-1}$. It is smooth by Corollary 5.38, because $\mathbb{S}^{n-1}$ is an embedded submanifold of $\mathbb{R}^n$.

(d) The natural action of $\mathrm{O}(n)$ restricts to an action of $\mathrm{SO}(n)$ on $\mathbb{S}^{n-1}$. When $n = 1$, this action is trivial because $\mathrm{SO}(1)$ is the trivial group consisting of the matrix (1) alone. But when $n > 1$, $\mathrm{SO}(n)$ acts transitively on $\mathbb{S}^{n-1}$. To see this, it suffices to show that for any $v \in \mathbb{S}^n$, there is a matrix $A \in \mathrm{SO}(n)$ taking the first standard basis vector e_1 to v. Since $\mathrm{O}(n)$ acts transitively, there is a matrix $A \in \mathrm{O}(n)$ taking e_1 to v. Either $\det A = 1$, in which case $A \in \mathrm{SO}(n)$, or $\det A = -1$, in which case the matrix obtained by multiplying the last column of A by -1 is in $\mathrm{SO}(n)$ and still takes e_1 to v.

(e) Any representation of a Lie group G on a finite-dimensional vector space V is a smooth action of G on V.

(f) Any Lie group G acts smoothly, freely, and transitively on itself by left or right translation. More generally, if H is a Lie subgroup of G, then the restriction of the multiplication map to $H \times G \to G$ defines a smooth, free (but generally not transitive) left action of H on G; similarly, restriction to $G \times H \to G$ defines a free right action of H on G.

(g) An action of a discrete group Γ on a manifold M is smooth if and only if for each $g \in \Gamma$, the map $p \mapsto g \cdot p$ is a smooth map from M to itself. Thus, for example, $\mathbb{Z}^n$ acts smoothly on the left on $\mathbb{R}^n$ by translation:

$$(m^1, \dots, m^n) \cdot (x^1, \dots, x^n) = (m^1 + x^1, \dots, m^n + x^n).$$

Equivariant Maps

Suppose M and N are both (left or right) G-spaces. A smooth map $F\colon M \to N$ is said to be *equivariant* with respect to the given G-actions if for each $g \in G$,

$$F(g \cdot p) = g \cdot F(p) \qquad \text{(for left actions)},$$

$$F(p \cdot g) = F(p) \cdot g \qquad \text{(for right actions)}.$$

Equivalently, if θ and φ are the given actions on M and N, respectively, F is equivariant if the following diagram commutes for each $g \in G$:

$$\begin{tikzcd} M \arrow[r, "F"] \arrow[d, "\theta_g"'] & N \arrow[d, "\varphi_g"] \\ M \arrow[r, "F"'] & N. \end{tikzcd}$$

This condition is also expressed by saying that F *intertwines* the two G-actions.

Example 7.4. Let G and H be Lie groups, and let $F\colon G \to H$ be a Lie homomorphism. There is a natural left action of G on itself by left translation. Define a left action θ of G on H by

$$\theta_g(h) = F(g)h.$$

To check that this is an action, we just observe that $\theta_e(h) = F(e)h = h$, and

$$\theta_{g_1} \circ \theta_{g_2}(h) = F(g_1)(F(g_2)h) = (F(g_1)F(g_2))h = F(g_1g_2)h = \theta_{g_1g_2}(h)$$

because F is a homomorphism. With respect to these G-actions, F is equivariant because

$$\theta_g \circ F(g') = F(g)F(g') = F(gg') = F \circ L_g(g').$$

The following theorem is an extremely useful tool for proving that certain sets are embedded submanifolds.

Theorem 7.5 (Equivariant Rank Theorem). *Let M and N be smooth manifolds and let G be a Lie group. Suppose $F\colon M \to N$ is a smooth map that is equivariant with respect to a transitive smooth G-action on M and any smooth G-action on N. Then F has constant rank. In particular, its level sets are closed embedded submanifolds of M.*

Proof. Let θ and φ denote the G-actions on M and N, respectively, and let p_0 be any point in M. For any other point $p \in M$, choose $g \in G$ such that $\theta_g(p_0) = p$. (Such a g exists because we are assuming G acts transitively on M.) Because $\varphi_g \circ F = F \circ \theta_g$, the following diagram commutes:

$$\begin{tikzcd} T_{p_0}M \arrow[r, "F_*"] \arrow[d, "\theta_{g*}"'] & T_{F(p_0)}N \arrow[d, "\varphi_{g*}"] \\ T_pM \arrow[r, "F_*"'] & T_{F(p)}N. \end{tikzcd}$$

Because the vertical linear maps in this diagram are isomorphisms, the horizontal ones have the same rank. In other words, the rank of F_* at an arbitrary point p is the same as its rank at p_0, so F has constant rank. □

Here are some applications of the equivariant rank theorem.

Proposition 7.6. *Let $F\colon G \to H$ be a Lie group homomorphism. The kernel of F is an embedded Lie subgroup of G, whose codimension is equal to the rank of F.*

Proof. As in Example 7.4, F is equivariant with respect to suitable G-actions on G and H. Since the action on G by left translation is transitive, it follows that F has constant rank, so its kernel $F^{-1}(0)$ is an embedded submanifold. It is thus a Lie subgroup by Proposition 5.41. □

As another application, we describe some important Lie subgroups of $\mathrm{GL}(n,\mathbb{C})$. For any complex matrix A, let A^* denote the *adjoint* or conjugate transpose of A: $A^* = \overline{A}^T$. Observe that $(AB)^* = (\overline{AB})^T = \overline{B}^T\overline{A}^T = B^*A^*$. Consider the following subgroups of $\mathrm{GL}(n,\mathbb{C})$:

- The Complex Special Linear Group:

$$\mathrm{SL}(n,\mathbb{C}) = \{A \in \mathrm{GL}(n,\mathbb{C}) : \det A = 1\}.$$

- The Unitary Group:

$$\mathrm{U}(n) = \{A \in \mathrm{GL}(n,\mathbb{C}) : A^*A = I_n\}.$$

- The Special Unitary Group:

$$\mathrm{SU}(n) = \mathrm{U}(n) \cap \mathrm{SL}(n,\mathbb{C}).$$

Exercise 7.1. Show that $\mathrm{SL}(n,\mathbb{C})$, $\mathrm{U}(n)$, and $\mathrm{SU}(n)$ are subgroups of $\mathrm{GL}(n,\mathbb{C})$ (in the algebraic sense).

Exercise 7.2. Show that a matrix is in $\mathrm{U}(n)$ if and only if its columns form an orthonormal basis for $\mathbb{C}^n$ with respect to the Hermitian dot product $z \cdot w = \sum_i z^i \overline{w^i}$.

Proposition 7.7. *The unitary group* $\mathrm{U}(n)$ *is an embedded* n^2*-dimensional Lie subgroup of* $\mathrm{GL}(n,\mathbb{C})$*.*

Proof. Clearly $\mathrm{U}(n)$ is a level set of the map $\Phi\colon \mathrm{GL}(n,\mathbb{C}) \to \mathrm{M}(n,\mathbb{C})$ defined by

$$\Phi(A) = A^*A.$$

To show that Φ has constant rank and therefore that $\mathrm{U}(n)$ is an embedded Lie subgroup, we will show that Φ is equivariant with respect to suitable right actions of $\mathrm{GL}(n,\mathbb{C})$. Let $\mathrm{GL}(n,\mathbb{C})$ act on itself by right multiplication, and define a right action of $\mathrm{GL}(n,\mathbb{C})$ on $\mathrm{M}(n,\mathbb{C})$ by

$$X \cdot B = B^*XB \qquad \text{for } X \in \mathrm{M}(n,\mathbb{C}),\ B \in \mathrm{GL}(n,\mathbb{C}).$$

It is easy to check that this is a smooth action, and Φ is equivariant because

$$\Phi(AB) = (AB)^*(AB) = B^*A^*AB = B^*\Phi(A)B = \Phi(A) \cdot B.$$

Thus $\mathrm{U}(n)$ is an embedded Lie subgroup of $\mathrm{GL}(n,\mathbb{C})$.

To determine its dimension, we need to compute the rank of Φ. Because the rank is constant, it suffices to compute it at the identity $I_n \in \mathrm{GL}(n,\mathbb{C})$. Thus for any $B \in T_{I_n}\mathrm{GL}(n,\mathbb{C}) = \mathrm{M}(n,\mathbb{C})$, let $\gamma\colon (-\varepsilon,\varepsilon) \to \mathrm{GL}(n,\mathbb{C})$ be the curve $\gamma(t) = I_n + tB$, and compute

$$\begin{aligned}
\Phi_*B &= \left.\frac{d}{dt}\right|_{t=0} \Phi \circ \gamma(t) \\
&= \left.\frac{d}{dt}\right|_{t=0} (I_n + tB)^*(I_n + tB) \\
&= B^* + B.
\end{aligned}$$

The image of this linear map is the set of all *Hermitian* $n \times n$ matrices, i.e., the set of $A \in \mathrm{M}(n,\mathbb{C})$ satisfying $A = A^*$. This is a (real) vector space of dimension n^2, as you can check. Therefore $\mathrm{U}(n)$ is an embedded Lie subgroup of dimension $2n^2 - n^2 = n^2$. □

Proposition 7.8. *The complex special linear group* $\mathrm{SL}(n,\mathbb{C})$ *is an embedded* $(2n^2-2)$*-dimensional Lie subgroup of* $\mathrm{GL}(n,\mathbb{C})$.

Proof. Just note that $\mathrm{SL}(n,\mathbb{C})$ is the kernel of the Lie group homomorphism $\det\colon \mathrm{GL}(n,\mathbb{C}) \to \mathbb{C}^*$. It is easy to check that the determinant is surjective onto $\mathbb{C}^*$, so it is a submersion by Proposition 6.5(a). Therefore $\mathrm{SL}(n,\mathbb{C}) = \mathrm{Ker}(\det)$ is an embedded Lie subgroup whose codimension is equal to $\dim \mathbb{C}^* = 2$. □

Proposition 7.9. *The special unitary group* $\mathrm{SU}(n)$ *is an embedded* (n^2-1)*-dimensional Lie subgroup of* $\mathrm{GL}(n,\mathbb{C})$.

Proof. We will show that $\mathrm{SU}(n)$ is an embedded submanifold of $\mathrm{U}(n)$. Since the composition of embeddings $\mathrm{SU}(n) \hookrightarrow \mathrm{U}(n) \hookrightarrow \mathrm{GL}(n,\mathbb{C})$ is again an embedding, $\mathrm{SU}(n)$ is also embedded in $\mathrm{GL}(n,\mathbb{C})$.

If $A \in \mathrm{U}(n)$, then

$$1 = \det I_n = \det(A^*A) = (\det A)(\det A^*) = (\det A)(\overline{\det A}) = |\det A|^2.$$

Thus $\det\colon \mathrm{U}(n) \to \mathbb{C}^*$ actually takes its values in $\mathbb{S}^1$. It is easy to check that it is surjective onto $\mathbb{S}^1$, so it is a submersion by Proposition 6.5(a). Therefore its kernel $\mathrm{SU}(n)$ is an embedded Lie subgroup of codimension 1 in $\mathrm{U}(n)$. □

Exercise 7.3. Use the techniques developed in this section to give simpler proofs that $\mathrm{O}(n)$ and $\mathrm{SL}(n,\mathbb{R})$ are Lie subgroups of $\mathrm{GL}(n,\mathbb{R})$.

Quotients of Manifolds by Group Actions

Suppose a Lie group G acts on a manifold M (on the left, say). The set of orbits of G in M is denoted by M/G; with the quotient topology, it is called the *orbit space* of the action. Equivalently, M/G is the quotient space of M determined by the equivalence relation $p_1 \sim p_2$ if and only if there exists $g \in G$ such that $g \cdot p_1 = p_2$. It is of great importance to determine conditions under which an orbit space is a smooth manifold.

One simple but important example to keep in mind is the action of $\mathbb{R}^k$ on $\mathbb{R}^k \times \mathbb{R}^n$ by translation in the $\mathbb{R}^k$ factor: $\theta_v(x,y) = (v+x,y)$. The orbits are the affine subspaces parallel to $\mathbb{R}^k$, and the orbit space $(\mathbb{R}^k \times \mathbb{R}^n)/\mathbb{R}^k$ is diffeomorphic to $\mathbb{R}^n$. The quotient map $\pi\colon \mathbb{R}^k \times \mathbb{R}^n \to (\mathbb{R}^k \times \mathbb{R}^n)/\mathbb{R}^k$ is a smooth submersion.

It is worth noting that some authors use distinctive notations such as M/G and $G\backslash M$ to distinguish between orbit spaces determined by left actions and right actions. We will rely on the context, not the notation, to distinguish between the two cases.

The following theorem gives a very general sufficient condition for the quotient of a smooth manifold by a group action to be a smooth manifold.

It is one of the most important applications of the inverse function theorem that we will see.

Theorem 7.10 (Quotient Manifold Theorem). *Suppose a Lie group G acts smoothly, freely, and properly on a smooth manifold M. Then the orbit space M/G is a topological manifold of dimension equal to $\dim M - \dim G$, and has a unique smooth structure with the property that the quotient map $\pi\colon M \to M/G$ is a smooth submersion.*

Proof. First we prove the uniqueness of the smooth structure. Suppose M/G has two different smooth structures such that $\pi\colon M \to M/G$ is a smooth submersion. Let $(M/G)_1$ and $(M/G)_2$ denote M/G with the first and second smooth structures, respectively. By Proposition 5.19, the identity map is smooth from $(M/G)_1$ to $(M/G)_2$:

$$\begin{tikzcd} M \arrow[d, "\pi"'] \arrow[dr, "\pi"] & \\ (M/G)_1 \arrow[r, "\mathrm{Id}"'] & (M/G)_2. \end{tikzcd}$$

The same argument shows that it is also smooth in the opposite direction, so the two smooth structures are identical.

Next we prove that M/G is a topological manifold. Assume for definiteness that G acts on the left, and let $\theta\colon G \times M \to M$ denote the action and $\Theta\colon G \times M \to M \times M$ the proper map $\Theta(g,p) = (g \cdot p, p)$. For any open set $U \subset M$, $\pi^{-1}(\pi(U))$ is equal to the union of all sets of the form $\theta_g(U)$ as g ranges over G. Since θ_g is a diffeomorphism, each such set is open, and therefore $\pi^{-1}(\pi(U))$ is open in M. Because π is a quotient map, this implies that $\pi(U)$ is open in M/G, and therefore π is an open map.

If $\{U_i\}$ is a countable basis for the topology of M, then $\{\pi(U_i)\}$ is a countable collection of open subsets of M/G, and it is easy to check that it is a basis for the topology of M/G. Thus M/G is second countable.

To show that M/G is Hausdorff, define the *orbit relation* $\mathcal{O} \subset M \times M$ by

$$\mathcal{O} = \Theta(G \times M) = \{(g \cdot p, p) \in M \times M : p \in M,\ g \in G\}.$$

(It is called the orbit relation because $(q,p) \in \mathcal{O}$ if and only if p and q are in the same G-orbit.) Since proper maps are closed, it follows that $\mathcal{O}$ is a closed subset of $M \times M$. If $\pi(p)$ and $\pi(q)$ are distinct points in M/G, then p and q lie in distinct orbits, so $(p,q) \notin \mathcal{O}$. If $U \times V$ is a product neighborhood of (p,q) in $M \times M$ that is disjoint from $\mathcal{O}$, then $\pi(U)$ and $\pi(V)$ are disjoint open subsets of M/G containing $\pi(p)$ and $\pi(q)$, respectively. Thus M/G is Hausdorff.

Before proving that M/G is locally Euclidean, we will show that the G-orbits are embedded submanifolds of M diffeomorphic to G. For any

$p \in M$, define a smooth map $\theta^{(p)}\colon G \to M$ by $\theta^{(p)}(g) = g \cdot p$. Note that the image of $\theta^{(p)}$ is exactly the orbit of p. We will show that $\theta^{(p)}$ is an embedding. First, if $\theta^{(p)}(g') = \theta^{(p)}(g)$, then $g' \cdot p = g \cdot p$, which implies $(g^{-1}g') \cdot p = p$. Since we are assuming G acts freely on M, this can only happen if $g^{-1}g' = e$, which means $g = g'$; thus $\theta^{(p)}$ is injective. Observe that

$$\theta^{(p)}(g'g) = (g'g) \cdot p = g' \cdot (g \cdot p) = g' \cdot \theta^{(p)}(g),$$

so $\theta^{(p)}$ is equivariant with respect to left translation on G and the given action on M. Since G acts transitively on itself, this implies that $\theta^{(p)}$ has constant rank. Since it is also injective, it is an immersion by Proposition 5.17.

If $K \subset M$ is a compact set, then $(\theta^{(p)})^{-1}(K)$ is closed in G by continuity, and since it is contained in $G_K = \{g \in G : (g \cdot K) \cap K \neq \varnothing\}$, it is compact by Lemma 7.1. Therefore, $\theta^{(p)}$ is a proper map. We have shown that $\theta^{(p)}$ is a proper injective immersion, so it is an embedding by Proposition 5.4(b).

Let $k = \dim G$ and $n = \dim M - \dim G$. Let us say that a coordinate chart (U, φ) on M, with coordinate functions $(x, y) = (x^1, \dots, x^k, y^1, \dots, y^n)$, is *adapted* to the G-action if

(i) $\varphi(U)$ is a product open set $U_1 \times U_2 \subset \mathbb{R}^k \times \mathbb{R}^n$, and

(ii) each orbit intersects U either in the empty set or in a single slice of the form $\{y^1 = c^1, \dots, y^n = c^n\}$.

We will show that for any $p \in M$, there exists an adapted coordinate chart centered at p. To prove this, we begin by choosing any slice chart (W, ψ) centered at p for the orbit $G \cdot p$ in M. Write the coordinate functions of ψ as $(u^1, \dots, u^k, v^1, \dots, v^n)$, so that $(G \cdot p) \cap U$ is the slice $\{v^1 = \dots = v^n = 0\}$. Let S be the submanifold of U defined by $u^1 = \dots = u^k = 0$. (This is the slice "perpendicular" to the orbit in these coordinates.) Thus T_pM decomposes as the following direct sum:

$$T_pM = T_p(G \cdot p) \oplus T_pS,$$

where $T_p(G \cdot p)$ is the span of $(\partial/\partial u^i)$ and T_pS is the span of $(\partial/\partial v^i)$.

Let $\psi\colon G \times S \to M$ denote the restriction of the action θ to $G \times S \subset G \times M$. We will use the inverse function theorem to show that ψ is a diffeomorphism in a neighborhood of $(e, p) \in G \times S$. Let $i_p\colon G \to G \times S$ be the embedding given by $i_p(g) = (g, p)$. The orbit map $\theta^{(p)}\colon G \to M$ is equal to the composition

$$G \xrightarrow{i_p} G \times S \xrightarrow{\psi} M.$$

Since $\theta^{(p)}$ is an embedding whose image is the orbit $G \cdot p$, it follows that $\theta^{(p)}_*(T_eG)$ is equal to the subspace $T_p(G \cdot p) \subset T_pM$, and thus the image of

$\psi_*\colon T_{(e,p)}(G \times S) \to T_pM$ contains $T_p(G \cdot p)$. Similarly, if $j_e\colon S \to G \times S$ is the embedding $j_e(q) = (e,q)$, then the inclusion $\iota\colon S \hookrightarrow M$ is equal to the composition

$$S \xrightarrow{j_e} G \times S \xrightarrow{\psi} M.$$

Therefore, the image of ψ_* also includes $T_pS \subset T_pM$. Since $T_p(G \cdot p)$ and T_pS together span T_pM, $\psi_*\colon T_{(e,p)}(G \times S) \to T_pM$ is surjective, and for dimensional reasons, it is bijective. By the inverse function theorem, there exist a neighborhood (which we may assume to be a product neighborhood) $X \times Y$ of (e,p) in $G \times S$ and a neighborhood U of p in M such that $\psi\colon X \times Y \to U$ is a diffeomorphism. Shrinking X and Y if necessary, we may assume that X and Y are precompact sets that are diffeomorphic to Euclidean balls in $\mathbb{R}^k$ and $\mathbb{R}^n$, respectively.

We need to show that Y can be chosen small enough that each G-orbit intersects Y in at most a single point. Suppose this is not true. Then if $\{Y_i\}$ is a countable neighborhood basis for Y at p (e.g., a sequence of Euclidean balls whose diameters decrease to 0), for each i there exist distinct points $p_i, p_i' \in Y_i$ that are in the same orbit, which is to say that $g_i \cdot p_i = p_i'$ for some $g_i \in G$. Now, the points $\Theta(g_i, p_i) = (g_i \cdot p_i, p_i) = (p_i', p_i)$ all lie in the compact set $\overline{Y} \times \overline{Y}$, so by properness of Θ, their inverse images (g_i, p_i) must lie in a compact set $L \subset G \times M$. Thus the points g_i all lie in the compact set $\pi_G(L) \subset G$. Passing to a subsequence, we may assume that $g_i \to g \in G$. Note also that both sequences $\{p_i\}$ and $\{p_i'\}$ converge to p, since $p_i, p_i' \in Y_i$ and $\{Y_i\}$ is a neighborhood basis at p. By continuity, therefore,

$$g \cdot p = \lim_{i\to\infty} g_i \cdot p_i = \lim_{i\to\infty} p_i' = p.$$

Since G acts freely, this implies $g = e$. When i gets large enough, therefore, $g_i \in X$. But this contradicts the fact that θ is injective on $X \times Y$, because

$$\theta_{g_i}(p_i) = p_i' = \theta_e(p_i'),$$

and we are assuming $p_i \neq p_i'$.

Choose diffeomorphisms $\alpha\colon \mathbb{B}^k \to X$ and $\beta\colon \mathbb{B}^n \to Y$ (where $\mathbb{B}^k$ and $\mathbb{B}^n$ are the open unit balls in $\mathbb{R}^k$ and $\mathbb{R}^n$, respectively), and define $\gamma\colon \mathbb{B}^k \times \mathbb{B}^n \to U$ by $\gamma(x,y) = \theta_{\alpha(x)}(\beta(y))$. Because γ is equal to the composition of diffeomorphisms

$$\mathbb{B}^k \times \mathbb{B}^n \xrightarrow{\alpha\times\beta} X \times Y \xrightarrow{\psi} U,$$

γ is a diffeomorphism. The map $\varphi = \gamma^{-1}$ is therefore a coordinate map on U. We will show that φ is adapted to the G-action. Condition (i) is obvious from the definition. Observe that each $y =$ constant slice is contained in a single orbit, because it is of the form $\theta(X \times \{p_0\}) \subset \theta(G \times \{p_0\}) = G \cdot p_0$, where $p_0 \in Y$ is the point whose y-coordinate is the given constant. Thus if an arbitrary orbit intersects U, it does so in a union of $y =$ constant slices. However, since an orbit can intersect Y at most once, and each $y =$ constant slice has a point in Y, it follows that each orbit intersects U in precisely one slice if at all. This completes the proof that adapted coordinate charts exist.

To finish the proof that M/G is locally Euclidean, let $q = \pi(p)$ be an arbitrary point of M/G, and let (U, φ) be an adapted coordinate chart centered at p, with $\varphi(U) = U_1 \times U_2 \subset \mathbb{R}^k \times \mathbb{R}^m$. Let $V = \pi(U)$, which is an open subset of M/G because π is an open map. Writing the coordinate functions of φ as $(x^1, \dots, x^k, y^1, \dots, y^n)$ as before, let $Y \subset U$ be the slice $\{x^1 = \cdots = x^k = 0\}$. Note that $\pi \colon Y \to V$ is bijective by the definition of an adapted chart. Moreover, if W is an open subset of Y, then

$$\pi(W) = \pi\big(\{(x, y) : (0, y) \in W\}\big)$$

is open in M/G, and thus $\pi|_Y$ is a homeomorphism. Let $\sigma = (\pi|_Y)^{-1} \colon V \to Y \subset U$, which is a local section of π.

Define a map $\eta \colon V \to U_2$ by sending the equivalence class of a point (x, y) to y; this is well defined by the definition of an adapted chart. Formally, $\eta = \pi_2 \circ \varphi \circ \sigma$, where $\pi_2 \colon U_1 \times U_2 \to U_2 \subset \mathbb{R}^n$ is the projection onto the second factor. Because σ is a homeomorphism from V to Y and $\pi_2 \circ \varphi$ is a homeomorphism from Y to U_2, it follows that η is a homeomorphism. This completes the proof that M/G is a topological n-manifold.

Finally, we need to show that M/G has a smooth structure such that π is a submersion. We will use the atlas consisting of all charts (V, η) as constructed in the preceding paragraph. With respect to any such chart for M/G and the corresponding adapted chart for M, π has the coordinate representation $\pi(x, y) = y$, which is certainly a submersion. Thus we need only show that any two such charts for M/G are smoothly compatible.

Let (U, φ) and $(\widetilde{U}, \widetilde{\varphi})$ be two adapted charts for M, and let (V, η) and $(\widetilde{V}, \widetilde{\eta})$ be the corresponding charts for M/G. First consider the case in which the two adapted charts are both centered at the same point $p \in M$. Writing the adapted coordinates as (x, y) and $(\widetilde{x}, \widetilde{y})$, the fact that the coordinates are adapted to the G-action means that two points with the same y-coordinate are in the same orbit, and therefore also have the same $\widetilde{y}$-coordinate. This means that the transition map between these coordinates can be written $(\widetilde{x}, \widetilde{y}) = (A(x, y), B(y))$, where A and B are smooth functions defined on some neighborhood of the origin. The transition map $\widetilde{\eta} \circ \eta^{-1}$ is just $\widetilde{y} = B(y)$, which is clearly smooth.

In the general case, suppose (U, φ) and $(\widetilde{U}, \widetilde{\varphi})$ are adapted charts for M, and $p \in U$, $\widetilde{p} \in \widetilde{U}$ are points such that $\pi(p) = \pi(\widetilde{p}) = q$. Modifying both charts by adding constant vectors, we can assume that they are centered at p and $\widetilde{p}$, respectively. Since p and $\widetilde{p}$ are in the same orbit, there is a group element g such that $g \cdot p = \widetilde{p}$. Because θ_g is a diffeomorphism taking orbits to orbits, it follows that $\widetilde{\varphi}' = \widetilde{\varphi} \circ \theta_g$ is another adapted chart centered at p. Moreover, $\widetilde{\sigma}' = \theta_g^{-1} \circ \widetilde{\sigma}$ is the local section corresponding to $\widetilde{\varphi}'$, and therefore $\widetilde{\eta}' = \pi_2 \circ \widetilde{\varphi}' \circ \widetilde{\sigma}' = \pi_2 \circ \widetilde{\varphi} \circ \theta_g \circ \theta_g^{-1} \circ \widetilde{\sigma} = \pi_2 \circ \widetilde{\varphi} \circ \widetilde{\sigma} = \widetilde{\eta}$. Thus we are back in the situation of the preceding paragraph, and the two charts are smoothly compatible. □

Covering Manifolds

Proposition 2.8 showed that any covering space of a smooth manifold is again a smooth manifold. It is often important to know when a space *covered by* a smooth manifold is itself a smooth manifold. To understand the answer to this question, we need to study the covering group of a covering space. In this section, we assume knowledge of the basic properties of topological covering maps, as developed for example in [Lee00, Chapters 11 and 12].

Let $\widetilde{M}$ and M be topological spaces, and let $\pi\colon \widetilde{M} \to M$ be a covering map. A *covering transformation* of π is a homeomorphism $\varphi\colon \widetilde{M} \to \widetilde{M}$ such that $\pi \circ \varphi = \pi$:

$$\begin{tikzcd} \widetilde{M} \arrow[rr, "\varphi"] \arrow[dr, "\pi"'] & & \widetilde{M} \arrow[dl, "\pi"] \\ & M. & \end{tikzcd}$$

The set $\mathcal{C}_\pi(\widetilde{M})$ of all covering transformations, called the *covering group* of π, is a group under composition, acting on $\widetilde{M}$ on the left. The covering group is the key to constructing smooth manifolds covered by $\widetilde{M}$.

We will see below that for a smooth covering $\pi\colon \widetilde{M} \to M$, the covering group acts smoothly, freely, and properly on the covering space $\widetilde{M}$. Before proceeding, it is useful to have an alternative characterization of properness for actions of discrete groups.

Lemma 7.11. *Suppose a discrete group Γ acts continuously on a topological manifold $\widetilde{M}$. The action is proper if and only if the following condition holds:*

$$\text{Any two points } p, p' \in \widetilde{M} \text{ have neighborhoods } U, U' \text{ such that the set } \{g \in \Gamma : (g \cdot U) \cap U' \neq \varnothing\} \text{ is finite.} \tag{7.3}$$

Proof. First suppose that Γ acts properly, and let $\Theta\colon \Gamma \times \widetilde{M} \to \widetilde{M} \times \widetilde{M}$ be the map $\Theta(g,p) = (g \cdot p, p)$. Let U, U' be precompact neighborhoods of p and p', respectively. If (7.3) does not hold, then there exist infinitely many distinct elements $g_i \in \Gamma$ and points $p_i \in U$ such that $g_i \cdot p_i \in U'$. Because the pairs $(g_i \cdot p_i, p_i) = \Theta(g_i, p_i)$ lie in the compact set $\overline{U'} \times \overline{U}$, the preimages (g_i, p_i) lie in a compact subset of $\Gamma \times \widetilde{M}$, and therefore have a convergent subsequence. But this is impossible, because $\{g_i\}$ is an infinite sequence of distinct points in a discrete space.

Conversely, suppose (7.3) holds. If L is any compact subset of $\widetilde{M} \times \widetilde{M}$, we need to show that $\Theta^{-1}(L) \subset \Gamma \times \widetilde{M}$ is compact. It suffices to show that any sequence $\{(g_i, p_i)\} \subset \Theta^{-1}(L)$ has a convergent subsequence. Thus suppose $\Theta(g_i, p_i) = (g_i \cdot p_i, p_i) \in L$ for all i. By compactness of L, we can replace this sequence by a subsequence such that $p_i \to p$ and $g_i \cdot p_i \to p'$. Let U, U' be neighborhoods of p and p', respectively, satisfying property (7.3). For all sufficiently large i, $p_i \in U$ and $g_i \cdot p_i \in U'$. Since there are only finitely many $g \in \Gamma$ for which $(g \cdot U) \cap U' \neq \varnothing$, this means that there is some $g \in \Gamma$ such that $g_i = g$ for infinitely many i; in particular, some subsequence of (g_i, p_i) converges. □

Exercise 7.4. Suppose Γ is a discrete group acting continuously on a topological manifold $\widetilde{M}$. Show that the action is proper if and only if both of the following conditions are satisfied:

(i) Each $p \in \widetilde{M}$ has a neighborhood U such that $(g \cdot U) \cap U = \varnothing$ for all but finitely many $g \in \Gamma$.

(ii) If $p, p' \in \widetilde{M}$ are not in the same Γ-orbit, there exist neighborhoods U of p and U' of p' such that $(g \cdot U) \cap U' = \varnothing$ for all $g \in \Gamma$.

A continuous discrete group action satisfying conditions (i) and (ii) of the preceding exercise (or condition (7.3) of Lemma 7.11, or something closely related to these) has traditionally been called *properly discontinuous*. Because the term "properly discontinuous" is self-contradictory (properly discontinuous group actions are, after all, continuous!), and because there is no general agreement about exactly what the term should mean, we will avoid using this terminology and stick with the more general term "proper action" in this book.

Proposition 7.12. *Let $\pi \colon \widetilde{M} \to M$ be a smooth covering map. With the discrete topology, the covering group $\mathcal{C}_\pi(\widetilde{M})$ is a zero-dimensional Lie group acting smoothly, freely, and properly on $\widetilde{M}$.*

Proof. To show that $\mathcal{C}_\pi(\widetilde{M})$ is a Lie group, we need only verify that it is countable. Let $\widetilde{p} \in \widetilde{M}$ be arbitrary and let $p = \pi(\widetilde{p})$. Because the fiber $\pi^{-1}(p)$ is a discrete subset of the manifold $\widetilde{M}$, it is countable. Since each element of $\mathcal{C}_\pi(\widetilde{M})$ is uniquely determined by what it does to $\widetilde{p}$ [Lee00, Proposition 11.27(a)], the map $\varphi \mapsto \varphi(\widetilde{p})$ is an injection of Γ into $\pi^{-1}(p)$; thus Γ is countable.

Smoothness of the action follows from the fact that any covering transformation φ can be written locally as $\varphi = \sigma \circ \pi$ for a suitable smooth local section σ. The action is free because the only covering transformation that fixes any point is the identity.

To show that the action is proper, we will show that it satisfies conditions (i) and (ii) of Exercise 7.4. If $p \in \widetilde{M}$, let U be an evenly covered neighborhood of $\pi(p)$, and let $\widetilde{U}$ be the component of $\pi^{-1}(U)$ containing p. Because each element of the covering group permutes the components of $\pi^{-1}(U)$ [Lee00, Proposition 11.27(d)], it follows that $\widetilde{U}$ satisfies (i). (In fact, it satisfies the stronger condition that $(g \cdot U) \cap U = \varnothing$ for all $g \in G$ except $g = e$.)

Let $p, p' \in \widetilde{M}$ be points in separate orbits. If $\pi(p) \neq \pi(p')$, then there are disjoint open sets U containing $\pi(p)$ and U' containing $\pi(p')$, so $\pi^{-1}(U), \pi^{-1}(U')$ are disjoint open sets satisfying (ii). If $\pi(p) = \pi(p')$, let U be an evenly covered neighborhood of $\pi(p)$, and let $\widetilde{U}$, $\widetilde{U}'$ be the components of $\pi^{-1}(U)$ containing p and p', respectively. If φ is a covering transformation such that $\varphi(\widetilde{U}) \cap \widetilde{U}' \neq \varnothing$, then $\varphi(\widetilde{U}) = \widetilde{U}'$ because covering transformations permute the components of $\pi^{-1}(U)$; therefore, since each component contains exactly one point of $\pi^{-1}(p)$, it follows that $\varphi(p) = p'$, which contradicts the assumption that p and p' are in different orbits. Thus $\widetilde{U}$ and $\widetilde{U}'$ satisfy (ii). □

The quotient manifold theorem yields the following converse to this proposition.

Theorem 7.13. *Suppose $\widetilde{M}$ is a connected smooth manifold, and a discrete Lie group Γ acts smoothly, freely, and properly on $\widetilde{M}$. Then $\widetilde{M}/\Gamma$ is a topological manifold and has a unique smooth structure such that $\pi\colon \widetilde{M} \to \widetilde{M}/\Gamma$ is a smooth covering map.*

Proof. It follows from the quotient manifold theorem that $\widetilde{M}/\Gamma$ has a unique smooth manifold structure such that π is a smooth submersion. Because $\dim \widetilde{M}/\Gamma = \dim \widetilde{M} - \dim \Gamma = \dim \widetilde{M}$, this implies that π is a local diffeomorphism. On the other hand, it follows from the theory of covering spaces [Lee00, Corollary 12.12] that π is a topological covering map. Thus π is a smooth covering map. Uniqueness of the smooth structure follows from the uniqueness assertion of the quotient manifold theorem, because a smooth covering map is in particular a submersion. □

Example 7.14 (Proper Discrete Group Actions).

(a) The discrete Lie group $\mathbb{Z}^n$ acts smoothly and freely on $\mathbb{R}^n$ by translation (Example 7.3(g)). To check that the action is proper, one can verify that condition (7.3) is satisfied by sufficiently small balls around p and p'. The quotient manifold $\mathbb{R}^n/\mathbb{Z}^n$ is homeomorphic to the n-torus $\mathbb{T}^n$, and Theorem 7.13 says that there is a unique smooth structure on $\mathbb{T}^n$ making the quotient map into a smooth covering map. To verify that this smooth structure on $\mathbb{T}^n$ is the same as the one we defined previously (thinking of $\mathbb{T}^n$ as the product manifold $\mathbb{S}^1 \times \cdots \times \mathbb{S}^1$), we just check that the covering map $\mathbb{R}^n \to \mathbb{T}^n$ given by $(x^1, \ldots, x^n) \mapsto (e^{2\pi i x^1}, \ldots, e^{2\pi i x^n})$ is a local diffeomorphism with respect to the product smooth structure on $\mathbb{T}^n$, and apply the result of Proposition 5.21.

(b) The two-element group $\{\pm 1\}$ acts on $\mathbb{S}^n$ by multiplication. This action is obviously smooth and free, and it is proper because the group is compact. This defines a smooth structure on $\mathbb{S}^n/\{\pm 1\}$. In fact, this quotient manifold is diffeomorphic to $\mathbb{P}^n$ with the smooth structure we defined in Chapter 1, which can be seen as follows. Consider the map $\pi' \colon \mathbb{S}^n \to \mathbb{P}^n$ defined as the composition of the inclusion $\iota \colon \mathbb{S}^n \hookrightarrow \mathbb{R}^{n+1} \smallsetminus \{0\}$ followed by the projection $\pi_0 \colon \mathbb{R}^{n+1} \smallsetminus \{0\} \mapsto \mathbb{P}^n$ defining $\mathbb{P}^n$. This is a smooth covering map (see Problem 7-1), and makes the same identifications as π. By the result of Problem 5.21, $\mathbb{S}^n/\{\pm 1\}$ is diffeomorphic to $\mathbb{P}^n$.

Quotients of Lie Groups

Another important application of the quotient manifold theorem is to the study quotients of Lie groups by Lie subgroups. Let G be a Lie group and let $H \subset G$ be a Lie subgroup. If we let H act on G by right translation, then an element of the orbit space is the orbit of an element $g \in G$, which is a set of the form $gH = \{gh : h \in H\}$. In other words, an orbit under the *right* action by H is a *left* coset of H. We will use the notation G/H to denote the orbit space by this right action (the left coset space).

Theorem 7.15. *Let G be a Lie group and let H be a closed Lie subgroup of G. The action of H on G by right translation is smooth, free, and proper. Therefore the left coset space G/H is a smooth manifold, and the quotient map $\pi \colon G \to G/H$ is a smooth submersion.*

Proof. We already observed in Example 7.3(f) that H acts smoothly and freely on G. To see that the action is proper, let $\Theta \colon G \times H \to G \times G$ be the map $\Theta(g,h) = (gh, g)$, and suppose $L \subset G \times G$ is a compact set. If $\{(g_i, h_i)\}$ is a sequence in $\Theta^{-1}(L)$, then passing to a subsequence if necessary we may assume that the sequences $\{g_i h_i\}$ and $\{g_i\}$ converge. By continuity, therefore, $h_i = g_i^{-1}(g_i h_i)$ converges to a point in G, and since H is closed in G it follows that $\{(g_i, h_i)\}$ converges in $G \times H$. □

A *discrete subgroup* of a Lie group is a subgroup that is a discrete set in the subspace topology (and is thus an embedded zero-dimensional Lie subgroup). The following corollary is an immediate consequence of Theorems 7.13 and 7.15.

Corollary 7.16. *Let G be a Lie group, and let $\Gamma \subset G$ be a discrete subgroup. Then the quotient map $\pi \colon G \to G/\Gamma$ is a smooth covering map.*

Example 7.17. Let C be the unit cube centered at the origin in $\mathbb{R}^3$. The set Γ of positive-determinant orthogonal transformations of $\mathbb{R}^3$ that take C to itself is a finite subgroup of $\mathrm{SO}(3)$, and the quotient $\mathrm{SO}(3)/\Gamma$ is a connected smooth 3-manifold whose universal cover is $\mathbb{S}^3$ (see Problem 7-13). Similar examples are obtained from the symmetry group of any regular polyhedron, such as a regular tetrahedron, dodecahedron, or icosahedron.

Homogeneous Spaces

One of the most interesting kinds of group action is that in which a group acts transitively. A smooth manifold endowed with a transitive smooth action by a Lie group G is called a *homogeneous G-space*, or a *homogeneous space* or *homogeneous manifold* if it is not important to specify the group.

In most examples, the group action preserves some property of the manifold (such as distances in some metric, or a class of curves such as straight lines in the plane); then the fact that the action is transitive means that the manifold "looks the same" everywhere from the point of view of this property. Often, homogeneous spaces are models for various kinds of geometric structures, and as such they play a central role in many areas of differential geometry.

Here are some important examples of homogeneous spaces.

Example 7.18 (Homogeneous Spaces).

(a) The natural action of $\mathrm{O}(n)$ on $\mathbb{S}^{n-1}$ is transitive, as we observed in Example 7.3. So is the natural action of $\mathrm{SO}(n)$ on $\mathbb{S}^{n-1}$ when $n \geq 2$. Thus for $n \geq 2$, $\mathbb{S}^{n-1}$ is a homogeneous space of either $\mathrm{O}(n)$ or $\mathrm{SO}(n)$.

(b) Let $E(n)$ denote the subgroup of $\mathrm{GL}(n+1,\mathbb{R})$ consisting of matrices of the form

$$\left\{ \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix} : A \in \mathrm{O}(n),\ b \in \mathbb{R}^n \right\},$$

where b is considered as an $n \times 1$ column matrix. It is straightforward to check that $E(n)$ is an embedded Lie subgroup. If $S \subset \mathbb{R}^{n+1}$ denotes the affine subspace defined by $x^{n+1} = 1$, then a simple computation shows that $E(n)$ takes S to itself. Identifying S with $\mathbb{R}^n$ in the obvious way, this induces an action of $E(n)$ on $\mathbb{R}^n$, in which the matrix $\left(\begin{smallmatrix} A & b \\ 0 & 1 \end{smallmatrix}\right)$ sends x to $Ax + b$. It is not hard to prove that these are precisely the transformations that preserve the Euclidean inner product (see Problem 7-17). For this reason, $E(n)$ is called the *Euclidean group*. Because any point in $\mathbb{R}^n$ can be taken to any other by a translation, $E(n)$ acts transitively on $\mathbb{R}^n$, so $\mathbb{R}^n$ is a homogeneous $E(n)$-space.

(c) The group $\mathrm{SL}(2,\mathbb{R})$ acts smoothly and transitively on the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}.$$

The resulting complex-analytic transformations of $\mathbb{H}$ are called *Möbius transformations*.

(d) If G is any Lie group and H is a closed Lie subgroup, the space G/H of left cosets is a smooth manifold by Theorem 7.15. We define a left action of G on G/H by

$$g_1 \cdot (g_2 H) = (g_1 g_2)H.$$

This action is obviously transitive, and Proposition 5.20 implies that it is smooth.

Exercise 7.5. Show that both $\mathrm{U}(n)$ and $\mathrm{SU}(n)$ act smoothly and transitively on $\mathbb{S}^{2n-1}$, thought of as the set of unit vectors in $\mathbb{C}^n$.

Example 7.18(d) above turns out to be of central importance because, as the next theorem shows, *every* homogeneous space is equivalent to one of this type.

Theorem 7.19 (Characterization of Homogeneous Spaces). *Let M be a homogeneous G-space, and let p be any point of M. Then the isotropy group G_p is a closed Lie subgroup of G, and the map $F\colon G/G_p \to M$ defined by $F(gG_p) = g \cdot p$ is an equivariant diffeomorphism.*

Proof. For simplicity, let us write $H = G_p$. First we will show that H is a closed Lie subgroup. Define a map $\Phi\colon G \to M$ by $\Phi(g) = g \cdot p$. This is obviously smooth, and $H = \Phi^{-1}(p)$. Observe that

$$\Phi(g'g) = (g'g) \cdot p = g' \cdot (g \cdot p) = g' \cdot \Phi(g),$$

so Φ is equivariant with respect to the action by G on itself by left multiplication and the given G-action on M. This implies that H is an embedded submanifold of G and therefore a closed Lie subgroup.

To see that F is well defined, assume that $g_1 H = g_2 H$, which means that $g_1^{-1} g_2 \in H$. Writing $g_1^{-1} g_2 = h$, we see that

$$F(g_2 H) = g_2 \cdot p = g_1 h \cdot p = g_1 \cdot p = F(g_1 H).$$

Also, F is equivariant, because

$$F(g'gH) = (g'g) \cdot p = g' \cdot F(gH).$$

It is smooth because it is obtained from Φ by passing to the quotient (see Proposition 5.20).

Next we show that F is bijective. Given any point $q \in M$ there is a group element $g \in G$ such that $F(gH) = g \cdot p = q$ by transitivity. On the other hand, if $F(g_1 H) = F(g_2 H)$, then $g_1 \cdot p = g_2 \cdot p$ implies $g_1^{-1} g_2 \cdot p = p$, so $g_1^{-1} g_2 \in H$, which implies $g_1 H = g_2 H$.

Because F is a bijective smooth map of constant rank, it is a diffeomorphism by Proposition 6.5. $\square$

This theorem shows that the study of homogeneous spaces can be reduced to the largely algebraic problem of understanding closed Lie subgroups of Lie groups. Because of this, some authors *define* a homogeneous space to be a quotient manifold of the form G/H, where G is a Lie group and H is a closed Lie subgroup of G.

Applying this theorem to the examples of transitive group actions we developed earlier, we see that some familiar spaces can be expressed as quotients of Lie groups by closed Lie subgroups.

Example 7.20 (Homogeneous Spaces Revisited).

(a) Consider again the natural action of $\mathrm{O}(n)$ on $\mathbb{S}^{n-1}$. If we choose our base point in $\mathbb{S}^{n-1}$ to be the "north pole" $N = (0, \dots, 0, 1)$, it is easy to check that the isotropy group is $\mathrm{O}(n-1)$, thought of as orthogonal transformations of $\mathbb{R}^n$ that fix the last variable. Thus $\mathbb{S}^{n-1}$ is diffeomorphic to the quotient manifold $\mathrm{O}(n)/\mathrm{O}(n-1)$. For the action of $\mathrm{SO}(n)$ on $\mathbb{S}^{n-1}$, the isotropy group is $\mathrm{SO}(n-1)$, so $\mathbb{S}^{n-1}$ is also diffeomorphic to $\mathrm{SO}(n)/\mathrm{SO}(n-1)$.

(b) Similarly, using the result of Exercise 7.5, we conclude that $\mathbb{S}^{2n-1} \approx \mathrm{U}(n)/\mathrm{U}(n-1) \approx \mathrm{SU}(n)/\mathrm{SU}(n-1)$.

(c) Because the Euclidean group $E(n)$ acts smoothly and transitively on $\mathbb{R}^n$, and the isotropy group of the origin is the subgroup $\mathrm{O}(n) \subset E(n)$ (identified with the $(n+1) \times (n+1)$ matrices of the form $\left(\begin{smallmatrix} A & 0 \\ 0 & 1 \end{smallmatrix}\right)$ with $A \in \mathrm{O}(n)$), $\mathbb{R}^n$ is diffeomorphic to $E(n)/\mathrm{O}(n)$.

Application: Sets with Transitive Group Actions

A highly useful application of the characterization theorem is to put smooth structures on *sets* that admit transitive Lie group actions.

Proposition 7.21. *Suppose X is a set, and we are given a transitive action of a Lie group G on X, such that the isotropy group of a point $p \in X$ is a closed Lie subgroup of G. Then X has a unique manifold topology and smooth structure such that the given action is smooth.*

Proof. Let H denote the isotropy group of p, so that G/H is a smooth manifold by Theorem 7.15. The map $F\colon G/H \to X$ defined by $F(gH) = g \cdot p$ is an equivariant bijection by exactly the same argument as we used in the proof of the characterization theorem. (That part did not use the fact that M was a manifold at all.) If we define a topology and smooth structure on X by declaring F to be a diffeomorphism, then the given action of G on X is smooth because it can be written $(g, x) \mapsto F(g \cdot F^{-1}(x))$.

If $\widetilde{X}$ denotes the set X with any smooth manifold structure such that the given action is smooth, then by the homogeneous space characterization theorem, $\widetilde{X}$ is equivariantly diffeomorphic to G/H and therefore to X, so the topology and smooth structure are unique. □

Example 7.22 (Grassmannians). Let $\mathrm{G}(k, n)$ denote the set of k-dimensional subspaces of $\mathbb{R}^n$ as in Example 1.15. The general linear group $\mathrm{GL}(n, \mathbb{R})$ acts transitively on $\mathrm{G}(k, n)$: Given two subspaces A and A', choose bases for both subspaces and extend them to bases for $\mathbb{R}^n$, and then the linear transformation taking the first basis to the second also takes A to A'. The isotropy group of the subspace $\mathbb{R}^k \subset \mathbb{R}^n$ is

$$H = \left\{ \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} : A \in \mathrm{GL}(k, \mathbb{R}),\ D \in \mathrm{GL}(n-k, \mathbb{R}), \right.$$
$$\left. B \in \mathrm{M}(k \times (n-k), \mathbb{R}) \right\},$$

which is a closed Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$. Therefore $\mathrm{G}(k, n)$ has a unique smooth manifold structure making the natural $\mathrm{GL}(n, \mathbb{R})$ action smooth. Problem 7-19 shows that this is the same smooth structure we defined in Example 1.15.

Example 7.23 (Flag Manifolds). Let V be a real vector space of dimension $n > 1$, and let $K = (k_1, \dots, k_m)$ be a finite sequence of integers satisfying $0 < k_1 < \cdots < k_m < n$. A *flag* in V of type K is a sequence of linear subspaces $S_1 \subset S_2 \subset \cdots \subset S_m \subset V$, with $\dim S_i = k_i$ for each i. The set of all flags of type K in V is denoted $F_K(V)$. (For example, if $K = (k)$, then $F_K(V)$ is the Grassmannian $\mathrm{G}_k(V)$.) It is not hard to show that $\mathrm{GL}(V)$ acts transitively on $F_K(V)$ with a closed Lie subgroup as isotropy group (see Problem 7-23), so $F_K(V)$ has a unique smooth manifold structure making it into a homogeneous $\mathrm{GL}(V)$-space. With this structure, $F_K(V)$ is called a *flag manifold*.

Application: Connectivity of Lie Groups

Another application of homogeneous space theory is to identify the connected components of many familiar Lie groups. The key result is the following proposition.

Proposition 7.24. *Suppose a Lie group G acts smoothly, freely, and properly on a manifold M. If G and M/G are connected, then M is connected.*

Proof. Suppose M is not connected. This means that there are nonempty, disjoint open sets $U, V \subset M$ whose union is M. Because the quotient map $\pi\colon M \to M/G$ is an open map (Proposition 5.18), $\pi(U)$ and $\pi(V)$ are nonempty open subsets of M/G. If $\pi(U) \cap \pi(V) \neq \varnothing$, there is a G-orbit that contains points of both U and V. However, each orbit is an embedded submanifold diffeomorphic to G, which is connected, so each orbit lies entirely in one of the sets U or V. Thus $\{\pi(U), \pi(V)\}$ would be a separation of M/G, which contradicts the assumption that M/G is connected. □

Proposition 7.25. *For any n, the Lie groups $\mathrm{SO}(n)$, $\mathrm{U}(n)$, and $\mathrm{SU}(n)$ are connected. The group $\mathrm{O}(n)$ has exactly two components, one of which is $\mathrm{SO}(n)$.*

Proof. We begin by proving that $\mathrm{SO}(n)$ is connected by induction on n. For $n = 1$ this is obvious, because $\mathrm{SO}(1)$ is the trivial group. Now suppose we have shown that $\mathrm{SO}(n-1)$ is connected for some $n \geq 2$. Because the homogeneous space $\mathrm{SO}(n)/\mathrm{SO}(n-1)$ is diffeomorphic to $\mathbb{S}^{n-1}$ and therefore is connected, Proposition 7.24 and the induction hypothesis imply that $\mathrm{SO}(n)$ is connected. A similar argument applies to $\mathrm{U}(n)$ and $\mathrm{SU}(n)$, using the facts that $\mathrm{U}(n)/\mathrm{U}(n-1) \approx \mathrm{SU}(n)/\mathrm{SU}(n-1) \approx \mathbb{S}^{2n-1}$.

Note that $\mathrm{O}(n)$ is equal to the union of the two open sets $\mathrm{O}^+(n)$ and $\mathrm{O}^-(n)$ consisting of orthogonal matrices whose determinant is $+1$ or -1, respectively. As we noted earlier, $\mathrm{O}^+(n) = \mathrm{SO}(n)$, which is connected. On the other hand, if A is any orthogonal matrix whose determinant is -1, then left translation L_A is a diffeomorphism from $\mathrm{O}^+(n)$ to $\mathrm{O}^-(n)$, so $\mathrm{O}^-(n)$ is connected as well. Therefore $\{\mathrm{O}^+(n), \mathrm{O}^-(n)\}$ are exactly the components of $\mathrm{O}(n)$. □

Determining the components of the general linear groups is a bit more involved. Let $\mathrm{GL}^+(n,\mathbb{R})$ and $\mathrm{GL}^-(n,\mathbb{R})$ denote the subsets of $\mathrm{GL}(n,\mathbb{R})$ consisting of matrices with positive determinant and negative determinant, respectively.

Proposition 7.26. *The components of $\mathrm{GL}(n,\mathbb{R})$ are $\mathrm{GL}^+(n,\mathbb{R})$ and $\mathrm{GL}^-(n,\mathbb{R})$.*

Proof. We begin by showing that $\mathrm{GL}^+(n,\mathbb{R})$ is connected. It suffices to show that it is path connected, which will follow once we show that there is a continuous path in $\mathrm{GL}^+(n,\mathbb{R})$ from any $A \in \mathrm{GL}^+(n,\mathbb{R})$ to the identity matrix I_n.

Let $A \in \mathrm{GL}^+(n,\mathbb{R})$ be arbitrary, and let $(A_1, \dots, A_n)$ denote the columns of A, considered as vectors in $\mathbb{R}^n$. The Gram-Schmidt algorithm

(Proposition A.17 in the Appendix) shows that there is an orthonormal basis $(Q_1, \dots, Q_n)$ for $\mathbb{R}^n$ with the property that $\operatorname{span}(Q_1, \dots, Q_k) = \operatorname{span}(A_1, \dots, A_k)$ for each $k = 1, \dots, n$. Thus we can write

$$\begin{aligned} A_1 &= R_1^1 Q_1, \\ A_2 &= R_2^1 Q_1 + R_2^2 Q_2, \\ &\vdots \\ A_n &= R_n^1 Q_1 + R_n^2 Q_2 + \dots + R_n^n Q_n, \end{aligned}$$

for some constants R_i^j. Replacing each Q_i by $-Q_i$ if necessary, we may assume that $R_i^i > 0$ for each i. In matrix notation, this is equivalent to $A = QR$, where R is upper triangular with positive entries on the diagonal. Since the determinant of R is the product of its diagonal entries and $\det A = (\det Q)(\det R) > 0$, it follows that $Q \in \mathrm{SO}(n)$. (This *QR decomposition* plays an important role in numerical linear algebra.)

Let $R_t = tI_n + (1-t)R$. It is immediate that R_t is upper triangular with positive diagonal entries for all $t \in [0,1]$, so $R_t \in \mathrm{GL}^+(n,\mathbb{R})$. Therefore, the path $\gamma\colon [0,1] \to \mathrm{GL}^+(n,\mathbb{R})$ given by $\gamma(t) = QR_t$ satisfies $\gamma(0) = A$ and $\gamma(1) = Q \in \mathrm{SO}(n)$. Because $\mathrm{SO}(n)$ is connected, there is a path in $\mathrm{SO}(n)$ from Q to the identity matrix. This shows that $\mathrm{GL}^+(n,\mathbb{R})$ is path connected.

Now, as in the case of $\mathrm{O}(n)$, any matrix B with $\det B < 0$ yields a diffeomorphism $L_B\colon \mathrm{GL}^+(n,\mathbb{R}) \to \mathrm{GL}^-(n,\mathbb{R})$, so $\mathrm{GL}^-(n,\mathbb{R})$ is connected as well. This completes the proof. $\square$

Problems

7-1. Let $\pi\colon \mathbb{S}^n \to \mathbb{P}^n$ be the map that sends $x \in \mathbb{S}^n$ to the line through the origin and x, thought of as a point in $\mathbb{P}^n$. Show that π is a smooth covering map.

7-2. Define a map $F\colon \mathbb{P}^2 \to \mathbb{R}^4$ by

$$F[x,y,z] = (x^2 - y^2, xy, xz, yz).$$

Show that F is a smooth embedding.

7-3. Let G be a Lie group and $H \subset G$ a closed normal Lie subgroup. Show that G/H is a Lie group and the quotient map $\pi\colon G \to G/H$ is a Lie homomorphism.

7-4. If $F\colon G \to H$ is a surjective Lie group homomorphism, show that H is Lie isomorphic to $G/\operatorname{Ker} F$.

7-5. Let G be a connected Lie group, and suppose $F\colon G \to H$ is a surjective Lie group homomorphism with discrete kernel. Show that F is a smooth covering map.

7-6. Let G be a Lie group, and suppose $\pi\colon \widetilde{G} \to G$ is any covering map. For any point $\widetilde{e} \in \pi^{-1}(e)$, show that $\widetilde{G}$ has a unique Lie group structure such that $\widetilde{e}$ is the identity element of $\widetilde{G}$ and π is a Lie group homomorphism.

7-7. (a) Show that there exists a Lie group homomorphism $\rho : \mathbb{S}^1 \to \mathrm{U}(n)$ such that $\det \circ \rho = \mathrm{Id}_{\mathbb{S}^1}$.

(b) Show that $\mathrm{U}(n)$ is diffeomorphic to $\mathbb{S}^1 \times \mathrm{SU}(n)$. [Hint: Consider the map $\varphi : \mathbb{S}^1 \times \mathrm{SU}(n) \to \mathrm{U}(n)$ given by $\varphi(z, A) = \rho(z)A$.]

(c) Show that $\mathrm{U}(n)$ and $\mathbb{S}^1 \times \mathrm{SU}(n)$ are not isomorphic Lie groups.

7-8. Show that $\mathrm{SU}(2)$ is diffeomorphic to $\mathbb{S}^3$.

7-9. Let G be a Lie group, and let G_0 denote the connected component of the identity (called the *identity component* of G).

(a) Show that G_0 is an embedded Lie subgroup of G, and that each connected component of G is diffeomorphic to G_0.

(b) If H is any connected open subgroup of G, show that $H = G_0$.

7-10. Suppose a Lie group acts smoothly on a manifold M.

(a) Show that each orbit is an immersed submanifold of M.

(b) Give an example of a Lie group acting smoothly on a manifold M in which two different orbits have different dimensions even though neither orbit has dimension equal to zero or to the dimension of M.

7-11. Prove the following partial converse to the quotient manifold theorem: If a Lie group G acts smoothly and freely on a smooth manifold M and the orbit space M/G has a smooth manifold structure such that the quotient map $\pi\colon M \to M/G$ is a smooth submersion, then G acts properly.

7-12. Give an example of a smooth, proper action of a Lie group on a smooth manifold such that the orbit space is not a topological manifold.

7-13. Prove that $\mathrm{SO}(3)$ is Lie isomorphic to $\mathrm{SU}(2)/\{\pm I\}$ and diffeomorphic to $\mathbb{P}^3$, as follows.

(a) Let $\mathcal{H}$ denote the set of 2×2 Hermitian matrices whose trace is zero. (The *trace* of a matrix is the sum of its diagonal entries.) Show that $\mathcal{H}$ is a 3-dimensional vector space over $\mathbb{R}$, and

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

is a basis for $\mathcal{H}$.

(b) If we give $\mathcal{H}$ the inner product for which (E_1, E_2, E_3) is an orthonormal basis, show that $|A|^2 = -\det A$ for all $A \in \mathcal{H}$.

(c) Identifying $\mathrm{GL}(3, \mathbb{R})$ with the set of invertible real-linear maps $\mathcal{H} \to \mathcal{H}$ by means of the basis (E_1, E_2, E_3), define a map $\rho\colon \mathrm{SU}(2) \to \mathrm{GL}(3, \mathbb{R})$ by

$$\rho(X)A = XAX^{-1}, \qquad X \in \mathrm{SU}(2),\ A \in \mathcal{H}.$$

Show that ρ is a Lie group homomorphism whose image is $\mathrm{SO}(3)$ and whose kernel is $\{\pm I\}$. [Hint: To show that the image is all of $\mathrm{SO}(3)$, show that ρ is open and closed and use the results of Problem 7-9.]

(d) Prove the result.

7-14. Determine which of the following Lie groups are compact: $\mathrm{GL}(n, \mathbb{R})$, $\mathrm{SL}(n, \mathbb{R})$, $\mathrm{GL}(n, \mathbb{C})$, $\mathrm{SL}(n, \mathbb{C})$, $\mathrm{U}(n)$, $\mathrm{SU}(n)$.

7-15. Show that $\mathrm{GL}(n, \mathbb{C})$ is connected.

7-16. Show that $\mathrm{SL}(n, \mathbb{R})$ and $\mathrm{SL}(n, \mathbb{C})$ are connected.

7-17. Prove that the set of maps from $\mathbb{R}^n$ to itself given by the action of $E(n)$ on $\mathbb{R}^n$ described in Example 7.18(b) is exactly the set of all maps from $\mathbb{R}^n$ to itself that preserve the Euclidean inner product.

7-18. Prove that the Grassmannian $\mathrm{G}(k,n)$ is compact for any k and n.

7-19. Show that the smooth structure on the Grassmannian $\mathrm{G}(k,n)$ defined in Example 7.22 is the same as the one defined in Example 1.15.

7-20. Show that the image of a Lie group homomorphism is a Lie subgroup.

7-21. (a) Let G and H be Lie groups. Suppose $\rho\colon H\times G\to G$ is a smooth left action of H on G with the property that $\rho_h\colon G\to G$ is a Lie group homomorphism for every $h\in H$. Define a group structure on the manifold $G\times H$ by

$$(g,h)(g',h')=(g\rho_h(g'),hh').$$

Show that this turns $G\times H$ into a Lie group, called the *semidirect product* of G and H induced by ρ, and denoted by $G\rtimes_\rho H$.

(b) If G is any Lie group, show that G is Lie isomorphic to a semidirect product of a connected Lie group with a discrete group.

7-22. Define an action of $\mathbb{Z}$ on $\mathbb{R}^2$ by

$$n\cdot(x,y)=(x+n,(-1)^n y).$$

(a) Show that the action is smooth, free and proper. Let $E=\mathbb{R}^2/\mathbb{Z}$ denote the quotient manifold.

(b) Show that the projection on the first coordinate $\pi_1\colon\mathbb{R}^2\to\mathbb{R}$ descends to a smooth map $\pi\colon E\to\mathbb{S}^1$.

(c) Show that E is a rank-1 vector bundle over $\mathbb{S}^1$ with projection π. (It is called the *Möbius bundle*.)

(d) Show that E is not a trivial bundle.

7-23. Let $F_K(V)$ be the set of flags of type K in a finite-dimensional vector space V as in Example 7.23. Show that $\mathrm{GL}(V)$ acts transitively on $F_K(V)$, and that the isotropy group of a particular flag is a closed Lie subgroup of $\mathrm{GL}(V)$.

7-24. The n-dimensional *complex projective space*, denoted by $\mathbb{CP}^n$, is the set of 1-dimensional complex subspaces of $\mathbb{C}^{n+1}$. Show that $\mathbb{CP}^n$ has a unique topology and smooth structure making it into a $2n$-dimensional compact manifold and a homogeneous space of $\mathrm{U}(n)$.

7-25. Show that $\mathbb{CP}^1$ is diffeomorphic to $\mathbb{S}^2$.

7-26. Considering $\mathbb{S}^{2n+1}$ as the unit sphere in $\mathbb{C}^{n+1}$, define an action of $\mathbb{S}^1$ on $\mathbb{S}^{2n+1}$ by

$$z \cdot (w^1, \dots, w^{n+1}) = (zw^1, \dots, zw^{n+1}).$$

Show that this action is smooth, free, and proper, and that the orbit space $\mathbb{S}^{2n+1}/\mathbb{S}^1$ is diffeomorphic to $\mathbb{CP}^n$. [Hint: Consider the restriction of the natural quotient map $\mathbb{C}^{n+1} \smallsetminus \{0\} \to \mathbb{CP}^n$ to $\mathbb{S}^{2n+1}$. The quotient map $\pi\colon \mathbb{S}^{2n+1} \to \mathbb{CP}^n$ is known as the *Hopf map*.]

7-27. Let c be an irrational number, and let $\mathbb{R}$ act on $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ by

$$t \cdot (w, z) = (e^{2\pi i t} w, e^{2\pi i c t} z).$$

Show that this is a smooth free action, but the quotient $\mathbb{T}^2/\mathbb{R}$ is not Hausdorff.

8
Tensors

Much of the machinery of smooth manifold theory is designed to allow the concepts of linear algebra to be applied to smooth manifolds. For example, a tangent vector can be thought of as a linear approximation to a curve; the tangent space to a submanifold can be thought of as a linear approximation to the submanifold; and the push-forward of a smooth map can be thought of as a linear approximation to the map itself. Calculus tells us how to approximate smooth objects by linear ones, and the abstract definitions of manifold theory give a way to interpret these linear approximations in a coordinate-invariant way. In this chapter, we carry this idea much further, by generalizing from linear objects to multlinear ones. This leads to the concepts of tensors and tensor fields on manifolds.

We begin with tensors on a vector space, which are multilinear generalizations of covectors; a covector is the special case of a tensor of rank one. We give two alternative definitions of tensors on a vector space: On the one hand, they are real-valued multilinear functions of several vectors; on the other hand, they are elements of the abstract "tensor product" of the dual vector space with itself. Each definition is useful in certain contexts. We then discuss the difference between covariant and contravariant tensors, and give a brief introduction to tensors of mixed variance.

We then move to smooth manifolds, and define tensors, tensor fields, and tensor bundles. After describing the coordinate representations of tensor fields, we describe how they can be pulled back by smooth maps. We introduce a special class of tensors, the symmetric ones, whose values are unchanged by permutations of their arguments.

The last section of the chapter is an introduction to one of the most important kinds of tensor fields, Riemannian metrics. A thorough treatment of Riemannian geometry is beyond the scope of this book, but we can at least lay the groundwork by giving the basic definitions and proving that every manifold admits Riemannian metrics.

The Algebra of Tensors

Suppose $V_1, \ldots, V_k$ and W are vector spaces. A map $F\colon V_1 \times \cdots \times V_k \to W$ is said to be *multilinear* if it is linear as a function of each variable separately:

$$F(v_1, \ldots, av_i + a'v_i', \ldots, v_k)$$
$$= aF(v_1, \ldots, v_i, \ldots, v_k) + a'F(v_1, \ldots, v_i', \ldots, v_k).$$

(A multilinear function of two variables is generally called *bilinear*.) Although linear maps are paramount in differential geometry, there are many situations in which multilinear maps play an important geometric role.

Here are a few examples to keep in mind:

- *The dot product* in $\mathbb{R}^n$ is a scalar-valued bilinear function of two vectors, used to compute lengths of vectors and angles between them.
- *The cross product* in $\mathbb{R}^3$ is a vector-valued bilinear function of two vectors, used to compute areas of parallelograms and to find a third vector orthogonal to two given ones.
- *The determinant* is a real-valued multilinear function of n vectors in $\mathbb{R}^n$, used to detect linear independence and to compute the volume of the parallelepiped spanned by the vectors.

In this section, we will develop a unified language for talking about multilinear functions—the language of tensors. In a little while, we will give a very general and abstract definition of tensors. But it will help to clarify matters if we start with a more concrete definition.

Let V be a finite-dimensional real vector space, and k a natural number. (Many of the concepts we will introduce in this section—at least the parts that do not refer explicitly to finite bases—work equally well in the infinite-dimensional case; but we will restrict our attention to the finite-dimensional case in order to keep things simple.)

A *covariant k-tensor* on V is a real-valued multilinear function of k elements of V:

$$T\colon \underbrace{V \times \cdots \times V}_{k \text{ copies}} \to \mathbb{R}.$$

The number k is called the *rank* of T. A 0-tensor is, by convention, just a real number (a real-valued function depending multilinearly on no vectors!). The set of all covariant k-tensors on V, denoted by $T^k(V)$, is a vector space under the usual operations of pointwise addition and scalar multiplication:

$$(aT)(X_1, \dots, X_k) = a(T(X_1, \dots, X_k)),$$

$$(T + T')(X_1, \dots, X_k) = T(X_1, \dots, X_k) + T'(X_1, \dots, X_k).$$

Let us look at some examples.

Example 8.1 (Covariant Tensors).

(a) Every linear map $\omega\colon V \to \mathbb{R}$ is multilinear, so a covariant 1-tensor is just a covector. Thus $T^1(V)$ is naturally identified with V^*.

(b) A covariant 2-tensor on V is a real-valued bilinear function of two vectors, also called a *bilinear form*. One example is the dot product on $\mathbb{R}^n$. More generally, any inner product on V is a covariant 2-tensor.

(c) The determinant, thought of as a function of n vectors, is a covariant n-tensor on $\mathbb{R}^n$.

(d) Suppose $\omega, \eta \in V^*$. Define a map $\omega \otimes \eta\colon V \times V \to \mathbb{R}$ by

$$\omega \otimes \eta(X, Y) = \omega(X)\eta(Y),$$

where the product on the right is just ordinary multiplication of real numbers. The linearity of ω and η guarantees that $\omega \otimes \eta$ is a bilinear function of X and Y, i.e., a 2-tensor.

The last example can be generalized to tensors of any rank as follows. Let V be a finite-dimensional real vector space and let $S \in T^k(V)$, $T \in T^l(V)$. Define a map

$$S \otimes T\colon \underbrace{V \times \cdots \times V}_{k+l \text{ copies}} \to \mathbb{R}$$

by

$$S \otimes T(X_1, \dots, X_{k+l}) = S(X_1, \dots, X_k)T(X_{k+1}, \dots, X_{k+l}).$$

It is immediate from the multilinearity of S and T that $S \otimes T$ depends linearly on each argument X_i separately, so it is a covariant $(k+l)$-tensor, called the *tensor product* of S and T.

Exercise 8.1. Show that the tensor product operation is bilinear and associative. More precisely, show that $S \otimes T$ depends linearly on each of the tensors S and T, and that $(R \otimes S) \otimes T = R \otimes (S \otimes T)$.

Because of the result of the preceding exercise, we can write the tensor product of three or more tensors unambiguously without parentheses. If $T_1, \dots, T_l$ are tensors of ranks $k_1, \dots, k_l$ respectively, their tensor product $T_1 \otimes \cdots \otimes T_l$ is a tensor of rank $k = k_1 + \cdots + k_l$, whose action on k vectors is given by inserting the first k_1 vectors into T_1, the next k_2 vectors into T_2, and so forth, and multiplying the results together. For example, if R and S are 2-tensors and T is a 3-tensor, then

$$R \otimes S \otimes T(X_1, \dots, X_7) = R(X_1, X_2) S(X_3, X_4) T(X_5, X_6, X_7).$$

Proposition 8.2. *Let V be a real vector space of dimension n, let (E_i) be any basis for V, and let (ε^i) be the dual basis. The set of all k-tensors of the form $\varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k}$ for $1 \le i_1, \dots, i_k \le n$ is a basis for $T^k(V)$, which therefore has dimension n^k.*

Proof. Let $\mathcal{B}$ denote the set $\{\varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k} : 1 \le i_1, \dots, i_k \le n\}$. We need to show that $\mathcal{B}$ is independent and spans $T^k(V)$. Suppose $T \in T^k(V)$ is arbitrary. For any k-tuple $(i_1, \dots, i_k)$ of integers such that $1 \le i_j \le n$, define a number $T_{i_1 \dots i_k}$ by

$$T_{i_1 \dots i_k} = T(E_{i_1}, \dots, E_{i_k}). \tag{8.1}$$

We will show that

$$T = T_{i_1 \dots i_k} \varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k}$$

(with the summation convention in effect as usual), from which it follows that $\mathcal{B}$ spans V. We compute

$$\begin{aligned} T_{i_1 \dots i_k} \varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k}(E_{j_1}, \dots, E_{j_k}) &= T_{i_1 \dots i_k} \varepsilon^{i_1}(E_{j_1}) \cdots \varepsilon^{i_k}(E_{j_k}) \\ &= T_{i_1 \dots i_k} \delta^{i_1}_{j_1} \cdots \delta^{i_k}_{j_k} \\ &= T_{j_1 \dots j_k} \\ &= T(E_{j_1}, \dots, E_{j_k}). \end{aligned}$$

By multilinearity, a tensor is determined by its action on sequences of basis vectors, so this proves the claim.

To show that $\mathcal{B}$ is independent, suppose some linear combination equals zero:

$$T_{i_1 \dots i_k} \varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k} = 0.$$

Apply this to any sequence $(E_{j_1}, \dots, E_{j_k})$ of basis vectors. By the same computation as above, this implies that each coefficient $T_{j_1 \dots j_k}$ is zero. Thus the only linear combination of elements of $\mathcal{B}$ that sums to zero is the trivial one. $\square$

The proof of this proposition shows, by the way, that the components $T_{i_1\dots i_k}$ of a tensor T in terms of the basis tensors in $\mathcal{B}$ are given by (8.1).

It is useful to see explicitly what this proposition means for tensors of low rank.

- $k = 0$: $T^0(V)$ is just $\mathbb{R}$, so $\dim T^0(V) = 1 = n^0$.
- $k = 1$: $T^1(V) = V^*$ has dimension $n = n^1$.
- $k = 2$: $T^2(V)$ is the space of bilinear forms on V. Any bilinear form can be written uniquely as $T = T_{ij}\varepsilon^i \otimes \varepsilon^j$, where (T_{ij}) is an arbitrary $n \times n$ matrix. Thus $\dim T^2(V) = n^2$.

Abstract Tensor Products of Vector Spaces

Because every covariant k-tensor can be written as a linear combination of tensor products of covectors, it is suggestive to write

$$T^k(V) = V^* \otimes \cdots \otimes V^*,$$

where we think of the expression on the right-hand side as a shorthand for the set of all linear combinations of tensor products of elements of V^*.

We will now give a construction that makes sense of this notation in a much more general setting. The construction is a bit involved, but the idea is simple: Given vector spaces V and W, we will construct a vector space $V \otimes W$ that consists of linear combinations of objects of the form $v \otimes w$ for $v \in V$, $w \in W$, defined in such a way that $v \otimes w$ depends bilinearly on v and w.

Let S be a set. The *free vector space* on S, denoted $\mathbb{R}\langle S\rangle$, is the set of all finite formal linear combinations of elements of S with real coefficients. More precisely, a finite formal linear combination is a function $\mathcal{F}\colon S \to \mathbb{R}$ such that $\mathcal{F}(s) = 0$ for all but finitely many $s \in S$. Under pointwise addition and scalar multiplication, $\mathbb{R}\langle S\rangle$ becomes a real vector space. Identifying each element $x \in S$ with the function that takes the value 1 on x and zero on all other elements of S, any element $\mathcal{F} \in \mathbb{R}\langle S\rangle$ can be written uniquely in the form $\mathcal{F} = \sum_{i=1}^m a_i x_i$, where $x_1, \dots, x_m$ are the elements of S for which $\mathcal{F}(x_i) \neq 0$, and $a_i = \mathcal{F}(x_i)$. Thus S is a basis for $\mathbb{R}\langle S\rangle$, which is therefore finite-dimensional if and only if S is a finite set.

Exercise 8.2 (Characteristic Property of Free Vector Spaces). Let S be a set and W a vector space. Show that any map $F\colon S \to W$ has a unique extension to a linear map $\overline{F}\colon \mathbb{R}\langle S\rangle \to W$.

Now let V and W be finite-dimensional real vector spaces, and let $\mathcal{R}$ be the subspace of the free vector space $\mathbb{R}\langle V \times W\rangle$ spanned by all elements of the following forms:

$$\begin{gathered}
a(v,w) - (av,w),\\
a(v,w) - (v,aw),\\
(v,w) + (v',w) - (v+v',w),\\
(v,w) + (v,w') - (v,w+w'),
\end{gathered} \tag{8.2}$$

for $a \in \mathbb{R}$, $v, v' \in V$, and $w, w' \in W$. Define the *tensor product* of V and W, denoted by $V \otimes W$, to be the quotient space $\mathbb{R}\langle V \times W\rangle/\mathcal{R}$. The equivalence class of an element (v,w) in $V \otimes W$ is denoted by $v \otimes w$, and is called the *tensor product* of v and w. From the definition, tensor products satisfy

$$\begin{aligned}
a(v\otimes w) &= av \otimes w = v \otimes aw,\\
v \otimes w + v' \otimes w &= (v+v') \otimes w,\\
v \otimes w + v \otimes w' &= v \otimes (w + w').
\end{aligned}$$

Note that the definition implies that every element of $V \otimes W$ can be written as a linear combination of elements of the form $v \otimes w$ for $v \in V$, $w \in W$; but it is not true in general that *every* element of $V \otimes W$ is of this form.

Proposition 8.3 (Characteristic Property of Tensor Products). *Let V and W be finite-dimensional real vector spaces. If $A\colon V \times W \to Y$ is a bilinear map into any vector space Y, there is a unique linear map $\widetilde{A}\colon V \otimes W \to Y$ such that the following diagram commutes:*

$$\begin{tikzcd}
V \times W \arrow[r, "A"] \arrow[d, "\pi"'] & Y \\
V \otimes W, \arrow[ur, "\widetilde{A}"'] &
\end{tikzcd} \tag{8.3}$$

where $\pi(v,w) = v \otimes w$.

Proof. First note that any map $A\colon V \times W \to X$ extends uniquely to a linear map $\overline{A}\colon \mathbb{R}\langle V \times W\rangle \to X$ by the characteristic property of the free vector space. This map is characterized by the fact that $\overline{A}(v,w) = A(v,w)$ whenever $(v,w) \in V \times W \subset \mathbb{R}\langle V \times W\rangle$. The fact that A is bilinear means precisely that the subspace $\mathcal{R}$ is contained in the kernel of $\overline{A}$, because

$$\begin{aligned}
\overline{A}(av,w) &= A(av,w)\\
&= aA(v,w)\\
&= a\overline{A}(v,w)\\
&= \overline{A}(a(v,w)),
\end{aligned}$$

with similar considerations for the other expressions in (8.2). Therefore, $\overline{A}$ descends to a linear map $\widetilde{A}\colon V \otimes W = \mathbb{R}\langle V \times W\rangle/\mathcal{R} \to X$ satisfying

$\widetilde{A} \circ \pi = A$. Uniqueness follows from the fact that every element of $V \otimes W$ can be written as a linear combination of elements of the form $v \otimes w$, and $\widetilde{A}$ is uniquely determined on such elements by $\widetilde{A}(v \otimes w) = \overline{A}(v,w) = A(v,w)$. □

The reason this is called the characteristic property is that it uniquely characterizes the tensor product up to isomorphism; see Problem 8-1.

Proposition 8.4 (Other Properties of Tensor Products). *Let V, W, and X be finite-dimensional real vector spaces.*

(*a*) *The tensor product $V^* \otimes W^*$ is canonically isomorphic to the space $B(V,W)$ of bilinear maps from $V \times W$ into $\mathbb{R}$.*

(*b*) *If (E_i) is a basis for V and (F_j) is a basis for W, then the set of all elements of the form $E_i \otimes F_j$ is a basis for $V \otimes W$, which therefore has dimension equal to $(\dim V)(\dim W)$.*

(*c*) *There is a unique isomorphism $V \otimes (W \otimes X) \to (V \otimes W) \otimes X$ sending $v \otimes (w \otimes x)$ to $(v \otimes w) \otimes x$.*

Proof. The canonical isomorphism between $V^* \otimes W^*$ and $B(V,W)$ is constructed as follows. First, define a map $\Phi \colon V^* \times W^* \to B(V,W)$ by

$$\Phi(\omega,\eta)(v,w) = \omega(v)\eta(w).$$

It is easy to check that Φ is bilinear, so by the characteristic property it descends uniquely to a linear map $\widetilde{\Phi} \colon V^* \otimes W^* \to B(V,W)$.

To see that $\widetilde{\Phi}$ is an isomorphism, we will construct an inverse for it. Let (E_i) and (F_j) be any bases for V and W, respectively, with dual bases (ε^i) and (φ^j). Since $V^* \otimes W^*$ is spanned by elements of the form $\omega \otimes \eta$ for $\omega \in V^*$ and $\eta \in W^*$, every $\tau \in V^* \otimes W^*$ can be written in the form $\tau = \tau_{ij}\varepsilon^i \otimes \varphi^j$. (We are not claiming yet that this expression is unique.)

Define a map $\Psi \colon B(V,W) \to V^* \otimes W^*$ by setting

$$\Psi(b) = b(E_k, F_l)\varepsilon^k \otimes \varphi^l.$$

We will show that Ψ and $\widetilde{\Phi}$ are inverses. First, for $\tau = \tau_{ij}\varepsilon^i \otimes \varphi^j \in V^* \otimes W^*$,

$$\begin{aligned}
\Psi \circ \widetilde{\Phi}(\tau) &= \widetilde{\Phi}(\tau)(E_k, F_l)\varepsilon^k \otimes \varphi^l \\
&= \tau_{ij}\widetilde{\Phi}(\varepsilon^i \otimes \varphi^j)(E_k, F_l)\varepsilon^k \otimes \varphi^l \\
&= \tau_{ij}\Phi(\varepsilon^i, \varphi^j)(E_k, F_l)\varepsilon^k \otimes \varphi^l \\
&= \tau_{ij}\varepsilon^i(E_k)\varphi^j(F_l)\varepsilon^k \otimes \varphi^l \\
&= \tau_{ij}\varepsilon^i \otimes \varphi^j \\
&= \tau.
\end{aligned}$$

On the other hand, for $b \in B(V,W)$, $v \in V$, and $w \in W$,

$$\begin{aligned}\widetilde{\Phi} \circ \Psi(b)(v,w) &= \widetilde{\Phi}\big(b(E_k,F_l)\varepsilon^k \otimes \varphi^l\big)(v,w)\\ &= b(E_k,F_l)\widetilde{\Phi}(\varepsilon^k \otimes \varphi^l)(v,w)\\ &= b(E_k,F_l)\varepsilon^k(v)\varphi^l(w)\\ &= b(E_k,F_l)v^k w^l\\ &= b(v,w).\end{aligned}$$

Thus $\Psi = \widetilde{\Phi}^{-1}$. (Note that although we used bases to prove that $\widetilde{\Phi}$ is invertible, $\widetilde{\Phi}$ itself is canonically defined without reference to any basis.)

We have already observed above that the elements of the form $\varepsilon^i \otimes \varphi^j$ span $V^* \otimes W^*$. On the other hand, it is easy to check that $\dim B(V,W) = (\dim V)(\dim W)$ (because any bilinear form is uniquely determined by its action on pairs of basis elements), so for dimensional reasons the set $\{\varepsilon^i \otimes \varphi^j\}$ is a basis for $V^* \otimes W^*$.

Finally, the isomorphism between $V \otimes (W \otimes X)$ and $(V \otimes W) \otimes X$ is constructed as follows. For each $x \in X$, the map $\alpha_x \colon V \times W \to V \otimes (W \otimes X)$ defined by

$$\alpha_x(v,w) = v \otimes (w \otimes x)$$

is obviously bilinear, and thus by the characteristic property of the tensor product it descends uniquely to a linear map $\widetilde{\alpha}_x \colon V \otimes W \to V \otimes (W \otimes X)$ satisfying $\widetilde{\alpha}_x(v \otimes w) = v \otimes (w \otimes x)$. Similarly, the map $\beta \colon (V \otimes W) \times X \to V \otimes (W \otimes X)$ given by

$$\beta(\tau, x) = \widetilde{\alpha}_x(\tau)$$

determines a linear map $\widetilde{\beta} \colon (V \otimes W) \otimes X \to V \otimes (W \otimes X)$ satisfying

$$\widetilde{\beta}((v \otimes w) \otimes x) = v \otimes (w \otimes x).$$

Because $V \otimes (W \otimes X)$ is spanned by elements of the form $v \otimes (w \otimes x)$, β is clearly surjective, and therefore it is an isomorphism for dimensional reasons. It is clearly the unique such isomorphism, because any other would have to agree with β on the set of elements of the form $(v \otimes w) \otimes x$, which spans $(V \otimes W) \otimes X$. $\square$

The next corollary explains the relationship between this abstract tensor product of vector spaces and the more concrete covariant k-tensors we defined earlier.

Corollary 8.5. *If V is a finite-dimensional real vector space, the space $T^k(V)$ of covariant k-tensors on V is canonically isomorphic to the k-fold tensor product $V^* \otimes \cdots \otimes V^*$.*

Exercise 8.3. Prove Corollary 8.5.

Using these results, we can generalize the notion of covariant tensors on a vector space as follows. For any finite-dimensional real vector space V, define the space of *contravariant tensors* of rank k to be

$$T_k(V) = \underbrace{V \otimes \cdots \otimes V}_{k \text{ copies}}.$$

Because of the canonical identification $V = V^{**}$ and Corollary 8.5, an element of $T_k(V)$ can be canonically identified with a multilinear function from $V^* \times \cdots \times V^*$ into $\mathbb{R}$. In particular, $T_1(V) \cong V^{**} \cong V$, the space of "contravariant vectors."

More generally, for any $k, l \in \mathbb{N}$, the space of *mixed tensors* on V of type $\binom{k}{l}$ is defined as

$$T^k_l(V) = \underbrace{V^* \otimes \cdots \otimes V^*}_{k \text{ copies}} \otimes \underbrace{V \otimes \cdots \otimes V}_{l \text{ copies}}.$$

From the discussion above, $T^k_l(V)$ can be identified with the set of real-valued multilinear functions of k vectors and l covectors.

In this book, we will be concerned primarily with covariant tensors, which we will think of primarily as multilinear functions of vectors, in keeping with our original definition. Thus tensors will always be understood to be covariant unless we explicitly specify otherwise. However, it is important to be aware that contravariant and mixed tensors play an important role in more advanced parts of differential geometry, especially Riemannian geometry.

Tensors and Tensor Fields on Manifolds

Now let M be a smooth manifold. We define the *bundle of covariant k-tensors* on M by

$$T^k M = \coprod_{p \in M} T^k(T_p M).$$

Similarly, we define the *bundle of contravariant l-tensors* by

$$T_l M = \coprod_{p \in M} T_l(T_p M),$$

and the *bundle of mixed tensors of type* $\binom{k}{l}$ by

$$T^k_l M = \coprod_{p \in M} T^k_l(T_p M).$$

Clearly there are natural identifications

$$\begin{aligned} T^0M &= T_0M = M \times \mathbb{R}, \\ T^1M &= T^*M, \\ T_1M &= TM, \\ T^k_0M &= T^kM, \\ T^0_lM &= T_lM. \end{aligned}$$

Exercise 8.4. Show that T^kM, T_lM, and T^k_lM have natural structures as smooth vector bundles over M, and determine their ranks.

Any one of these bundles is called a *tensor bundle* over M. (Thus the tangent and cotangent bundles are special cases of tensor bundles.) A section of a tensor bundle is called a (*covariant, contravariant, or mixed*) *tensor field* on M. A *smooth tensor field* is a section that is smooth in the usual sense of smooth sections of vector bundles. We denote the vector spaces of smooth sections of these bundles by

$$\begin{aligned} \mathcal{T}^k(M) &= \{\text{smooth sections of } T^kM\}; \\ \mathcal{T}_l(M) &= \{\text{smooth sections of } T_lM\}; \\ \mathcal{T}^k_l(M) &= \{\text{smooth sections of } T^k_lM\}. \end{aligned}$$

In any local coordinates (x^i), sections of these bundles can be written (using the summation convention) as

$$\sigma = \begin{cases} \sigma_{i_1\dots i_k} dx^{i_1} \otimes \cdots \otimes dx^{i_k}; & \sigma \in \mathcal{T}^k(M), \\ \sigma^{j_1\dots j_l} \dfrac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \dfrac{\partial}{\partial x^{j_l}}; & \sigma \in \mathcal{T}_l(M), \\ \sigma^{j_1\dots j_l}_{i_1\dots i_k} dx^{i_1} \otimes \cdots \otimes dx^{i_k} \otimes \dfrac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \dfrac{\partial}{\partial x^{j_l}}; & \sigma \in \mathcal{T}^k_l(M). \end{cases}$$

The functions $\sigma_{i_1\dots i_k}$, $\sigma^{j_1\dots j_l}$, or $\sigma^{j_1\dots j_l}_{i_1\dots i_k}$ are called the *component functions* of σ in these coordinates.

Lemma 8.6. *Let M be a smooth manifold, and let $\sigma\colon M \to T^kM$ be a map (not assumed to be continuous) such that $\sigma_p \in T^k(T_pM)$ for each $p \in M$. The following are equivalent.*

(*a*) *σ is smooth.*

(*b*) *In any coordinate chart, the component functions of σ are smooth.*

(*c*) *If $X_1, \dots, X_k$ are vector fields defined on any open subset $U \subset M$, then the function $\sigma(X_1, \dots, X_k)\colon U \to \mathbb{R}$, defined by*

$$\sigma(X_1, \dots, X_k)(p) = \sigma_p(X_1|_p, \dots, X_k|_p),$$

is smooth.

Exercise 8.5. Prove Lemma 8.6.

Exercise 8.6. Formulate and prove smoothness criteria analogous to those of Lemma 8.6 for contravariant and mixed tensor fields.

Smooth covariant 1-tensor fields are just covector fields. Recalling that a 0-tensor is just a real number, a 0-tensor field is the same as a real-valued function.

Lemma 8.7. *Let M be a smooth manifold, and suppose $\sigma \in \mathcal{T}^k(M)$, $\tau \in \mathcal{T}^l(M)$, and $f \in C^\infty(M)$. Then $f\sigma$ and $\sigma \otimes \tau$ are also smooth tensor fields, whose components in any local coordinate chart are*

$$(f\sigma)_{i_1\dots i_k} = f\sigma_{i_1\dots i_k},$$
$$(\sigma \otimes \tau)_{i_1\dots i_{k+l}} = \sigma_{i_1\dots i_k}\tau_{i_{k+1}\dots i_{k+l}}.$$

Exercise 8.7. Prove Lemma 8.7.

Pullbacks

Just like smooth covector fields, smooth covariant tensor fields can be pulled back by smooth maps to yield smooth tensor fields.

If $F\colon M \to N$ is a smooth map and σ is a smooth covariant k-tensor field on N, we define a k-tensor field $F^*\sigma$ on M, called the *pullback* of σ, by

$$(F^*\sigma)_p(X_1, \dots, X_k) = \sigma_{F(p)}(F_*X_1, \dots, F_*X_k).$$

Proposition 8.8. *Suppose $F\colon M \to N$ and $G\colon N \to P$ are smooth maps, $\sigma \in \mathcal{T}^k(N)$, $\tau \in \mathcal{T}^l(N)$, and $f \in C^\infty(N)$.*

(*a*) *F^* is linear over $\mathbb{R}$.*

(*b*) *$F^*(f\sigma) = (f \circ F)F^*\sigma$.*

(*c*) *$F^*(\sigma \otimes \tau) = F^*\sigma \otimes F^*\tau$.*

(*d*) *$(G \circ F)^* = F^* \circ G^*$.*

(*e*) *$\mathrm{Id}^*\tau = \tau$.*

Exercise 8.8. Prove Proposition 8.8.

If f is a smooth function (i.e., a 0-tensor field) and σ is a smooth k-tensor field, then it is consistent with our definitions to interpret $f \otimes \sigma$ as $f\sigma$, and F^*f as $f \circ F$. With these interpretations, property (b) of this proposition is really just a special case of (c).

Observe that properties (d) and (e) imply that the assignments $M \mapsto T^kM$ and $F \mapsto F^*$ yield a contravariant functor from the category of smooth manifolds to itself. Because of this, the convention of calling elements of $T^k M$ covariant tensors is particularly unfortunate; but this terminology is so deeply entrenched that one has no choice but to go along with it.

The following corollary is an immediate consequence of Proposition 8.8.

Corollary 8.9. *Let $F\colon M \to N$ be smooth, and let $\sigma \in \mathcal{T}^k(N)$. If $p \in M$ and (y^j) are coordinates for N on a neighborhood of $F(p)$, then $F^*\sigma$ has the following expression near p:*

$$F^*(\sigma_{j_1\dots j_k} dy^{j_1} \otimes \cdots \otimes dy^{j_k}) = (\sigma_{j_1\dots j_k} \circ F) d(y^{j_1} \circ F) \otimes \cdots \otimes d(y^{j_k} \circ F).$$

Therefore $F^*\sigma$ is smooth.

In words, this corollary just says that $F^*\sigma$ is computed by the same technique we described in Chapter 4 for computing the pullback of a covector field: Wherever you see y^j in the expression for σ, just substitute the jth component function of F and expand. We will see examples of this in the next section.

Symmetric Tensors

Symmetric tensors—those whose values are unchanged by rearranging their arguments—play an extremely important role in differential geometry. We will describe only covariant symmetric tensors, but similar considerations apply to contravariant ones.

It is useful to start, as usual, in the linear algebraic setting. Let V be a finite-dimensional vector space. A covariant k-tensor T on V is said to be *symmetric* if its value is unchanged by interchanging any pair of arguments:

$$T(X_1, \dots, X_i, \dots, X_j, \dots, X_k) = T(X_1, \dots, X_j, \dots, X_i, \dots, X_k)$$

whenever $1 \le i, j \le k$.

Exercise 8.9. Show that the following are equivalent for a covariant k-tensor T:

(a) T is symmetric.

(b) For any vectors $X_1, \dots, X_k \in V$, the value of $T(X_1, \dots, X_k)$ is unchanged when $X_1, \dots, X_k$ are rearranged in any order.

(c) The components $T_{i_1\dots i_k}$ of T with respect to any basis are unchanged by any permutation of the indices.

We denote the set of symmetric covariant k-tensors on V by $\Sigma^k(V)$. It is obviously a vector subspace of $T^k(V)$. There is a natural projection $\operatorname{Sym}\colon T^k(V) \to \Sigma^k(V)$ called *symmetrization*, defined as follows. First, let S_k denote the *symmetric group on k elements*, that is, the group of permutations of $\{1,\dots,k\}$. Given a k-tensor T and a permutation $\sigma \in S_k$, we define a new k-tensor T^σ by

$$T^\sigma(X_1,\dots,X_k) = T(X_{\sigma(1)},\dots,X_{\sigma(k)}).$$

Then we define $\operatorname{Sym} T$ by

$$\operatorname{Sym} T = \frac{1}{k!}\sum_{\sigma\in S_k} T^\sigma.$$

Lemma 8.10 (Properties of Symmetrization).

(a) *For any covariant tensor T, $\operatorname{Sym} T$ is symmetric.*

(b) *T is symmetric if and only if $\operatorname{Sym} T = T$.*

Proof. Suppose $T \in T^k(V)$. If $\tau \in S_k$ is any permutation, then

$$\begin{aligned}
(\operatorname{Sym} T)(X_{\tau(1)},\dots,X_{\tau(k)}) &= \frac{1}{k!}\sum_{\sigma\in S_k} T^\sigma(X_{\tau(1)},\dots,X_{\tau(k)})\\
&= \frac{1}{k!}\sum_{\sigma\in S_k} T^{\sigma\tau}(X_1,\dots,X_k)\\
&= \frac{1}{k!}\sum_{\eta\in S_k} T^{\eta}(X_1,\dots,X_k)\\
&= (\operatorname{Sym} T)(X_1,\dots,X_k),
\end{aligned}$$

where we have substituted $\eta = \sigma\tau$ in the second-to-last line and used the fact that η runs over all of S_k as σ does. This shows that $\operatorname{Sym} T$ is symmetric.

If T is symmetric, then Exercise 8.9 shows that $T^\sigma = T$ for every $\sigma \in S_k$, so it follows immediately that $\operatorname{Sym} T = T$. On the other hand, if $\operatorname{Sym} T = T$, then T is symmetric because part (a) shows that $\operatorname{Sym} T$ is. □

If S and T are symmetric tensors on V, then $S \otimes T$ is not symmetric in general. However, using the symmetrization operator, it is possible to define a new product that takes symmetric tensors to symmetric tensors. If $S \in \Sigma^k(V)$ and $T \in \Sigma^l(V)$, we define their *symmetric product* to be the $(k+l)$-tensor ST (denoted by juxtaposition with no intervening product symbol) given by

$$ST = \operatorname{Sym}(S \otimes T).$$

More explicitly, the action of ST on vectors $X_1,\dots,X_{k+l}$ is given by

$$\begin{aligned}
&ST(X_1,\dots,X_{k+l})\\
&\qquad = \frac{1}{(k+l)!}\sum_{\sigma\in S_{k+l}} S(X_{\sigma(1)},\dots,X_{\sigma(k)})T(X_{\sigma(k+1)},\dots,X_{\sigma(k+l)}).
\end{aligned}$$

Proposition 8.11 (Properties of the Symmetric Product).

(*a*) *The symmetric product is symmetric and bilinear: For all symmetric tensors R, S, T and all $a, b \in \mathbb{R}$,*

$$\begin{aligned} ST &= TS, \\ (aR + bS)T &= aRT + bST = T(aR + bS) \end{aligned}$$

(*b*) *If ω and η are covectors, then*

$$\omega\eta = \tfrac{1}{2}(\omega \otimes \eta + \eta \otimes \omega).$$

Exercise 8.10. Prove Proposition 8.11.

A *symmetric tensor field* on a manifold is simply a covariant tensor field whose value at any point is a symmetric tensor. The symmetric product of two or more tensor fields is defined pointwise, just like the tensor product.

Riemannian Metrics

The most important examples of symmetric tensors on a vector space are inner products (see the Appendix). Any inner product allows us to define lengths of vectors and angles between them, and thus to do Euclidean geometry.

Transferring these ideas to manifolds, we obtain one of the most important applications of tensors to differential geometry. Let M be a smooth manifold. A *Riemannian metric* on M is a smooth symmetric 2-tensor field that is positive definite at each point. A *Riemannian manifold* is a pair (M, g), where M is a smooth manifold and g is a Riemannian metric on M. One sometimes simply says "M is a Riemannian manifold" if M is understood to be endowed with a specific Riemannian metric.

Note that a Riemannian metric is not the same thing as a metric in the sense of metric spaces, although the two concepts are closely related, as we will see below. Because of this ambiguity, we will usually use the term "distance function" when considering a metric in the metric space sense, and reserve "metric" for a Riemannian metric. In any event, which type of metric is being considered should always be clear from the context.

If g is a Riemannian metric on M, then for each $p \in M$, g_p is an inner product on T_pM. Because of this, we will often use the notation $\langle X, Y\rangle_g$ to denote the real number $g_p(X, Y)$ for $X, Y \in T_pM$.

In any local coordinates (x^i), a Riemannian metric can be written

$$g = g_{ij}dx^i \otimes dx^j,$$

where g_{ij} is a symmetric positive definite matrix of smooth functions. Observe that the symmetry of g allows us to write g also in terms of symmetric products as follows:

$$
\begin{aligned}
g &= g_{ij} dx^i \otimes dx^j \\
&= \tfrac{1}{2}(g_{ij} dx^i \otimes dx^j + g_{ji} dx^i \otimes dx^j) && (\text{since } g_{ij} = g_{ji}) \\
&= \tfrac{1}{2}(g_{ij} dx^i \otimes dx^j + g_{ij} dx^j \otimes dx^i) && (\text{switch } i \leftrightarrow j \text{ in the second term}) \\
&= g_{ij} dx^i dx^j && (\text{definition of symmetric product}).
\end{aligned}
$$

Example 8.12. The simplest example of a Riemannian metric is the *Euclidean metric* $\overline{g}$ on $\mathbb{R}^n$, defined in standard coordinates by

$$\overline{g} = \delta_{ij} dx^i dx^j.$$

It is common to use the abbreviation ω^2 for the symmetric product of a tensor ω with itself, so the Euclidean metric can also be written

$$\overline{g} = (dx^1)^2 + \cdots + (dx^n)^2.$$

Applied to vectors $v, w \in T_p\mathbb{R}^n$, this yields

$$\overline{g}_p(v, w) = \delta_{ij} v^i w^j = \sum_{i=1}^{n} v^i w^i = v \cdot w.$$

In other words, $\overline{g}$ is the 2-tensor field whose value at each point is the Euclidean dot product. (As you may recall, we warned in Chapter 1 that expressions involving the Euclidean dot product are likely to violate our index conventions and therefore to require explicit summation signs. This can usually be avoided by writing the metric coefficients δ_{ij} explicitly, as in $\delta_{ij} v^i w^j$.)

To transform a Riemannian metric under a change of coordinates, we use the same technique as we used for covector fields: Think of the change of coordinates as the identity map expressed in terms of different coordinates for the domain and range, and use the formula of Corollary 8.9. As before, in practice this just amounts to substituting the formulas for one set of coordinates in terms of the other.

Example 8.13. To illustrate, let us compute the coordinate expression for the Euclidean metric on $\mathbb{R}^2$ in polar coordinates. The Euclidean metric is $\overline{g} = dx^2 + dy^2$. (By convention, the notation dx^2 means the symmetric product $dx\, dx$, not $d(x^2)$). Substituting $x = r\cos\theta$ and $y = r\sin\theta$ and expanding, we obtain

$$\begin{aligned}\overline{g} &= dx^2 + dy^2 \\ &= d(r\cos\theta)^2 + d(r\sin\theta)^2 \\ &= (\cos\theta\, dr - r\sin\theta\, d\theta)^2 + (\sin\theta\, dr + r\cos\theta\, d\theta)^2 \\ &= (\cos^2\theta + \sin^2\theta)dr^2 + (r^2\sin^2\theta + r^2\cos^2\theta)d\theta^2 \\ &\quad + (-2r\cos\theta\sin\theta + 2r\sin\theta\cos\theta)dr\, d\theta \\ &= dr^2 + r^2\, d\theta^2.\end{aligned} \tag{8.4}$$

Below are just a few of the geometric constructions that can be defined on a Riemannian manifold (M, g).

- The *length* or *norm* of a tangent vector $X \in T_pM$ is defined to be
$$|X|_g = \langle X, X\rangle_g^{1/2} = g_p(X, X)^{1/2}.$$
- The *angle* between two nonzero tangent vectors $X, Y \in T_pM$ is the unique $\theta \in [0, \pi]$ satisfying
$$\cos\theta = \frac{\langle X, Y\rangle_g}{|X|_g\,|Y|_g}.$$
- Two tangent vectors $X, Y \in T_pM$ are said to be *orthogonal* if $\langle X, Y\rangle_g = 0$.
- If $\gamma\colon [a, b] \to M$ is a piecewise smooth curve segment, the *length* of γ is
$$L_g(\gamma) = \int_a^b |\gamma'(t)|_g\, dt.$$
Because $|\gamma'(t)|_g$ is continuous at all but finitely many values of t, the integral is well-defined.

Exercise 8.11. If $\gamma\colon [a, b] \to M$ is a piecewise smooth curve segment and $a < c < b$, show that
$$L_g(\gamma) = L_g\left(\gamma|_{[a,c]}\right) + L_g\left(\gamma|_{[c,b]}\right).$$

It is an extremely important fact that length is independent of parametrization in the following sense. In chapter 4, we defined a reparametrization of a smooth curve segment $\gamma\colon [a, b] \to M$ to be a curve segment of the form $\widetilde{\gamma} = \gamma \circ \varphi$, where $\varphi\colon [c, d] \to [a, b]$ is a diffeomorphism. More generally, if γ is piecewise smooth, we allow φ to be a homeomorphism whose restriction to each subinterval $[c_{i-1}, c_i]$ is a diffeomorphism onto its image, where $c = c_0 < c_1 < \cdots < c_k = d$ is some finite subdivision of $[c, d]$.

Proposition 8.14 (Parameter Independence of Length). *Let (M, g) be a Riemannian manifold, and let $\gamma\colon [a, b] \to M$ be a piecewise smooth curve segment. If $\widetilde{\gamma}$ is any reparametrization of γ, then $L_g(\widetilde{\gamma}) = L_g(\gamma)$.*

Proof. First suppose that γ is smooth, and $\varphi\colon [c, d] \to [a, b]$ is a diffeomorphism such that $\widetilde{\gamma} = \gamma \circ \varphi$. The fact that φ is a diffeomorphism implies that either $\varphi' > 0$ or $\varphi' < 0$ everywhere. Let us assume first that $\varphi' > 0$. We have

$$\begin{aligned}
L_g(\widetilde{\gamma}) &= \int_c^d |\widetilde{\gamma}'(t)|_g \, dt \\
&= \int_c^d \left| \frac{d}{dt}(\gamma \circ \varphi)(t) \right|_g \, dt \\
&= \int_c^d |\varphi'(t)\gamma'(\varphi(t))|_g \, dt \\
&= \int_c^d |\gamma'(\varphi(t))|_g \, \varphi'(t) \, dt \\
&= \int_a^b |\gamma'(s)|_g \, ds \\
&= L_g(\gamma),
\end{aligned}$$

where the second-to-last equality follows from the change of variables formula for ordinary integrals.

In case $\varphi' < 0$, we just need to introduce two sign changes into the above calculation. The sign changes once when $\varphi'(t)$ is moved outside the absolute value signs, because $|\varphi'(t)| = -\varphi'(t)$. Then it changes again in the last step, because φ reverses the direction of the integral. Since the two sign changes cancel each other, the result is the same.

If γ and φ are only piecewise smooth, we can subdivide $[c, d]$ into finitely many subintervals on which both $\widetilde{\gamma}$ and φ are smooth, and then the result follows by applying the above argument on each such subinterval. □

Suppose (M, g) and $(\widetilde{M}, \widetilde{g})$ are Riemannian manifolds. A smooth map $F\colon M \to \widetilde{M}$ is called an *isometry* if it is a diffeomorphism that satisfies $F^*\widetilde{g} = g$. If there exists an isometry between M and $\widetilde{M}$, we say that M and $\widetilde{M}$ are *isometric* as Riemannian manifolds. More generally, F is called a *local isometry* if every point $p \in M$ has a neighborhood U such that $F|_U$ is an isometry of U onto an open subset of $\widetilde{M}$. A metric g on M is said to be *flat* if every point $p \in M$ has a neighborhood $U \subset M$ such that $(U, g|_U)$ is isometric to an open subset of $\mathbb{R}^n$ with the Euclidean metric.

Riemannian geometry is the study of properties of Riemannian manifolds that are invariant under isometries. See, for example, [Lee97] for an introduction to some of its main ideas and techniques.

Exercise 8.12. Show that lengths of curves are isometry invariants of Riemannian manifolds. More precisely, suppose (M, g) and $(\widetilde{M}, \widetilde{g})$ are Riemannian manifolds, and $F\colon M \to \widetilde{M}$ is an isometry. Show that $L_{\widetilde{g}}(F \circ \gamma) = L_g(\gamma)$ for any piecewise smooth curve segment γ in M.

Another extremely useful tool on Riemannian manifolds is orthonormal frames. Let (M, g) be an n-dimensional Riemannian manifold. A local frame $(E_1, \dots, E_n)$ for M defined on some open subset $U \subset M$ is said to be *orthonormal* if $(E_1|_p, \dots, E_n|_p)$ is an orthonormal basis for T_pM at each point $p \in U$, or in other words if $\langle E_i, E_j \rangle_g = \delta_{ij}$.

Example 8.15. The coordinate frame $(\partial/\partial x^i)$ is a global orthonormal frame on $\mathbb{R}^n$.

Proposition 8.16 (Existence of Orthonormal Frames). *Let (M, g) be a Riemannian manifold. For any $p \in M$, there is a smooth orthonormal frame on a neighborhood of p.*

Proof. Let (x^i) be any coordinates on a neighborhood U of p. Applying the Gram-Schmidt algorithm (Proposition A.17) to the coordinate frame $(\partial/\partial x^i)$, we obtain a new frame (E_i), given inductively by the formula

$$E_j = \frac{\partial/\partial x^j - \sum_{i=1}^{j-1} \langle E_j, E_i \rangle_g E_i}{\left| \partial/\partial x^j - \sum_{i=1}^{n-1} \langle E_j, E_i \rangle_g E_i \right|_g}.$$

Because $\operatorname{span}(E_1, \dots, E_{j-1}) = \operatorname{span}(\partial/\partial x^1, \dots, \partial/\partial x^{j-1})$, the vector whose norm appears in the demoninator above is nowhere zero on U. Thus this formula defines E_j as a smooth vector field on U, and a computation shows that the resulting frame (E_i) is orthonormal. □

Observe that Proposition 8.16 did *not* show that there are coordinates near p for which the *coordinate frame* is orthonormal. Problem 8-13 shows that there are such coordinates in a neighborhood of each point only if the metric is flat.

The Riemannian Distance Function

Using curve segments as "measuring tapes," we can define a notion of distance between points on a Riemannian manifold. If (M, g) is a connected Riemannian manifold and $p, q \in M$, the *(Riemannian) distance* between p and q, denoted by $d_g(p, q)$, is defined to be the infimum of $L_g(\gamma)$ over all piecewise smooth curve segments γ from p to q. Because any pair of points in a connected manifold can be joined by a piecewise smooth curve segment (Lemma 4.16), this is well-defined.

Example 8.17. On $\mathbb{R}^n$ with the Euclidean metric $\overline{g}$, one can show that any straight line segment is the shortest piecewise smooth curve segment between its endpoints (Problem 8-14). Therefore, the distance function $d_{\overline{g}}$ is equal to the usual Euclidean distance:

$$d_{\overline{g}}(x, y) = |x - y|.$$

Exercise 8.13. If (M, g) and $(\widetilde{M}, \widetilde{g})$ are connected Riemannian manifolds and $F \colon M \to \widetilde{M}$ is an isometry, show that $d_{\widetilde{g}}(F(p), F(q)) = d_g(p, q)$ for all $p, q \in M$.

We will see below that the Riemannian distance function turns M into a metric space whose topology is the same as the given manifold topology. The key is the following technical lemma, which shows that any Riemannian metric is locally comparable to the Euclidean metric in coordinates.

Lemma 8.18. *Let g be any Riemannian metric on an open set $U \subset \mathbb{R}^n$. For any compact subset $K \subset U$, there exist positive constants c, C such that for all $x \in K$ and all $v \in T_xM$,*

$$c|v|_{\overline{g}} \le |v|_g \le C|v|_{\overline{g}}. \tag{8.5}$$

Proof. For any compact subset $K \subset U$, let $L \subset T\mathbb{R}^n$ be the set

$$L = \{(x, v) \in T\mathbb{R}^n : x \in K,\ |v|_{\overline{g}} = 1\}.$$

Since L is a product of compact sets in $T\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$, L is compact. Because the norm $|v|_g$ is continuous and strictly positive on L, there are positive constants c, C such that $c \le |v|_g \le C$ whenever $(x, v) \in L$. If $x \in K$ and v is any nonzero vector in $T_x\mathbb{R}^n$, let $\lambda = |v|_{\overline{g}}$. Then $(x, \lambda^{-1}v) \in L$, so by homogeneity of the norm,

$$|v|_g = \lambda|\lambda^{-1}v|_g \le \lambda C = C|v|_{\overline{g}}.$$

A similar computation shows that $|v|_g \ge c|v|_{\overline{g}}$. The same inequalities are trivially true when $v = 0$. □

Proposition 8.19 (Riemannian Manifolds as Metric Spaces). *Let (M, g) be a connected Riemannian manifold. With the Riemannian distance function, M is a metric space whose metric topology is the same as the original manifold topology.*

Proof. It is immediate from the definition that $d_g(p, q) \ge 0$ for any $p, q \in M$. Because any constant curve segment has length zero, it follows that $d_g(p, p) = 0$, and $d_g(p, q) = d_g(q, p)$ follows from the fact that any curve segment from p to q can be reparametrized to go from q to p. Suppose γ_1 and γ_2 are piecewise smooth curve segments from p to q and q to r, respectively, and let γ be a piecewise smooth curve segment that first follows γ_1 and then follows γ_2 (reparametrized if necessary). Then

$$d_g(p,r) \leq L_g(\gamma) = L_g(\gamma_1) + L_g(\gamma_2).$$

Taking the infimum over all such γ_1 and γ_2, we find that $d_g(p,r) \leq d_g(p,q) + d_g(q,r)$. (This is one reason why it is important to define the distance function using piecewise smooth curves instead of just smooth ones.)

To complete the proof that (M, d_g) is a metric space, we need only show that $d_g(p,q) > 0$ if $p \neq q$. For this purpose, let $p, q \in M$ be distinct points, and let U be any coordinate domain containing p but not q. Use the coordinate map as usual to identify U with an open subset in $\mathbb{R}^n$, and let $\overline{g}$ denote the Euclidean metric in these coordinates. If V is a coordinate ball of radius ε centered at p such that $\overline{V} \subset U$, Lemma 8.18 shows that there are positive constants c, C such that

$$c|X|_{\overline{g}} \leq |X|_g \leq C|X|_{\overline{g}}$$

whenever $q \in \overline{V}$ and $X \in T_qM$. Then for any piecewise smooth curve segment γ lying entirely in $\overline{V}$, it follows that

$$cL_{\overline{g}}(\gamma) \leq L_g(\gamma) \leq CL_{\overline{g}}(\gamma).$$

Suppose $\gamma\colon [a,b] \to M$ is a piecewise smooth curve segment from p to q. Let t_0 be the infimum of all $t \in [a,b]$ such that $\gamma(t) \notin \overline{V}$. It follows that $\gamma(t_0) \in \partial V$ by continuity, and $\gamma(t) \in \overline{V}$ for $a \leq t \leq t_0$. Thus

$$L_g(\gamma) \geq L_g\left(\gamma|_{[a,t_0]}\right) \geq cL_{\overline{g}}\left(\gamma|_{[a,t_0]}\right) \geq cd_{\overline{g}}(p, \gamma(t_0)) = c\varepsilon.$$

Taking the infimum over all such γ, we conclude that $d_g(p,q) \geq c\varepsilon > 0$.

Finally, to show that the metric topology generated by d_g is the same as the given manifold topology on M, we will show that the open sets in the manifold topology are open in the metric topology and vice versa. Suppose first that $U \subset M$ is open in the manifold topology. Let p be any point of U, and let V be a coordinate ball of radius ε around p such that $\overline{V} \subset U$ as above. The argument in the previous paragraph shows that $d_g(p,q) \geq c\varepsilon$ whenever $q \notin \overline{V}$. The contrapositive of this statement is that $d_g(p,q) < c\varepsilon$ implies $q \in \overline{V} \subset U$, or in other words the metric ball of radius $c\varepsilon$ around p is contained in U. This shows that U is open in the metric topology.

Conversely, suppose that W is open in the metric topology, and let $p \in W$. Let $\overline{V}$ be any closed coordinate ball around p, let $\overline{g}$ be the Euclidean metric on $\overline{V}$ determined by the given coordinates, and let c, C be positive constants such that (8.5) is satisfied for $X \in T_qM$, $q \in \overline{V}$. For any $\varepsilon > 0$, let V_ε be the set of points whose Euclidean distance from p is less than ε. If $q \in V_\varepsilon$, let γ be the straight-line segment in coordinates from p to q. Arguing as above, (8.5) implies

$$d_g(p,q) \leq L_g(\gamma) \leq CL_{\overline{g}}(\gamma) = C\varepsilon.$$

If we choose ε small enough that the closed metric ball of radius $C\varepsilon$ around p is contained in W, this shows that $V_\varepsilon \subset W$. Since V_ε is a neighborhood of p in the manifold topology, this shows that W is open in the manifold topology as well. □

A topological space is said to be *metrizable* if it admits a distance function whose metric topology is the same as the given topology. The next corollary is an immediate consequence of the preceding proposition.

Corollary 8.20. *Every smooth manifold is metrizable.*

Riemannian Submanifolds

If (M, g) is a Riemannian manifold and $S \subset M$ is an immersed submanifold, we can define a smooth symmetric 2-tensor $g|_S$ on S by $g|_S = \iota^* g$, where $\iota\colon S \hookrightarrow M$ is the inclusion map. By definition, this means for $X, Y \in T_pS$

$$(g|_S)(X, Y) = \iota^* g(X, Y) = g(\iota_* X, \iota_* Y) = g(X, Y),$$

so $g|_S$ is just the restriction of g to vectors tangent to S. Since the restriction of an inner product to a subspace is still positive definite, $g|_S$ is a Riemannian metric on S, called the *induced metric*. In this case, S is called a *Riemannian submanifold* of M.

Example 8.21. The metric $\mathring{g} = \overline{g}|_{\mathbb{S}^n}$ induced on $\mathbb{S}^n$ from the Euclidean metric by the usual inclusion $\mathbb{S}^n \hookrightarrow \mathbb{R}^{n+1}$ is called the *round metric* on the sphere.

If S is a Riemannian submanifold of (M, g), it is usually easiest to compute the induced metric $g|_S$ in terms of a *local parametrization* of S, which is a smooth embedding $X\colon U \to M$ whose image is an open subset of S. The coordinate representation of $g|_S$ with respect to the coordinate chart $\varphi = X^{-1}$ is then the pullback metric $X^* g$. The next two examples will illustrate the procedure.

Example 8.22 (Riemannian Metrics in Graph Coordinates). Let $U \subset \mathbb{R}^n$ be an open set, and let $M \subset \mathbb{R}^{n+1}$ be the graph of the smooth function $f\colon U \to \mathbb{R}$. Then the map $X\colon U \to \mathbb{R}^{n+1}$ given by $X(u^1, \dots, u^n) = (u^1, \dots, u^n, f(u))$ is a (global) parametrization of M, and the induced metric on M is given in graph coordinates by

$$\begin{aligned} X^*\overline{g} &= X^*((dx^1)^2 + \cdots + (dx^{n+1})^2) \\ &= (du^1)^2 + \cdots + (du^n)^2 + df^2. \end{aligned}$$

Example 8.23. Let $D \subset \mathbb{R}^3$ be the embedded torus obtained by revolving the circle $(y-2)^2 + z^2 = 1$ around the z-axis. If $X\colon \mathbb{R}^2 \to \mathbb{R}^3$ is the map

$$X(\varphi, \theta) = ((2 + \cos\varphi)\cos\theta, (2 + \cos\varphi)\sin\theta, \sin\varphi),$$

then the restriction of X to any sufficiently small open set $U \subset \mathbb{R}^2$ is a local parametrization of D. The metric induced on D by the Euclidean metric is computed as follows:

$$\begin{aligned}
X^*\overline{g} &= X^*(dx^2 + dy^2 + dz^2) \\
&= d((2+\cos\varphi)\cos\theta)^2 + d((2+\cos\varphi)\sin\theta)^2 + d(\sin\varphi)^2 \\
&= (-\sin\varphi\cos\theta\, dt - (2+\cos\varphi)\sin\theta\, d\theta)^2 \\
&\quad + (-\sin\varphi\sin\theta\, dt + (2+\cos\varphi)\cos\theta\, d\theta)^2 \\
&\quad + (\cos\varphi\, dt)^2 \\
&= (\sin^2\varphi\cos^2\theta + \sin^2\varphi\sin^2\theta + \cos^2\varphi)d\varphi^2 \\
&\quad + ((2+\cos\varphi)\sin\varphi\cos\theta\sin\theta - (2+\cos\varphi)\sin\varphi\cos\theta\sin\theta)d\varphi\, d\theta \\
&\quad + ((2+\cos\varphi)^2\sin^2\theta + (2+\cos\varphi)^2\cos^2\theta)d\theta^2 \\
&= d\varphi^2 + (2+\cos\varphi)^2 d\theta^2.
\end{aligned}$$

If (M, g) is an n-dimensional Riemannian manifold and $S \subset M$ is a k-dimensional Riemannian submanifold, a local orthonormal frame $(E_1, \dots, E_n)$ for M on an open set $U \subset M$ is said to be *adapted* to S if $(E_1|_p, \dots, E_k|_p)$ is an orthonormal basis for T_pS at each $p \in U \cap S$.

Proposition 8.24 (Existence of Adapted Orthonormal Frames). *Let $S \subset M$ be an embedded Riemannian submanifold of the Riemannian manifold (M, g). For each $p \in S$, there is an adapted orthonormal frame on a neighborhood U of p in M.*

Proof. Let $(x^1, \dots, x^n)$ be slice coordinates for S on a neighborhood U of p, so that $S \cap U$ is the set where $x^{k+1} = \cdots = x^n = 0$. Applying the Gram-Schmidt algorithm to the frame $(\partial/\partial x^i)$, we obtain an orthonormal frame $(E_1, \dots, E_n)$ with the property that $\operatorname{span}(E_1|_p, \dots, E_k|_p) = \operatorname{span}(\partial/\partial x^1|_p, \dots, \partial/\partial x^k|_p) = T_pS$ at each $p \in S$. □

The Tangent-Cotangent Isomorphism

Another very important feature of Riemannian metrics is that they provide a natural correspondence between tangent and cotangent vectors. Given a Riemannian metric g on a manifold M, define a bundle map $\widetilde{g}\colon TM \to T^*M$ by

$$\widetilde{g}(X)(Y) = g_p(X, Y) \quad \text{for } X, Y \in T_pM.$$

(Recall that a bundle map is a smooth map whose restriction to each fiber is a linear map from T_pM to T_p^*M.)

Exercise 8.14. Show that $\widetilde{g}$ is a bundle map.

Note that $\widetilde{g}$ is injective, because $\widetilde{g}(X) = 0$ implies $0 = \widetilde{g}(X)(X) = g_p(X,X)$, which in turn implies $X = 0$. For dimensional reasons, therefore, $\widetilde{g}$ is bijective, and so it is a bundle isomorphism (see Problem 5-12).

In coordinates,

$$\widetilde{g}(X)(Y) = g_{ij}(p)X^iY^j,$$

which implies that the covector $\widetilde{g}(X)$ has the coordinate expression

$$\widetilde{g}(X) = g_{ij}(p)X^i dy^j.$$

In other words, the restriction of $\widetilde{g}$ to T_pM is the linear map whose matrix with respect to the coordinate bases for T_pM and T_p^*M is just the same as the matrix of g.

It is customary to denote the components of the covector $\widetilde{g}(X)$ by

$$X_j = g_{ij}(p)X^i,$$

so that

$$\widetilde{g}(X) = X_j dy^j.$$

Because of this, one says that $\widetilde{g}(X)$ is obtained from X by *lowering an index*. The notation $X^\flat$ is frequently used for $\widetilde{g}(X)$, because the symbol $\flat$ ("flat") is used in musical notation to indicate that a tone is to be lowered. Similarly, the inverse map $\widetilde{g}^{-1}\colon T_p^*M \to T_pM$ is represented by the inverse of the matrix (g_{ij}). The components of this inverse matrix are usually denoted by g^{ij}, so that

$$g^{ij}g_{jk} = g_{kj}g^{ji} = \delta^i_k.$$

Thus for a cotangent vector $\xi \in T_p^*M$, $\widetilde{g}^{-1}(\xi)$ has the coordinate representation

$$\widetilde{g}^{-1}(\xi) = \xi^i \frac{\partial}{\partial x^i}, \qquad \text{where } \xi^i = g^{ij}(p)\xi_j.$$

We use the notation $\xi^\#$ ("ξ-sharp") for $\widetilde{g}^{-1}(\xi)$, and say that $\xi^\#$ is obtained from ξ by *raising an index.*

The most important use of the sharp operation is to recover the notion of the gradient as a vector field on Riemannian manifolds. For any smooth function f on a Riemannian manifold (M,g), we define a vector field $\operatorname{grad} f$, called the *gradient* of f, by

$$\operatorname{grad} f = (df)^\# = \widetilde{g}^{-1}(df).$$

Unraveling the definitions, for any $X \in T_pM$, it satisfies

$$\langle \operatorname{grad} f|_p, X\rangle_g = \widetilde{g}(\operatorname{grad} f|_p)(X) = df_p(X) = Xf.$$

Thus $\operatorname{grad} f$ is the unique vector field that satisfies

$$\langle \operatorname{grad} f, X \rangle_g = Xf \quad \text{for every vector field } X,$$

or equivalently,

$$\langle \operatorname{grad} f, \cdot \rangle_g = df.$$

In coordinates, $\operatorname{grad} f$ has the expression

$$\operatorname{grad} f = g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^j}.$$

In particular, on $\mathbb{R}^n$ with the Euclidean metric, this is just

$$\operatorname{grad} f = \delta^{ij} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^i}.$$

Thus our new definition of the gradient in this case coincides with the gradient from elementary calculus, which is the vector field whose components are the partial derivatives of f. In other coordinates, however, the gradient will not generally have the same form.

Example 8.25. Let us compute the gradient of a function $f \in C^\infty(\mathbb{R}^2)$ in polar coordinates. From (8.4), we see that the matrix of $\overline{g}$ in polar coordinates is $\left(\begin{smallmatrix} 1 & 0 \\ 0 & r^2 \end{smallmatrix}\right)$, so its inverse matrix is $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/r^2 \end{smallmatrix}\right)$. Inserting this into the formula for the gradient, we obtain

$$\operatorname{grad} f = \frac{\partial f}{\partial r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \frac{\partial}{\partial \theta}.$$

Existence of Riemannian Metrics

We end this section by proving the following important result.

Proposition 8.26 (Existence of Riemannian Metrics). *Every smooth manifold admits a Riemannian metric.*

Proof. We give two proofs. For the first, we begin by covering M by coordinate charts $(U_\alpha, \varphi_\alpha)$. In each coordinate domain, there is a Riemannian metric g_α given by the Euclidean metric $\delta_{ij} dx^i dx^j$ in coordinates. Now let $\{\psi_\alpha\}$ be a partition of unity subordinate to the cover $\{U_\alpha\}$, and define

$$g = \sum_\alpha \psi_\alpha g_\alpha.$$

Because of the local finiteness condition for partitions of unity, there are only finitely many nonzero terms in a neighborhood of any point, so this expression defines a smooth tensor field. It is obviously symmetric, so only positivity needs to be checked. If $X \in T_pM$ is any nonzero vector, then

$$g_p(X, X) = \sum_\alpha \psi_\alpha(p) g_\alpha|_p(X, X).$$

This sum is nonnegative, because each term is nonnegative. At least one of the functions ψ_α is strictly positive at p (because they sum to 1). Because $g_\alpha|_p(X, X) > 0$, it follows that $g_p(X, X) > 0$.

The second proof is shorter, but relies on the Whitney embedding theorem, which is far less elementary. We simply embed M in $\mathbb{R}^N$ for some N, and then the Euclidean metric induces a Riemannian metric $\overline{g}|_M$ on M. □

Pseudo-Riemannian Metrics

An important generalization of Riemannian metrics is obtained by relaxing the requirement that the metric be positive definite. A 2-tensor g on a vector space V is said to be *nondegenerate* if it satisfies any of the following three equivalent conditions:

- $g(X, Y) = 0$ for all $Y \in V$ if and only if $X = 0$.
- The map $\widetilde{g}\colon V \to V^*$ defined by $\widetilde{g}(X)(Y) = g(X, Y)$ is invertible.
- The matrix of g with respect to any basis is nonsingular.

Just as any inner product can be transformed to the Euclidean one by switching to an orthonormal basis, every nondegenerate symmetric 2-tensor can be transformed by a change of basis to one whose matrix is diagonal with all entries equal to ± 1. The numbers of positive and negative diagonal entries are independent of the choice of basis; thus the *signature* of g, defined as the sequence $(-1, \ldots, -1, +1, \ldots, +1)$ of diagonal entries in nondecreasing order, is an invariant of g.

A *pseudo-Riemannian metric* on a manifold M is a smooth symmetric 2-tensor field that is nondegenerate at each point. Pseudo-Riemannian metrics whose signature is $(-1, +1, \ldots, +1)$ are called *Lorentz metrics*; they play a central role in physics, where they are used to model gravitation in Einstein's general theory of relativity.

We will not pursue the subject of pseudo-Riemannian metrics any further, except to note that neither of the proofs above of the existence of Riemannian metrics carries over to the pseudo-Riemannian case: in particular, it is not always true that the restriction of a nondegenerate 2-tensor to a subspace is nondegenerate, nor is it true that a linear combination of nondegenerate 2-tensors with positive coefficients is necessarily nondegenerate. Indeed, it is not true that every manifold admits a Lorentz metric.

Problems

8-1. Let V and W be finite-dimensional real vector spaces. Show that the tensor product $V \otimes W$ is uniquely determined up to canonical isomorphism by its characteristic property (Proposition 8.3). More precisely, suppose $\widetilde{\pi}\colon V \times W \to Z$ is a bilinear map into a vector space Z with following property: For any bilinear map $A\colon V \times W \to Y$, there is a unique linear map $\widetilde{A}\colon Z \to Y$ such that the following diagram commutes:

$$\begin{tikzcd} V \times W \arrow[r, "A"] \arrow[d, "\widetilde{\pi}"'] & Y \\ Z. \arrow[ur, "\widetilde{A}"'] & \end{tikzcd}$$

Then there is a unique isomorphism $\Phi\colon V \otimes W \to Z$ such that $\widetilde{\pi} = \Phi \circ \pi$. [This shows that the details of the construction used to define the tensor product are irrelevant, as long as the resulting space satisfies the characteristic property.]

8-2. If V is any finite-dimensional real vector space, prove that there are canonical isomorphisms $\mathbb{R} \otimes V \cong V \cong V \otimes \mathbb{R}$.

8-3. Let V and W be finite-dimensional real vector spaces. Prove that there is a canonical (basis-independent) isomorphism between $V^* \otimes W$ and the space $\operatorname{Hom}(V, W)$ of linear maps from V to W.

8-4. Let M be a smooth n-manifold, and σ a covariant k-tensor field on M. If (x^i) and $(\widetilde{x}^j)$ are overlapping coordinate charts on M, we can write

$$\sigma = \sigma_{i_1 \dots i_k} dx^{i_1} \otimes \cdots \otimes dx^{i_k} = \sigma = \widetilde{\sigma}_{j_1 \dots j_k} d\widetilde{x}^{j_1} \otimes \cdots \otimes d\widetilde{x}^{j_k}.$$

Compute a transformation law analogous to (4.4) expressing the component functions $\sigma_{i_1 \dots i_k}$ in terms of $\widetilde{\sigma}_{j_1 \dots j_k}$.

8-5. Generalize the change of coordinate formula of Problem 8-4 to mixed tensors of any rank.

8-6. Let M be a smooth manifold.

(a) Given a smooth covariant k-tensor field $\tau \in \mathcal{T}^k(M)$, show that the map $\mathcal{T}(M) \times \cdots \times \mathcal{T}(M) \to C^\infty(M)$ defined by

$$(X_1, \dots, X_k) \mapsto \tau(X_1, \dots, X_k)$$

is multilinear over $C^\infty(M)$, in the sense that for any smooth functions $f, f' \in C^\infty(M)$ and smooth vector fields X_i, X_i',

$$\begin{aligned}\tau(X_1, \dots, fX_i + f'X_i', \dots, X_k)&\\ &= f\tau(X_1, \dots, X_i, \dots, X_k) + f'\tau(X_1, \dots, X_i', \dots, X_k).\end{aligned}$$

(b) Show that a map

$$\widetilde{\tau}\colon \mathfrak{T}(M) \times \cdots \times \mathfrak{T}(M) \to C^\infty(M)$$

is induced by a smooth tensor field as above if and only if it is multilinear over $C^\infty(M)$.

8-7. Let V be an n-dimensional real vector space. Show that

$$\dim \Sigma^k(V) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}.$$

8-8. (a) Let T be a covariant k-tensor on a finite-dimensional real vector space V. Show that $\operatorname{Sym} T$ is the unique symmetric k-tensor satisfying

$$(\operatorname{Sym} T)(X, \dots, X) = T(X, \dots, X)$$

for all $X \in V$.

(b) Show that the symmetric product is associative: For all symmetric tensors R, S, T,

$$(RS)T = R(ST).$$

(c) If $\omega^1, \dots, \omega^k$ are covectors, show that

$$\omega^1 \cdots \omega^k = \frac{1}{k!} \sum_{\sigma \in S_k} \omega^{\sigma(1)} \otimes \cdots \otimes \omega^{\sigma(k)}.$$

8-9. Let $\mathring{g} = \overline{g}|_{\mathbb{S}^n}$ denote the round metric on the n-sphere, i.e., the metric induced from the Euclidean metric by the usual inclusion of $\mathbb{S}^n$ into $\mathbb{R}^{n+1}$.

(a) Derive an expression for $\mathring{g}$ in stereographic coordinates by computing the pullback $(\sigma^{-1})^*\overline{g}$.

(b) In the case $n = 2$, do the analogous computation in spherical coordinates $(x, y, z) = (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$.

8-10. Let M be any smooth manifold.

(a) Show that TM and T^*M are isomorphic vector bundles.

(b) Show that the isomorphism of part (a) is *not* canonical, in the following sense: There does not exist a rule that assigns to every smooth manifold M a bundle isomorphism $\lambda_M\colon TM \to T^*M$ in such a way that for every smooth map $F\colon M \to N$, the following diagram commutes:

$$\begin{tikzcd} TM \arrow[r, "F_*"] \arrow[d, "\lambda_M"'] & TN \arrow[d, "\lambda_N"] \\ T^*M & T^*N. \arrow[l, "F^*"] \end{tikzcd}$$

8-11. Let Γ be a discrete group acting smoothly, freely, and properly on a smooth manifold $\widetilde{M}$, and let $M = \widetilde{M}/\Gamma$. Show that a Riemannian metric $\widetilde{g}$ on $\widetilde{M}$ is the pullback of a metric on M by the quotient map $\pi\colon \widetilde{M} \to M$ if and only if $\widetilde{g}$ is invariant under Γ (i.e., $\gamma^*\widetilde{g} = \widetilde{g}$ for every $\gamma \in \Gamma$).

8-12. Let (M, g) and $(\widetilde{M}, \widetilde{g})$ be Riemannian manifolds. Suppose $F\colon M \to \widetilde{M}$ is a smooth map such that $F^*\widetilde{g} = g$. Show that F is an immersion.

8-13. Let (M, g) be a Riemannian manifold. Show that the following are equivalent:

(a) Each point of M has a coordinate neighborhood in which the coordinate frame is orthonormal.

(b) g is flat.

8-14. Show that the shortest path between two points in Euclidean space is a straight line. More precisely, for $x, y \in \mathbb{R}^n$, let $\gamma\colon [0,1] \to \mathbb{R}^n$ be the curve segment

$$\gamma(t) = (1-t)x + ty,$$

and show that any other piecewise smooth curve segment $\widetilde{\gamma}$ from x to y satisfies $L_{\overline{g}}(\widetilde{\gamma}) \geq L_{\overline{g}}(\gamma)$. [Hint: First consider the case in which both x and y lie on the x^1-axis.]

8-15. Let $M = \mathbb{R}^2 \smallsetminus \{0\}$ with the Euclidean metric $\overline{g}$, and let $p = (1, 0)$, $q = (-1, 0)$. Show that there is no piecewise smooth curve segment γ from p to q in M such that $L_{\overline{g}}(\gamma) = d_{\overline{g}}(p, q)$.

8-16. Let (M, g) be a Riemannian manifold, and let $f \in C^\infty(M)$.

(a) For any $p \in M$, show that among all unit vectors $X \in T_pM$, the directional derivative Xf is greatest when X points in the same direction as $\operatorname{grad} f|_p$, and the length of $\operatorname{grad} f|_p$ is equal to the value of the directional derivative in that direction.

(b) If p is a regular point of f, show that $\operatorname{grad} f|_p$ is orthogonal to the level set of f through p.

8-17. Let $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1 \subset \mathbb{C}^n$, and let g be the metric on $\mathbb{T}^n$ induced from the Euclidean metric on $\mathbb{C}^n$ (identified with $\mathbb{R}^{2n}$). Show that g is flat.

8-18. Let (M, g) be a Riemannian manifold and let $S \subset M$ be a Riemannian submanifold. If $p \in S$, a vector $N \in T_pM$ is said to be *normal to S* if N is orthogonal to T_pS with respect to g. Show that the set of all vectors normal to S is a smooth vector bundle over S, called the *normal bundle* to S. [Hint: use adapted orthonormal frames.]

8-19. If $S \subset M$ is an embedded submanifold, a smooth map $N\colon S \to TM$ such that $N_p \in T_pM$ for each $p \in S$ is called a *vector field along S*. If (M, g) is a Riemannian manifold and $S \subset M$ is a Riemannian submanifold of codimension 1, show that every $p \in S$ has a neighborhood on which there exist exactly two unit-length vector fields along S that are normal to S.

9

Differential Forms

In the previous chapter, we introduced symmetric tensors—those whose values are unchanged by interchanging any pair of arguments. In this chapter, we explore the complementary notion of alternating tensors, whose values change sign whenever two arguments are interchanged. The main focus of the chapter is differential forms, which are just alternating tensor fields. These innocent-sounding objects play an unexpectedly important role in smooth manifold theory, through two applications. First, as we will see in Chapter 10, they are the objects that can be integrated in a coordinate-independent way over manifolds or submanifolds; second, as we explore in Chapter 11, they provide a link between analysis and topology by way of the de Rham theorem.

We begin the chapter with a heuristic discussion of the measurement of volume, to motivate the central role played by alternating tensors. We then proceed to study the algebra of alternating tensors. The most important algebraic construction is a product operation called the wedge product, which takes alternating tensors to alternating tensors. Then we transfer this to manifolds, and introduce the exterior derivative, which is a natural differential operator on differential forms.

At the end of the chapter, we introduce symplectic forms, which are a particular type of differential form that play an important role in geometry, analysis, and mathematical physics.

The Heuristics of Volume Measurement

In Chapter 4, we introduced line integrals of covector fields, which generalize ordinary integrals to curves in manifolds. As we will see in subsequent chapters, it is also useful to generalize the theory of *multiple* integrals to manifolds.

How might we make coordinate-independent sense of multiple integrals? First, observe that there is no way to define integrals of *functions* in a coordinate-independent way on a manifold. It is easy to see why, even in the simplest possible case: Suppose $C \subset \mathbb{R}^n$ is an n-dimensional cube, and $f\colon C \to \mathbb{R}$ is the constant function $f(x) \equiv 1$. Then

$$\int_C f\, dV = \operatorname{Vol}(C),$$

which is clearly not invariant under coordinate transformations, even if we just restrict attention to linear ones.

Let us think a bit more geometrically about why covector fields are the natural fields to integrate along curves. A covector field assigns a number to each tangent vector, in such a way that multiplying the tangent vector by a constant has the effect of multiplying the resulting number by the same constant. Thus a covector field can be thought of as assigning a "signed length meter" to each one-dimensional subspace of the tangent space, and it does so in a coordinate-independent way. Computing the line integral of a covector field, in effect, assigns a "length" to a curve by using this varying measuring scale along the points of the curve.

Now we wish to seek a kind of "field" that can be integrated in a coordinate-independent way over submanifolds of dimension $k > 1$. Its value at each point should be something that we can interpret as a "signed volume meter" on k-dimensional subspaces of the tangent space—a machine Ω that accepts any k tangent vectors $(X_1, \dots, X_k)$ at a point and returns a number $\Omega(X_1, \dots, X_k)$ that we might think of as the "signed volume" of the parallelepiped spanned by those vectors, measured according to a scale determined by Ω.

The most obvious example of such a machine is the determinant in $\mathbb{R}^n$. For example, it is shown in most linear algebra texts that for any two vectors $X_1, X_2 \in \mathbb{R}^2$, $\det(X_1, X_2)$ is, up to a sign, the area of the parallelogram spanned by X_1, X_2. It is not hard to show (see Problem 9-1) that the analogous fact is true in all dimensions. The determinant, remember, is an example of a tensor. In fact, it is a tensor of a very specific type: It changes sign whenever two of its arguments are interchanged. A covariant k-tensor T on a finite-dimensional vector space V is said to be *alternating* if it has this property:

$$T(X_1, \dots, X_i, \dots, X_j, \dots, X_k) = -T(X_1, \dots, X_j, \dots, X_i, \dots, X_k).$$

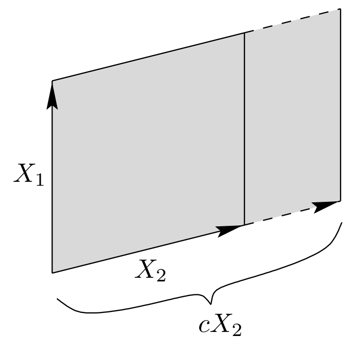

FIGURE 9.1. Scaling by a constant.

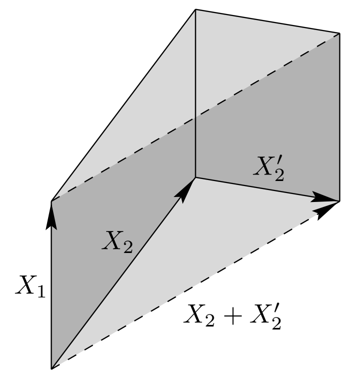

FIGURE 9.2. Sum of two vectors.

Let us consider what properties we might expect a general "signed volume meter" Ω to have. To be consistent with our ordinary ideas of volume, we would expect that multiplying any one of the vectors by a constant c should cause the volume to be scaled by that same constant (Figure 9.1), and that the parallelepiped formed by adding together two vectors in the ith place results in a volume that is the sum of the volumes of the two parallelepipeds with the original vectors in the ith place (Figure 9.2):

$$\begin{aligned}\Omega(X_1, \dots, cX_i, \dots, X_n) &= c\Omega(X_1, \dots, X_i, \dots, X_n),\\ \Omega(X_1, \dots, X_i + X_i', \dots, X_n) &= \Omega(X_1, \dots, X_i, \dots, X_n)\\ &\quad + \Omega(X_1, \dots, X_i', \dots, X_n).\end{aligned}$$

These two requirements suggest that Ω should be multilinear, and thus should be a covariant k-tensor.

There is one more essential property that we should expect: Since n linearly dependent vectors span a parallepiped of zero n-dimensional volume, Ω should give the value zero whenever it is applied to n linearly dependent vectors. As the next lemma shows, this forces Ω to be an alternating tensor.

Lemma 9.1. *Suppose Ω is a k-tensor on a vector space V with the property that $\Omega(X_1, \dots, X_k) = 0$ whenever $X_1, \dots, X_k$ are linearly dependent. Then Ω is alternating.*

Proof. The hypothesis implies, in particular, that Ω gives the value zero whenever two of its arguments are the same. This in turn implies

$$\begin{aligned}
0 &= \Omega(X_1, \ldots, X_i + X_j, \ldots, X_i + X_j, \ldots, X_n) \\
&= \Omega(X_1, \ldots, X_i, \ldots, X_i, \ldots, X_n) + \Omega(X_1, \ldots, X_i, \ldots, X_j, \ldots, X_n) \\
&\quad + \Omega(X_1, \ldots, X_j, \ldots, X_i, \ldots, X_n) + \Omega(X_1, \ldots, X_j, \ldots, X_j, \ldots, X_n) \\
&= \Omega(X_1, \ldots, X_i, \ldots, X_j, \ldots, X_n) + \Omega(X_1, \ldots, X_j, \ldots, X_i, \ldots, X_n).
\end{aligned}$$

Thus Ω is alternating. □

Because of these considerations, alternating tensor fields are promising candidates for objects that can be integrated in a coordinate-independent way. We will develop these ideas rigorously in the remainder of this chapter and the next; as we do, you should keep this geometric motivation in mind.

The Algebra of Alternating Tensors

In this section, we set aside heuristics and start developing the technical machinery for working with alternating tensors. For any finite-dimensional real vector space V, let $\Lambda^k(V)$ denote the subspace of $T^k(V)$ consisting of alternating tensors. [Warning: Some authors use the notation $\Lambda^k(V^*)$ in place of $\Lambda^k(V)$ for this space; see Problem 9-8 for a discussion of the reasons why.] An alternating k-tensor is sometimes called a *k-covector*.

Recall that for any permutation $\sigma \in S_k$, the *sign* of σ, denoted by $\operatorname{sgn}\sigma$, is equal to $+1$ if σ is even (i.e., can be written as a composition of an even number of transpositions), and -1 if σ is odd.

The following exercise is an analogue of Exercise 8.9.

Exercise 9.1. Show that the following are equivalent for a covariant k-tensor T:

(a) T is alternating.

(b) For any vectors $X_1, \ldots, X_k$ and any permutation $\sigma \in S_k$,

$$T(X_{\sigma(1)}, \ldots, X_{\sigma(k)}) = (\operatorname{sgn}\sigma) T(X_1, \ldots, X_k).$$

(c) T gives zero whenever two of its arguments are equal:

$$T(X_1, \ldots, Y, \ldots, Y, \ldots, X_k) = 0.$$

(d) $T(X_1, \ldots, X_k) = 0$ whenever the vectors $(X_1, \ldots, X_k)$ are linearly independent.

(e) With respect to any basis, the components $T_{i_1 \ldots i_k}$ of T change sign whenever two indices are interchanged.

Notice that part (d) implies that there are no nonzero alternating k-tensors on V if $k > \dim V$, for then every k-tuple of vectors is dependent.

Every 0-tensor (which is just a real number) is alternating, because there are no arguments to interchange. Similarly, every 1-tensor is alternating. An alternating 2-tensor is just a skew-symmetric bilinear form on V. It is interesting to note that any 2-tensor T can be expressed as the sum of an alternating tensor and a symmetric one, because

$$\begin{aligned} T(X,Y) &= \tfrac{1}{2}(T(X,Y) - T(Y,X)) + \tfrac{1}{2}(T(X,Y) + T(Y,X)) \\ &= A(X,Y) + S(X,Y), \end{aligned}$$

where $A(X,Y) = \frac{1}{2}(T(X,Y) - T(Y,X))$ is alternating, and $S(X,Y) = \frac{1}{2}(T(X,Y) + T(Y,X))$ is symmetric. This is not true for tensors of higher rank, as Problem 9-2 shows.

The tensor S defined above is just $\operatorname{Sym} T$, the symmetrization of T defined in the preceding chapter. We define a similar projection $\operatorname{Alt}\colon T^k(V) \to \Lambda^k(V)$, called the *alternating projection*, as follows:

$$\operatorname{Alt} T = \frac{1}{k!} \sum_{\sigma \in S_k} (\operatorname{sgn} \sigma) T^\sigma.$$

More explicitly, this means

$$(\operatorname{Alt} T)(X_1, \dots, X_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\operatorname{sgn} \sigma) T(X_{\sigma(1)}, \dots, X_{\sigma(k)}).$$

Example 9.2. If T is any 1-tensor, then $\operatorname{Alt} T = T$. If T is a 2-tensor, then

$$\operatorname{Alt} T(X,Y) = \tfrac{1}{2}(T(X,Y) - T(Y,X)).$$

For a 3-tensor T,

$$\begin{aligned} \operatorname{Alt} T(X,Y,Z) = \tfrac{1}{6}(&T(X,Y,Z) + T(Y,Z,X) + T(Z,X,Y) \\ &- T(Y,X,Z) - T(X,Z,Y) - T(Z,Y,X)). \end{aligned}$$

The next lemma is the analogue of Lemma 8.10.

Lemma 9.3 (Properties of the Alternating Projection).

(*a*) *For any tensor T,* $\operatorname{Alt} T$ *is alternating.*

(*b*) *T is alternating if and only if* $\operatorname{Alt} T = T$.

Exercise 9.2. Prove Lemma 9.3.

Elementary Alternating Tensors

Let k be a positive integer. An ordered k-tuple $I = (i_1, \dots, i_k)$ of positive integers is called a *multi-index* of length k. If I and J are multi-indices such that J is obtained from I by a permutation $\sigma \in S_k$, in the sense that

$$j_1 = i_{\sigma(1)}, \ \dots, \ j_k = i_{\sigma(k)},$$

then we write $J = \sigma I$. It is useful to extend the Kronecker delta notation in the following way. If I and J are multi-indices of length k, we define

$$\delta_I^J = \begin{cases} \operatorname{sgn}\sigma & \text{if neither } I \text{ nor } J \text{ has a repeated index} \\ & \qquad \text{and } J = \sigma I \text{ for some } \sigma \in S_k, \\ 0 & \text{if } I \text{ or } J \text{ has a repeated index} \\ & \qquad \text{or } J \text{ is not a permutation of } I. \end{cases}$$

Let V be an n-dimensional vector space, and suppose $(\varepsilon^1, \dots, \varepsilon^n)$ is any basis for V^*. We will define a collection of alternating tensors on V that generalize the determinant function on $\mathbb{R}^n$. For each multi-index $I = (i_1, \dots, i_k)$ of length k such that $1 \le i_1, \dots, i_k \le n$, define a covariant k-tensor ε^I by

$$\begin{aligned} \varepsilon^I(X_1, \dots, X_k) &= \det \begin{pmatrix} \varepsilon^{i_1}(X_1) & \dots & \varepsilon^{i_1}(X_k) \\ \vdots & & \vdots \\ \varepsilon^{i_k}(X_1) & \dots & \varepsilon^{i_k}(X_k) \end{pmatrix} \\ &= \det \begin{pmatrix} X_1^{i_1} & \dots & X_k^{i_1} \\ \vdots & & \vdots \\ X_1^{i_k} & \dots & X_k^{i_k} \end{pmatrix}. \end{aligned} \tag{9.1}$$

In other words, if $\mathbb{X}$ denotes the matrix whose columns are the components of the vectors $X_1, \dots, X_k$ with respect to the basis (E_i) dual to (ε^i), then $\varepsilon^I(X_1, \dots, X_k)$ is the determinant of the $k \times k$ minor consisting of rows $i_1, \dots, i_k$ of $\mathbb{X}$. Because the determinant changes sign whenever two columns are interchanged, it is clear that ε^I is an alternating k-tensor. We will call ε^I an *elementary alternating tensor* or *elementary k-covector*.

For example, in terms of the standard dual basis (e^1, e^2, e^3) for $(\mathbb{R}^3)^*$, we have

$$\begin{aligned} e^{13}(X, Y) &= X^1 Y^3 - Y^1 X^3; \\ e^{123}(X, Y, X) &= \det(X, Y, Z). \end{aligned}$$

Lemma 9.4. *Let (E_i) be a basis for V, let (ε^i) be the dual basis for V^*, and let ε^I be as defined above.*

(*a*) *If I has a repeated index, then $\varepsilon^I = 0$.*

(*b*) *If* $J = \sigma I$ *for some* $\sigma \in S_k$, *then* $\varepsilon^I = (\operatorname{sgn}\sigma)\varepsilon^J$.

(*c*) *The result of evaluating* ε^I *on a sequence of basis vectors is*

$$\varepsilon^I(E_{j_1}, \dots, E_{j_k}) = \delta^I_J.$$

Proof. If I has a repeated index, then for any vectors $X_1, \dots, X_k$, the determinant in (9.1) has two identical rows and thus is equal to zero, which proves (a). On the other hand, if J is obtained from I by interchanging two indices, then the corresponding determinants have opposite signs; this implies (b).

To prove (c), we consider several cases. First, if I has a repeated index, then $\varepsilon^I = 0$ by part (a). If J has a repeated index, then $\varepsilon^I(E_{j_1}, \dots, E_{j_k}) = 0$ by Exercise 9.1(c). If neither multi-index has any repeated indices but J is not a permutation of I, then the determinant in the definition of $\varepsilon^I(E_{j_1}, \dots, E_{j_k})$ has at least one row of zeros, so it is zero. If $J = I$, then $\varepsilon^I(E_{j_1}, \dots, E_{j_k})$ is the determinant of the identity matrix, which is 1. Finally, if $J = \sigma I$, then

$$\varepsilon^I(E_{j_1}, \dots, E_{j_k}) = (\operatorname{sgn}\sigma)\varepsilon^J(E_{j_1}, \dots, E_{j_k}) = \operatorname{sgn}\sigma$$

by part (b). □

The significance of the elementary k-covectors is that they provide a convenient basis for $\Lambda^k(V)$. Of course, the ε^I are not all independent, because some of them are zero and the ones corresponding to different permutations of the same multi-index are constant multiples of each other. But, as the next lemma shows, we can get a basis by restricting attention to an appropriate subset of multi-indices. A multi-index $I = (i_1, \dots, i_k)$ is said to be *increasing* if $i_1 < \cdots < i_k$. It will be useful to use a primed summation sign to denote a sum over only increasing multi-indices, so that, for example,

$$\sum_I{}' T_I \varepsilon^I = \sum_{\{I : 1 \le i_1 < \cdots < i_k \le n\}} T_I \varepsilon^I.$$

Lemma 9.5. *Let* V *be an* n*-dimensional vector space. If* (ε^i) *is any basis for* V^*, *then the collection of* k*-covectors*

$$\{\varepsilon^I : I \textit{ is increasing}\}$$

is a basis for $\Lambda^k(V)$. *Therefore,*

$$\dim \Lambda^k(V) = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proof. Let (E_i) be the basis for V dual to (ε^i), and let $\mathcal{E} = \{\varepsilon^I : I \text{ is increasing}\}$. We need to show that the set $\mathcal{E}$ spans $\Lambda^k(V)$ and is independent.

To show that $\mathcal{E}$ spans, let $T \in \Lambda^k(V)$ be arbitrary. For each multi-index $I = (i_1, \ldots, i_k)$, define a real number T_I by

$$T_I = T(E_{i_1}, \ldots, E_{i_k}).$$

The fact that T is alternating implies that $T_I = 0$ if I contains a repeated multi-index, and $T_J = (\operatorname{sgn}\sigma)T_I$ if $J = \sigma I$ for $\sigma \in S_k$. For any multi-index J, Lemma 9.4 gives

$$\sum_I{}' T_I \varepsilon^I(E_{j_1}, \ldots, E_{j_k}) = \sum_I{}' T_I \delta^I_J = T_J = T(E_{j_1}, \ldots, E_{j_k}).$$

Therefore, $\sum_I' T_I \varepsilon^I = T$, so $\mathcal{E}$ spans $\Lambda^k(V)$.

To show that $\mathcal{E}$ is an independent set, suppose

$$\sum_I{}' T_I \varepsilon^I = 0$$

for some coefficients T_I. Let J be any increasing multi-index. Applying both sides to $(E_{j_1}, \ldots, E_{j_k})$ and using Lemma 9.4,

$$0 = \sum_I{}' T_I \varepsilon^I(E_{j_1}, \ldots, E_{j_k}) = T_J.$$

Thus each coefficient T_J is zero. $\square$

In particular, for an n-dimensional vector space V, this lemma implies that $\Lambda^n(V)$ is 1-dimensional, and is spanned by $\varepsilon^{1\ldots n}$. By definition, this elementary n-covector acts on vectors $(X_1, \ldots, X_n)$ by taking the determinant of the component matrix $\mathbb{X} = (X^i_j)$. For example, on $\mathbb{R}^n$ with the standard basis, $e^{1\ldots n}$ is precisely the determinant function. Since there are no increasing multi-indices of length greater than n, the space $\Lambda^k(V)$ is trivial for $k > n$.

The Wedge Product

In Chapter 8, we defined the symmetric product, which takes a pair of symmetric tensors S, T and yields another symmetric tensor $ST = \operatorname{Sym}(S \otimes T)$ whose rank is the sum of the ranks of the original ones.

In this section, we will define a similar product operation for alternating tensors. One way to define it would be to mimic what we did in the symmetric case and define the product of alternating tensors ω and η to be $\operatorname{Alt}(\omega \otimes \eta)$. However, we will use a different definition that looks more complicated at first but turns out to be much better suited to computation.

If $\omega \in \Lambda^k(V)$ and $\eta \in \Lambda^l(V)$, we define the *wedge product* or *exterior product* of ω and η to be the alternating $(k+l)$-tensor

$$\omega \wedge \eta = \frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega \otimes \eta). \tag{9.2}$$

The mysterious coefficient is motivated by the simplicity of the statement of the following lemma.

Lemma 9.6. *For any multi-indices $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_l)$,*

$$\varepsilon^I \wedge \varepsilon^J = \varepsilon^{IJ}, \tag{9.3}$$

where IJ is the multi-index $(i_1, \ldots, i_k, j_1, \ldots, j_l)$ obtained by concatenating I and J.

Proof. By multilinearity, it suffices to show that

$$\varepsilon^I \wedge \varepsilon^J (E_{p_1}, \ldots, E_{p_{k+l}}) = \varepsilon^{IJ}(E_{p_1}, \ldots, E_{p_{k+l}}) \tag{9.4}$$

for any sequence $(E_{p_1}, \ldots, E_{p_{k+l}})$ of basis vectors. We consider several cases.

Case I: *$P = (p_1, \ldots, p_{k+l})$ has a repeated index.* In this case, both sides of (9.4) are zero by Exercise 9.1(c).

Case II: *P contains an index that does not appear in either I or J.* In this case, the right-hand side is zero by Lemma 9.4(c). Similarly, each term in the expansion of the left-hand side involves either ε^I or ε^J evaluated on a sequence of basis vectors that is not a permutation of I or J, respectively, so the left-hand side is also zero.

Case III: *$P = IJ$ and P has no repeated indices.* In this case, the right-hand side of (9.4) is equal to 1 by Lemma 9.4(c), so we need to show that the left-hand side is also equal to 1. By definition,

$$\begin{aligned}
&\varepsilon^I \wedge \varepsilon^J (E_{p_1}, \ldots, E_{p_{k+l}}) \\
&\quad = \frac{(k+l)!}{k!l!} \operatorname{Alt}(\varepsilon^I \otimes \varepsilon^J)(E_{p_1}, \ldots, E_{p_{k+l}}) \\
&\quad = \frac{1}{k!l!} \sum_{\sigma \in S_{k+l}} (\operatorname{sgn} \sigma) \varepsilon^I (E_{p_{\sigma(1)}}, \ldots, E_{p_{\sigma(k)}}) \varepsilon^J (E_{p_{\sigma(k+1)}}, \ldots, E_{p_{\sigma(k+l)}}).
\end{aligned}$$

By Lemma 9.4 again, the only terms in the sum above that give nonzero values are those in which σ permutes the first k indices and the last l indices of P separately. In other words, σ must be of the form $\sigma = \tau\eta$, where $\tau \in S_k$ acts by permuting $\{1, \ldots, k\}$ and $\eta \in S_l$ acts by permuting

$\{k+1,\ldots,k+l\}$. Since $\operatorname{sgn}(\tau\eta) = (\operatorname{sgn}\tau)(\operatorname{sgn}\eta)$, we have

$$\begin{aligned}
&\varepsilon^I \wedge \varepsilon^J(E_{p_1},\ldots,E_{p_{k+l}}) \\
&\quad = \frac{1}{k!l!}\sum_{\substack{\tau\in S_k\\ \eta\in S_l}} (\operatorname{sgn}\tau)(\operatorname{sgn}\eta)\varepsilon^I(E_{p_{\tau(1)}},\ldots,E_{p_{\tau(k)}})\varepsilon^J(E_{p_{\eta(k+1)}},\ldots,E_{p_{\eta(k+l)}}) \\
&\quad = \left(\frac{1}{k!}\sum_{\tau\in S_k}(\operatorname{sgn}\tau)\varepsilon^I(E_{p_{\tau(1)}},\ldots,E_{p_{\tau(k)}})\right)\times \\
&\qquad\qquad\qquad \left(\frac{1}{l!}\sum_{\eta\in S_l}(\operatorname{sgn}\eta)\varepsilon^J(E_{p_{\eta(k+1)}},\ldots,E_{p_{\eta(k+l)}})\right) \\
&\quad = (\operatorname{Alt}\varepsilon^I)(E_{p_1},\ldots,E_{p_k})\,(\operatorname{Alt}\varepsilon^J)(E_{p_{k+1}},\ldots,E_{p_{k+l}}) \\
&\quad = \varepsilon^I(E_{p_1},\ldots,E_{p_k})\varepsilon^J(E_{p_{k+1}},\ldots,E_{p_{k+l}}) \\
&\quad = 1.
\end{aligned}$$

Case IV: *P is a permutation of IJ.* In this case, applying a permutation to P brings us back to Case III. Since the effect of the permutation is to multiply both sides of (9.4) by the same sign, the result holds in this case as well. □

Proposition 9.7 (Properties of the Wedge Product).

(*a*) Bilinearity:

$$\begin{aligned}
(a\omega + a'\omega')\wedge\eta &= a(\omega\wedge\eta) + a'(\omega'\wedge\eta),\\
\eta\wedge(a\omega + a'\omega') &= a(\eta\wedge\omega) + a'(\eta\wedge\omega').
\end{aligned}$$

(*b*) Associativity:

$$\omega\wedge(\eta\wedge\xi) = (\omega\wedge\eta)\wedge\xi.$$

(*c*) Anticommutativity: *For $\omega\in\Lambda^k(V)$ and $\eta\in\Lambda^l(V)$,*

$$\omega\wedge\eta = (-1)^{kl}\eta\wedge\omega.$$

(*d*) *If $(\varepsilon^1,\ldots,\varepsilon^n)$ is any basis for V^* and $I=(i_1,\ldots,i_k)$ is any multi-index,*

$$\varepsilon^{i_1}\wedge\cdots\wedge\varepsilon^{i_k} = \varepsilon^I. \tag{9.5}$$

(*e*) *For any covectors $\omega^1,\ldots,\omega^k$ and vectors $X_1,\ldots,X_k$,*

$$\omega^1\wedge\cdots\wedge\omega^k(X_1,\ldots,X_k) = \det(\omega^i(X_j)). \tag{9.6}$$

Proof. Bilinearity follows immediately from the definition, because the tensor product is bilinear and Alt is linear. To prove associativity, note that Lemma 9.6 gives

$$(\varepsilon^I \wedge \varepsilon^J) \wedge \varepsilon^K = \varepsilon^{IJ} \wedge \varepsilon^K = \varepsilon^{IJK} = \varepsilon^I \wedge \varepsilon^{JK} = \varepsilon^I \wedge (\varepsilon^J \wedge \varepsilon^K).$$

The general case follows from bilinearity. Similarly, using Lemma 9.6 again, we get

$$\varepsilon^I \wedge \varepsilon^J = \varepsilon^{IJ} = (\operatorname{sgn}\tau)\varepsilon^{JI} = (\operatorname{sgn}\tau)\varepsilon^J \wedge \varepsilon^I,$$

where τ is the permutation that sends IJ to JI. It is easy to check that $\operatorname{sgn}\tau = (-1)^{kl}$, because τ can be decomposed as a composition of kl transpositions (each index of I must be moved past each of the indices of J). Anticommutativity then follows from bilinearity.

Part (d) is an immediate consequence of Lemma 9.6 and induction. To prove part (e), we note that the special case in which each ω^j is one of the basis covectors ε^{i_j} just reduces to (9.5). Since both sides of (9.6) are multilinear in $(\omega^1, \dots, \omega^k)$, this suffices. □

Because of part (d) of this lemma, we will generally use the notations ε^I and $\varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_k}$ interchangeably.

The definition and computational properties of the wedge product can seem daunting at first sight. However, the only properties that you need to remember for most practical purposes are that it is bilinear, associative, and anticommutative, and satisfies (9.6). In fact, these properties determine the wedge product uniquely, as the following exercise shows.

Exercise 9.3. Show that the wedge product is the unique associative, bilinear, and anticommutative map $\Lambda^k(V) \times \Lambda^l(V) \to \Lambda^{k+l}(V)$ satisfying (9.6).

As we observed at the beginning of this section, one could also define the wedge product without the unwieldy coefficient of (9.2). Many authors choose this alternative definition of the wedge product, which we denote by $\overline{\wedge}$:

$$\omega \,\overline{\wedge}\, \eta = \operatorname{Alt}(\omega \otimes \eta). \tag{9.7}$$

Using this definition, (9.3) is replaced by

$$\varepsilon^I \,\overline{\wedge}\, \varepsilon^J = \frac{k!l!}{(k+l)!}\varepsilon^{IJ}$$

and (9.6) is replaced by

$$\omega^1 \,\overline{\wedge} \cdots \overline{\wedge}\, \omega^k(X_1, \dots, X_k) = \frac{1}{k!}\det(\omega^i(X_j)) \tag{9.8}$$

whenever $\omega^1, \dots, \omega^k$ are covectors, as you can check.

Because of (9.6), we will call the wedge product defined by (9.2) the *determinant convention* for the wedge product, and the wedge product defined by (9.7) the *Alt convention*. Although the definition of the Alt convention is perhaps a bit more natural, the computational advantages of the determinant convention make it preferable for most applications, and we will use it exclusively in this book.

Differential Forms on Manifolds

Now we turn our attention to an n-dimensional smooth manifold M. The subset of T^kM consisting of alternating tensors is denoted by $\Lambda^k M$:

$$\Lambda^k M = \coprod_{p \in M} \Lambda^k(T_pM).$$

Exercise 9.4. Show that $\Lambda^k M$ is a smooth subbundle of T^kM, and is therefore a smooth vector bundle of rank $\binom{n}{k}$ over M.

A smooth section of $\Lambda^k M$ is called a *differential k-form*, or just a *k-form*; this is just a smooth tensor field whose value at each point is an alternating tensor. We denote the vector space of sections of $\Lambda^k M$ by $\mathcal{A}^k(M)$. (We ordinarily denote the space of sections of a vector bundle by the upper-case script letter corresponding to the name of the bundle; in this case, we use $\mathcal{A}$ because of the typographical similarity between A and Λ.)

In any coordinate chart, a k-form ω can be written locally as

$$\omega = \sum_I{}' \omega_I dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \sum_I{}' \omega_I dx^I,$$

where the coefficients ω_I are smooth functions defined on the coordinate neighborhood, and we use dx^I as an abbreviation for $dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ (not to be mistaken for the differential of a function x^I). In terms of differential forms, the result of Lemma 9.4(c) translates to

$$dx^{i_1} \wedge \cdots \wedge dx^{i_k} \left(\frac{\partial}{\partial x^{j_1}}, \ldots, \frac{\partial}{\partial x^{j_k}} \right) = \delta^I_J.$$

The wedge product of a k-form with an l-form is a $(k+l)$-form. A 0-form is just a real-valued function, and we interpret the wedge product $f \wedge \eta$ of a 0-form f with a k-form η to mean the product $f\eta$.

Example 9.8. A 1-form is just a smooth covector field. On $\mathbb{R}^3$, some examples of 2-forms are given by

$$\begin{aligned} \omega &= (\sin xy)\, dy \wedge dz; \\ \eta &= dx \wedge dy + dx \wedge dz + dy \wedge dz. \end{aligned}$$

Every n-form on $\mathbb{R}^n$ is a smooth function times $dx^1 \wedge \cdots \wedge dx^n$, because there is only one increasing multi-index of length n.

If $F\colon M \to N$ is a smooth map and ω is a differential form on N, the pullback $F^*\omega$ is a differential form on M, defined as for any smooth tensor field:

$$F^*\omega(X_1, \ldots, X_k) = \omega(F_*X_1, \ldots, F_*X_k).$$

In particular, if $\iota\colon N \hookrightarrow M$ is the inclusion map of an immersed submanifold, then we usually use the notation $\omega|_N$ for $\iota^*\omega$.

Lemma 9.9. *Suppose $F\colon M \to N$ is smooth.*

(*a*) $F^*(\omega \wedge \eta) = (F^*\omega) \wedge (F^*\eta)$.

(*b*) *In any coordinate chart,*

$$F^*\left(\sum_I{}' \omega_I dy^{i_1} \wedge \cdots \wedge dy^{i_k}\right) = \sum_I{}' (\omega_I \circ F)\, d(y^{i_1} \circ F) \wedge \cdots \wedge d(y^{i_k} \circ F).$$

Exercise 9.5. Prove this lemma.

This lemma gives a computational rule for pullbacks of differential forms similar to the one we developed for arbitrary tensor fields in the preceding chapter. As before, it can also be used to compute the expression for a differential form in another coordinate chart.

Example 9.10. Let ω be the 2-form $dx \wedge dy$ on $\mathbb{R}^2$. Thinking of the transformation to polar coordinates $x = r\cos\theta$, $y = r\sin\theta$ as an expression for the identity map with respect to different coordinates on the domain and range, we find

$$\begin{aligned}
\omega &= dx \wedge dy \\
&= d(r\cos\theta) \wedge d(r\sin\theta) \\
&= (\cos\theta\, dr - r\sin\theta\, d\theta) \wedge (\sin\theta\, dr + r\cos\theta\, d\theta) \\
&= r\cos^2\theta\, dr \wedge d\theta - r\sin^2\theta\, d\theta \wedge dr,
\end{aligned}$$

where we have used the fact that $dr \wedge dr = d\theta \wedge d\theta = 0$ by anticommutativity. Because $d\theta \wedge dr = -dr \wedge d\theta$, this simplifies to

$$dx \wedge dy = r\, dr \wedge d\theta.$$

The similarity between this formula and the formula for changing a double integral from Cartesian to polar coordinates is striking. More generally, we have the following lemma:

Lemma 9.11. *Let $F\colon M \to N$ be a smooth map between n-manifolds. If (x^i) and (y^j) are coordinates on open sets $U \subset M$ and $V \subset N$, respectively, and u is a smooth function on V, then the following holds on $U \cap F^{-1}(V)$:*

$$F^*\left(u\,dy^1 \wedge \cdots \wedge dy^n\right) = (u \circ F)\det\left(\frac{\partial F^j}{\partial x^i}\right) dx^1 \wedge \cdots \wedge dx^n. \tag{9.9}$$

Proof. Because the fiber of $\Lambda^n M$ is spanned by $dx^1 \wedge \cdots \wedge dx^n$ at each point, it suffices to show that both sides of (9.9) give the same result when evaluated on $(\partial/\partial x^1, \ldots, \partial/\partial x^n)$. From Lemma 9.9,

$$F^*\left(u\,dy^1 \wedge \cdots \wedge dy^n\right) = (u \circ F)dF^1 \wedge \cdots \wedge dF^n.$$

Proposition 9.7(e) shows that

$$dF^1 \wedge \cdots \wedge dF^n\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) = \det\left(dF^j\left(\frac{\partial}{\partial x^i}\right)\right) = \det\left(\frac{\partial F^j}{\partial x^i}\right).$$

Therefore, the left-hand side of (9.9) gives $(u \circ F)\det(\partial F^j/\partial x^i)$ when applied to $(\partial/\partial x^1, \ldots, \partial/\partial x^n)$. On the other hand, the right-hand side gives the same thing, because $dx^1 \wedge \cdots \wedge dx^n(\partial/\partial x^1, \ldots, \partial/\partial x^n) = 1$. □

Exterior Derivatives

In this section, we define a natural differential operator on forms, called the exterior derivative. It is a generalization of the differential of a function.

To give some idea where the motivation for the exterior derivative comes from, let us look back at a question we addressed in Chapter 4. Recall that not all smooth covector fields are differentials of functions: Given ω, a necessary condition for the existence of a function f such that $\omega = df$ is that ω be closed, which means that it satisfies

$$\frac{\partial \omega_j}{\partial x^i} - \frac{\partial \omega_i}{\partial x^j} = 0 \tag{9.10}$$

in every coordinate system. Since this is a coordinate-independent property by Lemma 4.25, one might hope to find a more invariant way to express it. The key is that the expression in (9.10) is antisymmetric in the indices i and j, so it can be interpreted as the ij-component of an alternating tensor field, i.e., a 2-form. We will define a 2-form $d\omega$ by

$$d\omega = \sum_{i<j}\left(\frac{\partial \omega_j}{\partial x^i} - \frac{\partial \omega_i}{\partial x^j}\right) dx^i \wedge dx^j,$$

so it follows that ω is closed if and only if $d\omega = 0$.

This formula has a significant generalization to differential forms of all degrees. For any manifold, we will show that there is a differential operator $d\colon \mathcal{A}^k(M) \to \mathcal{A}^{k+1}(M)$ satisfying $d(d\omega) = 0$ for all ω. Thus it will follow that a necessary condition for a k-form ω to be equal to $d\eta$ for some $(k-1)$-form η is that $d\omega = 0$.

The definition of d in coordinates is straightforward:

$$d\Bigl({\sum_I}' \omega_I dx^I\Bigr) = {\sum_I}' d\omega_I \wedge dx^I, \tag{9.11}$$

where $d\omega_I$ is just the differential of the function ω_I. In somewhat more detail, this is

$$d\Bigl({\sum_I}' \omega_I dx^{i_1} \wedge \cdots \wedge dx^{i_k}\Bigr) = {\sum_I}' \sum_i \frac{\partial \omega_I}{\partial x^i} dx^i \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

Observe that when ω is a 1-form, this becomes

$$d(\omega_j\, dx^j) = \frac{\partial \omega_j}{\partial x^i} dx^i \wedge dx^j = \sum_{i<j} \left(\frac{\partial \omega_j}{\partial x^i} - \frac{\partial \omega_i}{\partial x^j}\right) dx^i \wedge dx^j,$$

using the fact that $dx^i \wedge dx^j = -dx^j \wedge dx^i$, so this is consistent with our earlier definition.

Proving that this definition is independent of the choice of coordinates takes a little work. This is the content of the next theorem.

Theorem 9.12 (The Exterior Derivative). *On any smooth manifold M, there is a unique linear map $d\colon \mathcal{A}^k(M) \to \mathcal{A}^{k+1}(M)$ for each $k \geq 0$ satisfying the following conditions:*

(*i*) *If f is a smooth function (a 0-form), then df is the differential of f, defined as usual by*

$$df(X) = Xf.$$

(*ii*) *If $\omega \in \mathcal{A}^k(M)$ and $\eta \in \mathcal{A}^l(M)$, then*

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

(*iii*) $d^2 = 0$. *More precisely, for any k-form ω, $d(d\omega) = 0$.*

In any local coordinates, d is given by (9.11).

Proof. First we will prove uniqueness. Suppose $d\colon \mathcal{A}^k(M) \to \mathcal{A}^{k+1}(M)$ is a linear operator satisfying (i), (ii), and (iii). We will show that it also satisfies (9.11) in any local coordinate chart, which implies that it is uniquely determined.

We begin by showing that d is *local*, in the following sense: If ω and $\widetilde{\omega}$ are k-forms on M that agree on an open subset $U \subset M$, then $d\omega = d\widetilde{\omega}$ on U. Writing $\eta = \widetilde{\omega} - \omega$, it clearly suffices to show that $d\eta = 0$ on U if η vanishes on U. Let $p \in U$ be arbitrary, and let $\varphi \in C^\infty(M)$ be a bump function that is equal to 1 in a neighborhood of p and supported in U. Then $\varphi\eta$ is identically zero on M, so

$$0 = d(\varphi\eta)_p = d\varphi_p \wedge \eta_p + \varphi(p) d\eta_p = d\eta_p,$$

because $\varphi \equiv 1$ in a neighborhood of p. Since p was an arbitrary point of U, this shows that $d\eta = 0$ on U.

Suppose (x^i) are coordinates on an open subset $U \subset M$. Let $\omega \in \mathcal{A}^k(M)$, and write $\omega = \sum_I' \omega_I dx^I$ in coordinates on U. We will show that (9.11) holds at each point $p \in U$. By the extension lemma (Lemma 2.20), we can extend the coordinate functions x^i to smooth functions $\widetilde{x}^i$ on all of M that agree with x^i in some neighborhood of p. Likewise, we extend the component functions ω_I to functions $\widetilde{\omega}_I$ on M. The k-form $\widetilde{\omega} = \sum_I' \widetilde{\omega}_I d\widetilde{x}^{i_1} \wedge \cdots \wedge d\widetilde{x}^{i_k}$ is globally defined on M and agrees with ω near p. Using linearity of d together with (i) and (ii), we compute

$$\begin{aligned}
d\widetilde{\omega} &= d\left(\sum_I{}' \widetilde{\omega}_I d\widetilde{x}^{i_1} \wedge \cdots \wedge d\widetilde{x}^{i_k}\right) \\
&= \sum_I{}' d\widetilde{\omega}_I \wedge d\widetilde{x}^{i_1} \wedge \cdots \wedge d\widetilde{x}^{i_k} + (-1)^0 \sum_I{}' \widetilde{\omega}_I d(d\widetilde{x}^{i_1} \wedge \cdots \wedge d\widetilde{x}^{i_k}),
\end{aligned}$$

because $\widetilde{\omega}_I$ is a 0-form. Now using (ii) again, the last term expands into a sum of terms, each of which contains a factor of the form $d(d\widetilde{x}^{i_p})$, which is zero by (iii). Therefore, since d is local,

$$d\omega_p = d\widetilde{\omega}_p = \sum_I{}' d\widetilde{\omega}_I|_p \wedge d\widetilde{x}^{i_1}_p \wedge \cdots \wedge d\widetilde{x}^{i_k}_p = \sum_I{}' d\omega_I|_p \wedge dx^{i_1}_p \wedge \cdots \wedge dx^{i_k}_p.$$

Since p was arbitrary, (9.11) holds on all of U. This proves that $d\omega$ is uniquely determined.

Now to prove that such an operator exists, we begin by assuming that M is covered by a single coordinate chart, and define d by (9.11). It is clearly linear and satisfies (i). We need to check that it satisfies (ii) and (iii). Before doing so, we need to know that d satisfies

$$d(f\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}) = df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

for *any* multi-index I, not just increasing ones. If I has repeated indices, then clearly both sides are zero. If not, let σ be the permutation sending I to an increasing multi-index J. Then

$$\begin{aligned}
d(f\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}) &= (\operatorname{sgn}\sigma) d(f\, dx^{j_1} \wedge \cdots \wedge dx^{j_k}) \\
&= (\operatorname{sgn}\sigma) df \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_k} \\
&= df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.
\end{aligned}$$

To prove (ii), it suffices to consider terms of the form $\omega = f\,dx^I = f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ and $\eta = g\,dx^J$. We compute

$$\begin{aligned} d(\omega \wedge \eta) &= d((f\,dx^I) \wedge (g\,dx^J)) \\ &= d(fg\,dx^I \wedge dx^J) \\ &= (g\,df + f\,dg) \wedge dx^I \wedge dx^J \\ &= (df \wedge dx^I) \wedge (g\,dx^J) + (-1)^k (f\,dx^I) \wedge (dg \wedge dx^J) \\ &= d\omega \wedge \eta + (-1)^k \omega \wedge d\eta, \end{aligned}$$

where the $(-1)^k$ comes from $dg \wedge dx^I = (-1)^k dx^I \wedge dg$.

We will prove (iii) first for the special case of a 0-form, i.e., a smooth real-valued function. In this case,

$$\begin{aligned} d(df) &= d\left(\frac{\partial f}{\partial x^j} dx^j\right) \\ &= \frac{\partial^2 f}{\partial x^i \partial x^j} dx^i \wedge dx^j \\ &= \sum_{i<j} \left(\frac{\partial^2 f}{\partial x^i \partial x^j} - \frac{\partial^2 f}{\partial x^j \partial x^i}\right) dx^i \wedge dx^j \\ &= 0. \end{aligned}$$

For the general case, we use the $k = 0$ case together with (ii) to compute

$$\begin{aligned} d(d\omega) &= d\left(\sum_I{}' d\omega_I \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) \\ &= \sum_I{}' d(d\omega_I) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &\qquad + \sum_I{}' \sum_{j=1}^{k} (-1)^j d\omega_I \wedge dx^{i_1} \wedge \cdots \wedge d(dx^{i_j}) \wedge \cdots \wedge dx^{i_k} \\ &= 0. \end{aligned}$$

Finally, we consider the case of an arbitrary manifold M. On any coordinate domain $U \subset M$, we have a unique linear operator d_U defined as above and satisfying (i)–(iii). On any set $U \cap U'$ where two charts overlap, the restrictions of $d_U\omega$ and $d_{U'}\omega$ to $U \cap U'$ must agree by uniqueness; therefore, defining $d\omega$ by (9.11) in each coordinate chart, we get a globally-defined operator satisfying (i)–(iii). □

The operator d whose existence and uniqueness are asserted in this theorem is called *exterior differentiation*, and $d\omega$ is called the *exterior derivative* of ω. The exterior derivative of a real-valued function f is, of course, just its differential df.

Example 9.13. Let us work out the exterior derivatives of arbitrary 1-forms and 2-forms on $\mathbb{R}^3$. Any 1-form can be written

$$\omega = P\,dx + Q\,dy + R\,dz$$

for some smooth functions P, Q, R. Using (9.11) and the fact that the wedge product of any 1-form with itself is zero, we compute

$$\begin{aligned}
d\omega &= dP \wedge dx + dQ \wedge dy + dR \wedge dz \\
&= \left(\frac{\partial P}{\partial x}dx + \frac{\partial P}{\partial y}dy + \frac{\partial P}{\partial z}dz\right) \wedge dx + \left(\frac{\partial Q}{\partial x}dx + \frac{\partial Q}{\partial y}dy + \frac{\partial Q}{\partial z}dz\right) \wedge dy \\
&\quad + \left(\frac{\partial R}{\partial x}dx + \frac{\partial R}{\partial y}dy + \frac{\partial R}{\partial z}dz\right) \wedge dz \\
&= \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz \wedge dx \\
&\quad + \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz.
\end{aligned}$$

It is interesting to note that the components of this 2-form are exactly the components of the curl of the vector field with components (P, Q, R) (except perhaps in a different order and with different signs). We will explore this connection in more depth later in the book.

An arbitrary 2-form on $\mathbb{R}^3$ can be written

$$\omega = \alpha\, dx \wedge dy + \beta\, dx \wedge dz + \gamma\, dy \wedge dz.$$

A similar computation shows

$$d\omega = \left(\frac{\partial \alpha}{\partial z} - \frac{\partial \beta}{\partial y} + \frac{\partial \gamma}{\partial x}\right) dx \wedge dy \wedge dz.$$

One important feature of the exterior derivative is that it behaves well with respect to pullbacks, as the next lemma shows.

Lemma 9.14. *If $G\colon M \to N$ is a smooth map, then the pullback map $G^*\colon \mathcal{A}^k(N) \to \mathcal{A}^k(M)$ commutes with d: For all $\omega \in \mathcal{A}^k(N)$,*

$$G^*(d\omega) = d(G^*\omega) \tag{9.12}$$

Proof. Let $\omega \in \mathcal{A}^k(N)$ be arbitrary. Because d is local, if (9.12) holds in a neighborhood of each point, then it holds on all of M. In a coordinate neighborhood, ω can be written as a sum of terms like $f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, so by linearity it suffices to check (9.12) for a form of this type.

For such a form, the left-hand side of (9.12) is

$$\begin{aligned}
G^* d(f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}) &= G^*(df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}) \\
&= d(f \circ G) \wedge d(x^{i_1} \circ G) \wedge \cdots \wedge d(x^{i_k} \circ G),
\end{aligned}$$

while the right-hand side is

$$\begin{aligned} dG^*(f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}) &= d((f \circ G)\,d(x^{i_1} \circ G) \wedge \cdots \wedge d(x^{i_k} \circ G)) \\ &= d(f \circ G) \wedge d(x^{i_1} \circ G) \wedge \cdots \wedge d(x^{i_k} \circ G). \end{aligned}$$

□

Extending the terminology that we introduced for covector fields in Chapter 4, we say that a differential form $\omega \in \mathcal{A}^k(M)$ is *closed* if $d\omega = 0$, and *exact* if there exists a $(k-1)$-form η on M such that $\omega = d\eta$. The fact that $d^2 = 0$ implies that every exact form is closed. The converse may not be true, as we saw already in Chapter 4 the case of covector fields. We will return to these ideas in Chapter 11.

Symplectic Forms

In this section, we introduce symplectic forms, a special kind of 2-form that plays a leading role in many applications of smooth manifold theory to analysis and physics.

We begin with some linear algebra. Recall that a 2-tensor ω on a finite-dimensional real vector space V is said to be *nondegenerate* if $\omega(X, Y) = 0$ for all $Y \in V$ implies $X = 0$.

Exercise 9.6. Show that the following are equivalent for a 2-tensor ω on a finite-dimensional vector space V:

(a) ω is nondegenerate.

(b) The matrix (ω_{ij}) representing ω in terms of any basis is nonsingular.

(c) The linear map $\widetilde{\omega}\colon V \to V^*$ defined by $\widetilde{\omega}(X)(Y) = \omega(X, Y)$ is invertible.

A nondegenerate alternating 2-tensor is called a *symplectic tensor*. A vector space V endowed with a specific symplectic tensor is called a *symplectic vector space*.

Example 9.15. Let V be a vector space of dimension $2n$. Choose any basis for V, and denote the basis by $(A_1, B_1, \ldots, A_n, B_n)$ and the corresponding dual basis for V^* by $(\alpha^1, \beta^1, \ldots, \alpha^n, \beta^n)$. Let $\omega \in \Lambda^2(V)$ be the 2-covector defined by

$$\omega = \sum_{i=1}^{n} \alpha^i \wedge \beta^i. \tag{9.13}$$

Note that the action of ω on basis vectors is given by

$$\begin{aligned} \omega(A_i, B_j) &= -\omega(B_j, A_i) = \delta_{ij}, \\ \omega(A_i, A_j) &= \omega(B_i, B_j) = 0. \end{aligned} \tag{9.14}$$

Suppose $X = a^i A_i + b^i B_i \in V$ satisfies $\omega(X,Y) = 0$ for all $Y \in V$. Then $0 = \omega(X, B_i) = a^i$ and $0 = \omega(X, A_i) = -b^i$, which implies that $X = 0$. Thus ω is nondegenerate.

It is useful to consider the special case in which $\dim V = 2$. In this case, every 2-covector is a multiple of $\alpha^1 \wedge \beta^1$, which is nondegenerate by the argument above. Thus every nonzero 2-covector on a 2-dimensional vector space is symplectic.

If (V, ω) is a symplectic vector space and $S \subset V$ is any subspace, we define the *symplectic complement* of S, denoted by $S^\perp$, to be the subspace

$$S^\perp = \{X \in V : \omega(X,Y) = 0 \text{ for all } Y \in S\}.$$

As the notation suggests, the symplectic complement is analogous to the orthogonal complement in an inner product space. For example, just as in the inner product case, the dimension of $S^\perp$ is the codimension of S, as the next lemma shows.

Lemma 9.16. *Let (V,ω) be a symplectic vector space. For any subspace $S \subset V$, $\dim S + \dim S^\perp = \dim V$.*

Proof. If $(E_1, \dots, E_k)$ is any basis for S, then the covectors $\widetilde{\omega}(E_1), \dots, \widetilde{\omega}(E_k) \subset V^*$ are independent because $\widetilde{\omega}$ is injective. Therefore, the map $\Phi \colon V \to \mathbb{R}^k$ given by

$$\Phi(X) = (\widetilde{\omega}(E_1)(X), \dots, \widetilde{\omega}(E_k)(X))$$

is surjective, so $S^\perp = \operatorname{Ker}\Phi$ has dimension equal to $\dim V - \dim \mathbb{R}^k$ by the rank-nullity law. □

Symplectic complements differ from orthogonal complements in one important respect: Although it is always true that $S \cap S^\perp = \{0\}$ in an inner product space, this need not be true in a symplectic vector space. Indeed, if S is 1-dimensional, the fact that ω is alternating forces $\omega(X,X) = 0$ for every $X \in S$, so $S \subset S^\perp$. Carrying this idea a little further, subspaces of V can be classified in the following way. A subspace $S \subset V$ is said to be

- *symplectic* if $S \cap S^\perp = \{0\}$;
- *isotropic* if $S \subset S^\perp$;
- *coisotropic* if $S \supset S^\perp$;
- *Lagrangian* if $S = S^\perp$.

Exercise 9.7. Let (V,ω) be a symplectic vector space, and let $S \subset V$ be a subspace.

(a) Show that $(S^\perp)^\perp = S$.

(b) Show that S is symplectic if and only if $\omega|_S$ is nondegenerate.

(c) Show that S is isotropic if and only if $\omega|_S = 0$.

(d) Show that S is Lagrangian if and only if $\omega|_S = 0$ and $\dim S = n$.

The symplectic tensor ω defined in Example 9.15 turns out to be the prototype of all symplectic tensors, as the next proposition shows. This can be viewed as a symplectic version of the Gram-Schmidt algorithm.

Proposition 9.17 (Canonical Form for a Symplectic Tensor). *Let ω be a symplectic tensor on an m-dimensional vector space V. Then V has even dimension $m = 2n$, and there exists a basis for V in which ω has the form* (9.13)*.*

Proof. It is easy to check that ω has the form (9.13) with respect to a basis $(A_1, B_1, \dots, A_n, B_n)$ if and only if the action of ω on basis vectors is given by (9.14). Thus we will prove the theorem by induction on $m = \dim V$, by showing that there exists a basis of this form.

For $m = 0$ there is nothing to prove. Suppose (V, ω) is a symplectic vector space of dimension $m \geq 1$, and assume the proposition is true for all symplectic vector spaces of dimension less than m. Let A_1 be any nonzero vector in V. Since ω is nondegenerate, there exists $B_1 \in V$ such that $\omega(A_1, B_1) \neq 0$. Multiplying B_1 by a constant if necessary, we may assume that $\omega(A_1, B_1) = 1$. Because ω is alternating, B_1 cannot be a multiple of A_1, so the set $\{A_1, B_1\}$ is independent.

Let $S \subset V$ be the subspace spanned by $\{A_1, B_1\}$. Then $\dim S^\perp = m - 2$ by Lemma 9.16. Since $\omega|_S$ is obviously nondegenerate, by Exercise 9.7 it follows that S is symplectic. This means $S \cap S^\perp = \{0\}$, so $S^\perp$ is also symplectic. By induction $S^\perp$ is even-dimensional and there is a basis $(A_2, B_2, \dots, A_n, B_n)$ for $S^\perp$ such that (9.14) is satisfied for $2 \leq i, j \leq n$. It follows easily that $(A_1, B_1, A_2, B_2, \dots, A_n, B_n)$ is the required basis for V. $\square$

Because of this proposition, if (V, ω) is a symplectic vector space, a basis $(A_1, B_1, \dots, A_n, B_n)$ for V is called a *symplectic basis* if (9.14) holds, which is equivalent to ω being given by (9.13) in terms of the dual basis. The proposition then says that every symplectic vector space has a symplectic basis.

Now let us turn to manifolds. A *symplectic form* on a smooth manifold M is a closed, nondegenerate 2-form. In other words, a 2-form ω is symplectic if and only if it is closed and ω_p is a symplectic tensor for each $p \in M$. A smooth manifold endowed with a specific choice of symplectic form is called a *symplectic manifold*. A choice of symplectic form is also sometimes called a *symplectic structure* on M. Proposition 9.17 implies that a symplectic manifold must be even-dimensional. If (M, ω) and $(\widetilde{M}, \widetilde{\omega})$ are symplectic manifolds, a diffeomorphism $F\colon M \to \widetilde{M}$ satisfying $F^*\widetilde{\omega} = \omega$ is called a *symplectomorphism*. The study of properties of symplectic manifolds that are invariant under symplectomorphisms is known as *symplectic geometry*.

Example 9.18 (Symplectic Manifolds).

(a) If we denote the standard coordinates on $\mathbb{R}^{2n}$ by $(x^1, y^1, \ldots, x^n, y^n)$, the 2-form

$$\omega = \sum_{i=1}^{n} dx^i \wedge dy^i$$

is symplectic: It is obviously closed, and its value at each point is the standard symplectic tensor of Example 9.15. This is called the *standard symplectic structure* (or *standard symplectic form*) on $\mathbb{R}^{2n}$.

(b) Suppose Σ is any smooth 2-manifold and Ω is any nonvanishing 2-form on Σ. Then Ω is closed because $d\Omega$ is a 3-form, and every 3-form on a 2-manifold is zero. Moreover, as we observed above, every nonvanishing 2-form is nondegenerate, so (Σ, Ω) is a symplectic manifold.

Suppose (M, ω) is a symplectic manifold. An (immersed or embedded) submanifold $N \subset M$ is said to be symplectic, isotropic, coisotropic, or Lagrangian if T_pN (thought of as a subspace of T_pM) has this property at each point $p \in N$. More generally, an immersion $F\colon N \to M$ is said to have one of these properties if the subspace $F_*(T_pN) \subset T_{F(p)}M$ has the corresponding property for every $p \in N$. Thus a submanifold is symplectic (isotropic, etc.) if and only if its inclusion map has the same property.

Exercise 9.8. Suppose (M, ω) is a symplectic manifold, and $F\colon N \to M$ is an immersion. Show that F is isotropic if and only if $F^*\omega = 0$, and F is symplectic if and only if $F^*\omega$ is a symplectic form.

The most important example of a symplectic manifold is the total space of the cotangent bundle of any smooth manifold M, which carries a canonical symplectic structure that we now define. First, there is a natural 1-form τ on the total space of T^*M, called the *tautologous* 1*-form*, defined as follows. A point in T^*M is a covector $\eta \in T^*_pM$ for some $p \in M$; we will denote such a point by the notation (p, η). The natural projection $\pi\colon T^*M \to M$ is then just $\pi(p, \eta) = p$, and its pullback is a linear map $\pi^*\colon T^*_pM \to T_{(p,\eta)}(T^*M)$. We define $\tau \in \Lambda^1(T^*M)$ by

$$\tau_{(p,\eta)} = \pi^*\eta.$$

In other words, the value of τ at $(p, \eta) \in T^*M$ is the pullback with respect to π of the covector η itself. If X is a tangent vector in $T_{(p,\eta)}(T^*M)$, then

$$\tau_{(p,\eta)}(X) = \eta(\pi_* X).$$

Proposition 9.19. *Let M be a smooth manifold. The tautologous* 1*-form τ is smooth, and $\omega = -d\tau$ is a symplectic form on the total space of T^*M.*

Proof. Let (x^i) be any coordinates on M, and let (x^i, ξ_i) denote the corresponding standard coordinates on T^*M as defined in Proposition 4.4. Recall that the coordinates of $(p, \eta) \in T^*M$ are defined to be (x^i, ξ_i), where (x^i) is the coordinate representation of p and $\xi_i \, dx^i$ is the coordinate representation of η. In terms of these coordinates, the projection $\pi \colon T^*M \to M$ has the coordinate expression $\pi(x, \xi) = x$, and therefore the coordinate representation of τ is

$$\tau_{(x,\xi)} = \pi^*(\xi_i \, dx^i) = \xi_i \, dx^i.$$

It follows immediately that τ is smooth, because its component functions are linear.

Clearly ω is closed, because it is exact. Moreover,

$$\omega = -d\tau = \sum_i dx^i \wedge d\xi_i.$$

Under the identification of an open subset of T^*M with an open subset of $\mathbb{R}^{2n}$ by means of these coordinates, ω corresponds to the standard symplectic form on $\mathbb{R}^{2n}$ (with ξ_i substituted for y^i). It follows that ω is symplectic. □

The symplectic structure defined in this proposition is called the *canonical symplectic structure* on T^*M. One of its many uses is in giving a somewhat more "geometric" picture of what it means for a 1-form to be closed, as shown by the following proposition.

Proposition 9.20. *Let M be a smooth manifold, and let σ be a* 1*-form on M. Thought of as a smooth map from M to T^*M, σ is an embedding, and σ is closed if and only if its image $\sigma(M)$ is a Lagrangian submanifold of T^*M.*

Proof. Throughout this proof, we need to remember that σ is playing two roles: On the one hand, it is a 1-form on M, and on the other hand, it is a smooth map between manifolds. Since they are literally the same map, we will not use different notations to distinguish between them; but you should be careful to think about which role σ is playing at each step of the argument.

In terms of any local coordinates (x^i) for M and the corresponding standard coordinates (x^i, ξ_i) for T^*M, the map $\sigma \colon M \to T^*M$ has the coordinate representation

$$\sigma(x^1, \ldots, x^n) = (x^1, \ldots, x^n, \sigma_1(x), \ldots, \sigma_n(x)),$$

where $\sigma_i\,dx^i$ is the coordinate representation of σ as a 1-form. It follows immediately that σ is an immersion, and the fact that it is injective follows from $\pi \circ \sigma = \mathrm{Id}_M$.

To show that it is an embedding, it suffices by Proposition 5.4 to show that it is a proper map. This follows easily from the fact that $\pi\circ\sigma = \mathrm{Id}_M$: If $K \subset T^*M$ is a compact set, then $\sigma^{-1}(K)$ is a closed subset of the compact set $\pi(K)$, and so is compact.

Because $\sigma(M)$ is n-dimensional, it is Lagrangian if and only if it is isotropic, which is the case if and only if $\sigma^*\omega = 0$. The pullback of the tautologous form τ under σ is

$$\sigma^*\tau = \sigma^*(\xi_i\,dx^i) = \sigma_i\,dx^i = \sigma.$$

This can also be seen somewhat more invariantly from the computation

$$(\sigma^*\tau)_p(X) = \tau_{\sigma(p)}(\sigma_*X) = \sigma_p(\pi_*\sigma_*X) = \sigma_p(X),$$

which follows from the definition of τ and the fact that $\pi \circ \sigma = \mathrm{Id}_M$. Therefore,

$$\sigma^*\omega = -\sigma^*d\tau = -d(\sigma^*\tau) = -d\sigma.$$

It follows that σ is a Lagrangian embedding if and only if $d\sigma = 0$. $\square$

Problems

9-1. Let $v_1, \dots, v_n$ be any n vectors in $\mathbb{R}^n$, and let P be the n-dimensional parallelepiped spanned by them:

$$P = \{t_1 v_1 + \cdots + t_n v_n : 0 \le t_i \le 1\}.$$

Show that $\operatorname{Vol}(P) = \det(v_1, \dots, v_n)$.

9-2. Let (e^1, e^2, e^3) be the standard dual basis for $(\mathbb{R}^3)^*$. Show that $e^1 \otimes e^2 \otimes e^3$ is not equal to a sum of an alternating tensor and a symmetric tensor.

9-3. Show that covectors $\omega^1, \dots, \omega^k$ on a finite-dimensional vector space are linearly dependent if and only if $\omega^1 \wedge \cdots \wedge \omega^k = 0$.

9-4. Show that two k-tuples $\{\omega^1, \dots, \omega^k\}$ and $\{\eta^1, \dots, \eta^k\}$ of independent covectors have the same span if and only if

$$\omega^1 \wedge \cdots \wedge \omega^k = c\,\eta^1 \wedge \cdots \wedge \eta^k$$

for some nonzero real number c.

9-5. A k-covector η on a finite-dimensional vector space V is said to be *decomposable* if it can be written

$$\eta = \omega^1 \wedge \cdots \wedge \omega^k,$$

where $\omega^1, \dots, \omega^k$ are covectors. Is every 2-covector on V decomposable? Your answer will depend on the dimension of V.

9-6. Define a 2-form Ω on $\mathbb{R}^3$ by

$$\Omega = x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy.$$

(a) Compute Ω in spherical coordinates (ρ, φ, θ) defined by $(x, y, z) = (\sin\varphi \cos\theta, \sin\varphi \sin\theta, \cos\varphi)$.

(b) Compute $d\Omega$ in both Cartesian and spherical coordinates and verify that both expressions represent the same 3-form.

(c) Compute the restriction $\Omega|_{\mathbb{S}^2} = \iota^*\Omega$, using coordinates (φ, θ), on the open subset where these coordinates are defined.

(d) Show that $\Omega|_{\mathbb{S}^2}$ is nowhere zero.

9-7. In each of the following problems, $g\colon M \to N$ is a smooth map between manifolds M and N, and ω is a differential form on N. In each case, compute $g^*\omega$ and $d\omega$, and verify by direct computation that $g^*(d\omega) = d(g^*\omega)$.

(a) $g\colon \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\begin{aligned}(x, y) = g(s, t) &= (st, e^t);\\ \omega &= x\,dy.\end{aligned}$$

(b) $g\colon \{(r, \theta) : r > 0\} \to \mathbb{R}^2$ by

$$\begin{aligned}(x, y) &= (r\cos\theta, r\sin\theta);\\ \omega &= dy \wedge dx.\end{aligned}$$

(c) $g\colon \mathbb{R}^2 \to \mathbb{R}^3$ by

$$\begin{aligned}(x, y, z) = g(\theta, \varphi) &= ((\cos\varphi + 2)\cos\theta, (\cos\varphi + 2)\sin\theta, \sin\varphi);\\ \omega &= y\,dz \wedge dx.\end{aligned}$$

(d) $g\colon \{(u, v) : u^2 + v^2 < 1\} \to \mathbb{R}^3 \smallsetminus \{0\}$ by

$$\begin{aligned}(x, y, z) &= (u, v, \sqrt{1 - u^2 - v^2});\\ \omega &= (x^2 + y^2 + z^2)^{-3/2}(x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy).\end{aligned}$$

(e) $g\colon \{(r, \theta, \varphi) : r > 0\} \to \mathbb{R}^3$ by

$$\begin{aligned}(x, y, z) &= (r\cos\theta\sin\varphi, r\sin\theta\sin\varphi, r\cos\varphi);\\ \omega &= x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy.\end{aligned}$$

9-8. Let V be a finite-dimensional real vector space. We have two ways to think about the tensor space $T^k(V)$: concretely, as the space of k-multilinear functionals on V; and abstractly, as the tensor product space $V^* \otimes \cdots \otimes V^*$. However, we have defined alternating and symmetric tensors only in terms of the concrete definition. This problem outlines an abstract approach to alternating tensors.

Let $\mathcal{A}$ denote the subspace of $V^* \otimes \cdots \otimes V^*$ spanned by all elements of the form $\alpha \otimes \omega \otimes \omega \otimes \beta$ for covectors ω and arbitrary tensors α, β, and let $A^k(V^*)$ denote the quotient vector space $V^* \otimes \cdots \otimes V^*/\mathcal{A}$. Define a wedge product on $A^k(V^*)$ by $\omega \wedge \eta = \pi(\widetilde{\omega} \otimes \widetilde{\eta})$, where $\pi\colon V^* \otimes \cdots \otimes V^* \to A^k(V^*)$ is the projection, and $\widetilde{\omega}$, $\widetilde{\eta}$ are arbitrary tensors such that $\pi(\widetilde{\omega}) = \omega$, $\pi(\widetilde{\eta}) = \eta$. Show that this wedge product is well defined, and that there is a unique isomorphism $F\colon A^k(V^*) \to \Lambda^k(V)$ such that the following diagram commutes:

$$\begin{tikzcd} V^* \otimes \cdots \otimes V^* \arrow[r, "\cong"] \arrow[d, "\pi"'] & T^k(V) \arrow[d, "\mathrm{Alt}"] \\ A^k(V^*) \arrow[r, "F"'] & \Lambda^k(V), \end{tikzcd}$$

and show that F takes the wedge product we just defined on $A^k(V^*)$ to the Alt convention wedge product on $\Lambda^k(V)$. [This is one reason why some authors consider the Alt convention for the wedge product to be more natural than the determinant convention. It also explains why some authors prefer the notation $\Lambda^k(V^*)$ instead of $\Lambda^k(V)$ for the space of alternating covariant k-tensors, since it is a quotient of the kth tensor product of V^* with itself.]

9-9. Let (V, ω) be a symplectic vector space of dimension $2n$, and let $S \subset V$ be a subspace. Show that V has a symplectic basis (A_i, B_i) with the following property:

(a) If S is symplectic, $S = \operatorname{span}(A_1, B_1, \ldots, A_k, B_k)$ for some k.

(b) If S is isotropic, $S = \operatorname{span}(A_1, \ldots, A_k)$ for some k.

(c) If S is coisotropic, $S = \operatorname{span}(A_1, \ldots, A_n, B_1, \ldots, B_k)$ for some k.

(d) If S is Lagrangian, $S = \operatorname{span}(A_1, \ldots, A_n)$.

9-10. Let M be a smooth manifold, and let N be an embedded submanifold of the total space of T^*M. Show that M is the image of a closed 1-form on M if and only if N is Lagrangian and the projection map $\pi \colon T^*M \to M$ restricts to a diffeomorphism $N \to M$.

9-11. Show that there is no 1-form σ on M such that the tautologous form $\tau \in \Lambda^1(T^*M)$ is equal to the pullback $\pi^*\sigma$.

9-12. Let (M, ω) and $(\widetilde{M}, \widetilde{\omega})$ be symplectic manifolds. Define a 2-form Ω on $M \times \widetilde{M}$ by

$$\Omega = \pi^*\omega - \widetilde{\pi}^*\widetilde{\omega},$$

where $\pi \colon M \times \widetilde{M} \to M$ and $\widetilde{\pi} \colon M \times \widetilde{M} \to \widetilde{M}$ are the projections. For a smooth map $F \colon M \to N$, let $\Gamma(F) \subset M \times \widetilde{M}$ be the graph of F:

$$\Gamma(F) = \{(x, y) \in M \times \widetilde{M} : y = F(x)\}.$$

Show that F is a symplectomorphism if and only if $\Gamma(F)$ is a Lagrangian submanifold of $(M \times \widetilde{M}, \Omega)$.

9-13. The (*real*) *symplectic group* is the subgroup $\mathrm{Sp}(2n, \mathbb{R}) \subset \mathrm{GL}(2n, \mathbb{R})$ of matrices leaving the standard symplectic form $\omega = \sum_{i=1}^n dx^i \wedge dy^i$ invariant, that is, the set of invertible linear maps $A \colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ such that $A^*\omega = \omega$.

(a) Show that a matrix A is in $\mathrm{Sp}(2n, \mathbb{R})$ if and only if it takes the standard basis to a symplectic basis.

(b) Show that $A \in \mathrm{Sp}(2n, \mathbb{R})$ if and only if $A^T J A = J$, where J is given in block form as

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}.$$

(c) Show that $\mathrm{Sp}(2n, \mathbb{R})$ is an embedded Lie subgroup of $\mathrm{GL}(2n, \mathbb{R})$, and determine its dimension.

9-14. Let $\Lambda_n \subset \mathrm{G}(n, 2n)$ denote the set of Lagrangian subspaces of $\mathbb{R}^{2n}$.

(a) Show that $\mathrm{Sp}(2n, \mathbb{R})$ acts transitively on Λ_n.

(b) Show that Λ_n has a unique smooth manifold structure such that the action of $\mathrm{Sp}(2n, \mathbb{R})$ is smooth, and determine its dimension.

10
Integration on Manifolds

We introduced differential forms in the previous chapter with a promise that they would turn out to be objects that can be integrated on manifolds in a coordinate-independent way. In this chapter, we fulfill that promise by defining the integral of a differential n-form over a smooth n-manifold.

Before doing so, however, we need to address a serious issue that we have so far swept under the rug. This is the small matter of the positive and negative signs that arise when we try to interpret a k-form as a machine for measuring k-dimensional volumes. In the previous chapter, we brushed this aside by saying that the value of a k-form applied to k vectors has to be interpreted as a "signed volume" of the parallelepiped spanned by the vectors. These signs will cause problems, however, when we try to integrate differential forms on manifolds, for the simple reason that the transformation law (9.9) for an n-form involves the determinant of the Jacobian, while the change of variables formula for multiple integrals involves the *absolute value* of the determinant. In the first part of this chapter, we develop the theory of orientations, which which is a systematic way to restrict to coordinate transformations with positive determinant, thus eliminating the sign problem.

Next we address the theory of integration. First we define the integral of a differential form over a domain in Euclidean space, and then we show how to use diffeomorphism invariance and partitions of unity to extend this definition to the integral of a compactly supported k-form over an oriented k-manifold. The key feature of the definition is that it is invariant under orientation-preserving diffeomorphisms.

Next we prove one of the most fundamental theorems in all of differential geometry. This is Stokes's theorem, which is a generalization of the fundamental theorem of calculus, as well as of the three great classical theorems of vector analysis: Green's theorem for vector fields in the plane; the divergence theorem for vector fields in space; and (the classical version of) Stokes's theorem for surface integrals in $\mathbb{R}^3$. We also describe an extension of Stokes's theorem to manifolds with corners, which will be useful in our treatment of the de Rham theorem in Chapter 11.

In the last section of the chapter, we show how these ideas play out on a Riemannian manifold. In particular, we prove Riemannian versions of the divergence theorem and of Stokes's theorem for surface integrals, of which the classical theorems are special cases.

Orientations

The word "orientation" has some familiar meanings from our everyday experience, which can be interpreted as rules for singling out certain bases of $\mathbb{R}^1$, $\mathbb{R}^2$, and $\mathbb{R}^3$. For example, most people would understand an "orientation" of a line to mean a choice of preferred direction along the line, so we might declare an oriented basis for $\mathbb{R}^1$ to be one that points to the right (i.e., in the positive direction). A natural class of bases for $\mathbb{R}^2$ is the ones for which the rotation from the first vector to the second is in the counterclockwise direction. And every student of vector calculus encounters "right-handed" bases in $\mathbb{R}^3$: These are the bases (E_1, E_2, E_3) with the property that when the fingers of your right hand curl from E_1 to E_2, your thumb points in the direction of E_3.

Although "to the right," "counterclockwise," and "right-handed" are not mathematical terms, it is easy to translate the rules for selecting oriented bases of $\mathbb{R}^1$, $\mathbb{R}^2$, and $\mathbb{R}^3$ into rigorous mathematical terms: You can check that in all three cases, the preferred bases are the ones whose transition matrix from the standard basis has positive determinant.

In an abstract vector space for which there is no canonical basis, we no longer have any way to determine which bases are "correctly oriented." For example, if V is the space of polynomials of degree at most 2, who is to say which of the ordered bases $(1, x, x^2)$ or $(x^2, x, 1)$ is "right-handed"? All we can say in general is what it means for two bases to have the "same orientation."

Thus we are led to introduce the following definition. Let V be a vector space of dimension $n \geq 1$. We say two ordered bases $(E_1, \dots, E_n)$ and $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ are *consistently oriented* if the transition matrix (B^i_j), defined by

$$E_i = B^j_i \widetilde{E}_j, \tag{10.1}$$

has positive determinant.

Exercise 10.1. Show that being consistently oriented is an equivalence relation on the set of all ordered bases for V, and show that there are exactly two equivalence classes.

If $\dim V = n \geq 1$, we define an *orientation* for V as an equivalence class of ordered bases. If $(E_1, \dots, E_n)$ is any ordered basis for V, we denote the orientation that it determines by $[E_1, \dots, E_n]$. A vector space together with a choice of orientation is called an *oriented vector space*. If V is oriented, then any ordered basis $(E_1, \dots, E_n)$ that is in the given orientation is said to be *oriented* or *positively oriented*. A basis that is not oriented is said to be *negatively oriented*.

For the special case of a 0-dimensional vector space V, we define an orientation of V to be simply a choice of one of the numbers ± 1.

Example 10.1. The orientation $[e_1, \dots, e_n]$ of $\mathbb{R}^n$ determined by the standard basis is called the *standard orientation*. You should convince yourself that, in our usual way of representing the axes graphically, an oriented basis for $\mathbb{R}$ is one that points to the right; an oriented basis for $\mathbb{R}^2$ is one for which the rotation from the first vector to the second is counterclockwise; and an oriented basis for $\mathbb{R}^3$ is a right-handed one. (These can be taken as mathematical definitions for the words "right," "counterclockwise," and "right-handed.") The standard orientation for $\mathbb{R}^0$ is defined to be $+1$.

There is an important connection between orientations and alternating tensors, expressed in the following lemma.

Lemma 10.2. *Let V be a vector space of dimension $n \geq 1$, and suppose Ω is a nonzero element of $\Lambda^n(V)$. The set of ordered bases $(E_1, \dots, E_n)$ such that $\Omega(E_1, \dots, E_n) > 0$ is an orientation for V.*

Proof. Let $\mathcal{O}_\Omega$ denote the set of ordered bases on which Ω gives positive values. We need to show that $\mathcal{O}_\Omega$ is exactly one equivalence class.

Choose one basis $(E_1, \dots, E_n)$ such that $\Omega(E_1, \dots, E_n) > 0$. (Such a basis can always be found by starting with an arbitrary basis and replacing E_1 by $-E_1$ if necessary.) Let $(\varepsilon^1, \dots, \varepsilon^n)$ denote the dual basis. Since $\varepsilon^1 \wedge \dots \wedge \varepsilon^n$ is a basis for $\Lambda^n(V)$, there is some (necessarily positive) number c such that $\Omega = c\varepsilon^1 \wedge \dots \wedge \varepsilon^n$.

Now if $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ is any other basis, with transition matrix (A^i_j) defined by

$$\widetilde{E}_j = A^i_j E_i,$$

we have

$$\begin{aligned}
\Omega(\widetilde{E}_1, \dots, \widetilde{E}_n) &= c\varepsilon^1 \wedge \dots \wedge \varepsilon^n(\widetilde{E}_1, \dots, \widetilde{E}_n) \\
&= c \det(\varepsilon^i(\widetilde{E}_j)) \\
&= c \det(A^i_j).
\end{aligned}$$

It follows that $(\widetilde{E}_j)$ is consistently oriented with (E_i) if and only if $\Omega(\widetilde{E}_1, \dots, \widetilde{E}_n) > 0$, and therefore $\mathcal{O}_\Omega = [E_1, \dots, E_n]$. $\square$

If V is an oriented vector space and Ω is an n-covector that determines the orientation of V as described in this lemma, we say that Ω is an *oriented* (or *positively oriented*) n-covector. For example, the n-covector $e^1 \wedge \dots \wedge e^n$ is positively oriented for the standard orientation on $\mathbb{R}^n$.

Orientations of Manifolds

Let M be a smooth manifold. We define a *pointwise orientation* on M to be a choice of orientation of each tangent space. By itself, this is not a very useful concept, because the orientations of nearby points may have no relation to each other. For example, a pointwise orientation on $\mathbb{R}^n$ might switch randomly from point to point between the standard orientation and its opposite. In order for orientations to have some relationship with the smooth structure, we need an extra condition to ensure that the orientations of nearby tangent spaces are consistent with each other.

Suppose M is a smooth n-manifold with a given pointwise orientation. Recall that a local frame for M is an n-tuple of smooth vector fields $(E_1, \dots, E_n)$ on an open set $U \subset M$ such that $(E_i|_p)$ forms a basis for T_pM at each $p \in U$. We say that a local frame (E_i) is (*positively*) *oriented* if $(E_1|_p, \dots, E_n|_p)$ is a positively oriented basis for T_pM at each point $p \in U$. A *negatively oriented* frame is defined analogously.

A pointwise orientation is said to be *continuous* if every point is in the domain of an oriented local frame. An *orientation* of M is a continuous pointwise orientation. An *oriented manifold* is a smooth manifold together with a choice of orientation. We say M is *orientable* if there exists an orientation for it, and *nonorientable* if not.

If M is 0-dimensional, this definition just means that an orientation of M is a choice of ± 1 attached to each of its points. The local constancy condition is vacuous in this case, and the notion of oriented frames is not useful. Clearly every 0-manifold is orientable.

Exercise 10.2. If M is an oriented manifold of dimension $n \geq 1$, show that every local frame with connected domain is either positively oriented or negatively oriented.

The next two propositions give ways of specifying orientations on manifolds that are somewhat more practical to use than the definition. A smooth coordinate chart (U, φ) is said to be (*positively*) *oriented* if the coordinate frame $(\partial/\partial x^i)$ is positively oriented, and *negatively oriented* if the coordinate frame is negatively oriented. A collection of charts $\{(U_\alpha, \varphi_\alpha)\}$ is said to be *consistently oriented* if for each α, β, the transition map $\varphi_\beta \circ \varphi_\alpha^{-1}$ has positive Jacobian determinant everywhere on $\varphi_\alpha(U_\alpha \cap U_\beta)$.

Proposition 10.3. *Let M be a smooth positive-dimensional manifold, and suppose we are given an open cover of M by consistently oriented charts $\{(U_\alpha, \varphi_\alpha)\}$. Then there is a unique orientation for M with the property that each chart φ_α is oriented. Conversely, if M is oriented, then the collection of all oriented charts is a consistently oriented cover of M.*

Proof. For any $p \in M$, the consistency condition means that the transition matrix between the coordinate bases determined by any two of the charts in the given collection has positive determinant. Thus the coordinate bases for all of the given charts determine the same orientation on T_pM. This defines a pointwise orientation on M. Each point of M is in the domain of at least one of the given charts, and the corresponding coordinate frame is oriented by definition, so this pointwise orientation is continuous. The converse is similar, and is left as an exercise. □

Exercise 10.3. Complete the proof of Proposition 10.3.

Proposition 10.4. *Let M be a smooth manifold of dimension $n \geq 1$. A nonvanishing n-form $\Omega \in \mathcal{A}^n(M)$ determines a unique orientation of M for which Ω is positively oriented at each point. Conversely, if M is given an orientation, then there is a nonvanishing n-form on M that is positively oriented at each point.*

Remark. Because of this proposition, any nonvanishing n-form on an n-manifold is called an *orientation form.* If M is an oriented manifold and Ω is an orientation form determining the given orientation, we also say that Ω is (*positively*) *oriented*. It is easy to check that if Ω and $\widetilde{\Omega}$ are two positively oriented forms on the same orientated manifold M, then $\widetilde{\Omega} = f\Omega$ for some strictly positive smooth function f.

If M is a 0-manifold, the proposition remains true if we interpret an orientation form as a nonvanishing function Ω, which assigns the orientation $+1$ to points where $\Omega > 0$ and -1 to points where $\Omega < 0$.

Proof. Let Ω be a nonvanishing n-form on M. Then Ω defines a pointwise orientation by Lemma 10.2, so all we need to check is that it is continuous. Let (x^i) be any local coordinates on a connected domain $U \subset M$. Writing $\Omega = f\, dx^1 \wedge \cdots \wedge dx^n$ on U, the fact that Ω is nonvanishing means that f is nonvanishing, and therefore

$$\Omega\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) = f \neq 0$$

at all points of U. Since U is connected, it follows that this expression is either always positive or always negative on U, and therefore the coordinate chart is either positively oriented or negatively oriented. If negatively, we can replace x^1 by $-x^1$ to obtain a new coordinate chart for which the coordinate frame is positively oriented. Thus the pointwise orientation determined by Ω is continuous.

Conversely, suppose M is oriented. Let $\{(U_\alpha, \varphi_\alpha)\}$ be the collection of all oriented charts for M. In each coordinate domain U_α, the n-form $\Omega_\alpha = dx^1 \wedge \cdots \wedge dx^n$ is positively oriented. Let $\{\psi_\alpha\}$ be a partition of unity subordinate to the cover $\{U_\alpha\}$, and define

$$\Omega = \sum_\alpha \psi_\alpha \Omega_\alpha.$$

By the usual argument, Ω is a smooth n-form on M. To complete the proof, we need to show that Ω never vanishes.

Let $p \in M$ be arbitrary, and let $(E_1, \ldots, E_n)$ be an oriented basis for T_pM. For each α such that $p \in U_\alpha$, we have $\Omega_\alpha|_p(E_1, \ldots, E_n) > 0$. Since $\psi_\alpha(p) = 0$ for all other α and there is at least one α for which $\psi_\alpha(p) > 0$, we have

$$\Omega_p(E_1, \ldots, E_n) = \sum_{\{\alpha : p \in U_\alpha\}} \psi_\alpha(p) \Omega_\alpha|_p(E_1, \ldots, E_n) > 0,$$

Thus $\Omega_p \neq 0$. □

Exercise 10.4. Show that any open subset of an orientable manifold is orientable, and any product of orientable manifolds is orientable.

Recall that a smooth manifold is said to be parallelizable if it admits a global frame.

Proposition 10.5. *Every parallelizable manifold is orientable.*

Proof. Suppose M is parallelizable, and let $(E_1, \ldots, E_n)$ be a global frame for M. Define a pointwise orientation by declaring $(E_1|_p, \ldots, E_n|_p)$ to be positively oriented at each $p \in M$. This pointwise orientation is continuous, because every point of M is in the domain of the (global) oriented frame (E_i). □

Example 10.6. The preceding proposition shows that Euclidean spaces $\mathbb{R}^n$, the n-torus $\mathbb{T}^n$, the spheres $\mathbb{S}^1$ and $\mathbb{S}^3$, and products of them are all orientable, because they are all parallelizable. Therefore any open subset of one of these manifolds is also orientable.

Let M and N be oriented positive-dimensional manifolds, and let $F\colon M \to N$ be a local diffeomorphism. We say F is *orientation-preserving* if for each $p \in M$, F_* takes oriented bases of T_pM to oriented bases of $T_{F(p)}N$, and *orientation-reversing* if it takes oriented bases of T_pM to negatively oriented bases of $T_{F(p)}N$.

Exercise 10.5. Show that a smooth map $F\colon M \to N$ is orientation-preserving if and only if its Jacobian matrix with respect to any oriented coordinate charts for M and N has positive determinant, and orientation-reversing if and only if it has negative determinant.

Orientations of Hypersurfaces

If M is an oriented manifold and N is a submanifold of M, N may not inherit an orientation from M, even if N is embedded. Clearly it is not sufficient to restrict an orientation form from M to N, since the restriction of an n-form to a manifold of lower dimension must necessarily be zero. A useful example to consider is the Möbius band, which is not orientable (see Problem 10-5), even though it can be embedded in $\mathbb{R}^3$.

In this section, we will restrict our attention to (immersed or embedded) submanifolds of codimension 1, commonly called *hypersurfaces*. With one extra piece of information (a certain kind of vector field along the hypersurface), we can use an orientation on M to induce an orientation on any hypersurface $N \subset M$.

We start with some definitions. Let V be a finite-dimensional vector space, and let $X \in V$. We define a linear map $i_X \colon \Lambda^k V \to \Lambda^{k-1} V$, called *interior multiplication* or *contraction* with X, by

$$i_X \omega(Y_1, \ldots, Y_{k-1}) = \omega(X, Y_1, \ldots, Y_{k-1}).$$

In other words, $i_X \omega$ is obtained from ω by inserting X into the first slot. By convention, we interpret $i_X \omega$ to be zero when ω is a 0-covector (i.e., a number). Another common notation is

$$X \lrcorner\, \omega = i_X \omega.$$

Interior multiplication shares two important properties with exterior differentiation: They are both antiderivations whose square is zero, as the following lemma shows.

Lemma 10.7. *Let V be a finite-dimensional vector space and $X \in V$.*

(*a*) $i_X \circ i_X = 0$.

(*b*) i_X *is an antiderivation: If ω is a k-covector and η is an l-covector,*

$$i_X(\omega \wedge \eta) = (i_X \omega) \wedge \eta + (-1)^k \omega \wedge (i_X \eta).$$

Proof. On k-covectors for $k \geq 2$, part (a) is immediate from the definition, because any alternating tensor gives zero when two of its arguments are identical. On 1-covectors and 0-covectors, it follows from the fact that $i_X \equiv 0$ on 0-covectors.

To prove (b), it suffices to consider the case in which both ω and η are wedge products of covectors (such an alternating tensor is said to be *decomposable*), since every alternating tensor can be written locally as a linear combination of decomposable ones. It is easy to verify that (b) will follow in this special case from the following general formula for covectors

$\omega^1, \dots, \omega^k$:

$$X \lrcorner (\omega^1 \wedge \cdots \wedge \omega^k) = \sum_{i=1}^{k} (-1)^{i-1} \omega^i(X) \omega^1 \wedge \cdots \wedge \widehat{\omega^i} \wedge \cdots \wedge \omega^k, \quad (10.2)$$

where the hat indicates that ω^i is omitted.

To prove (10.2), let us write $X_1 = X$ and apply both sides to vectors $(X_2, \dots, X_k)$; then what we have to prove is

$$\begin{aligned}(\omega^1 \wedge \cdots \wedge \omega^k)&(X_1, \dots, X_k) \\ &= \sum_{i=1}^{k} (-1)^{i-1} \omega^i(X_1)(\omega^1 \wedge \cdots \wedge \widehat{\omega^i} \wedge \cdots \wedge \omega^k)(X_2, \dots, X_k). \quad (10.3)\end{aligned}$$

The left-hand side of (10.3) is the determinant of the matrix $\mathbb{X}$ whose (i, j)-entry is $\omega^i(X_j)$. To simplify the right-hand side, let $\mathbb{X}^i_j$ denote the $(k-1) \times (k-1)$ minor of $\mathbb{X}$ obtained by deleting the ith row and jth column. Then the right-hand side of (10.3) is

$$\sum_{i=1}^{k} (-1)^{i-1} \omega^i(X_1) \det \mathbb{X}^i_1.$$

This is just the expansion of $\det \mathbb{X}$ by minors along the first column $(\omega^1(X_1), \dots, \omega^k(X_1))$, and therefore is equal to $\det \mathbb{X}$. $\square$

It should be noted that when the wedge product is defined using the Alt convention, interior multiplication has to be defined with an extra factor of k:

$$\bar{i}_X \omega(Y_1, \dots, Y_{k-1}) = k\omega(X, Y_1, \dots, Y_{k-1}).$$

This definition ensures that interior multiplication i_X is still an antiderivation; the factor of k is needed to compensate for the difference between the factors of $1/k!$ and $1/(k-1)!$ that occur when the left-hand and right-hand sides of (10.3) are evaluated using the Alt convention.

On a smooth manifold M, interior multiplication extends naturally to vector fields and differential forms, simply by letting i_X act pointwise: if $X \in \mathfrak{T}(M)$ and $\omega \in \mathcal{A}^k(M)$, define a $(k-1)$-form $X \lrcorner \omega = i_X \omega$ by

$$(X \lrcorner \omega)_p = X_p \lrcorner \omega_p.$$

Exercise 10.6. If X is a smooth vector field and ω is a differential form, show that $X \lrcorner \omega$ is smooth.

Now suppose M is a smooth manifold and $S \subset M$ is a hypersurface (immersed or embedded). A *vector field along S* is a continuous map

$N: S \to TM$ with the property that $N_p \in T_pM$ for each $p \in S$. (Note the difference between this and a vector field *on* S, which would have the property that $N_p \in T_pS$ at each point.) A vector $N_p \in T_pM$ is said to be *transverse* (to S) if T_pM is spanned by N_p and T_pS for each $p \in S$. Similarly, a vector field N along S is transverse if N_p is transverse at each $p \in S$.

For example, any smooth vector field on M restricts to a smooth vector field along S; it is transverse if and only if it is nowhere tangent to S.

Proposition 10.8. *Suppose M is an oriented smooth n-manifold, S is an immersed hypersurface in M, and N is a smooth transverse vector field along S. Then S has a unique orientation with the property that $(E_1, \dots, E_{n-1})$ is an oriented basis for T_pS if and only if $(N_p, E_1, \dots, E_{n-1})$ is an oriented basis for T_pM. If Ω is an orientation form for M, then $(N \lrcorner \Omega)|_S$ is an orientation form for S with respect to this orientation.*

Remark. When $n = 1$, since S is a 0-manifold, this proposition should be interpreted as follows: At each point $p \in S$, we assign the orientation $+1$ to p if N_p is an oriented basis for T_pM, and -1 if N_p is negatively oriented. With this understanding, the proof below goes through in this case without modification.

Proof. Let Ω be a smooth orientation form for M. Then $\omega = (N \lrcorner \Omega)|_S$ is a smooth $(n-1)$-form on S. It will be an orientation form for S if we can show that it never vanishes. Given any basis $(E_1, \dots, E_{n-1})$ for T_pS, the fact that N is transverse to S implies that $(N_p, E_1, \dots, E_{n-1})$ is a basis for T_pM. The fact that Ω is nonvanishing implies that

$$\omega_p(E_1, \dots, E_{n-1}) = \Omega_p(N_p, E_1, \dots, E_{n-1}) \neq 0.$$

Since $\omega_p(E_1, \dots, E_{n-1}) > 0$ if and only if $\Omega_p(N_p, E_1, \dots, E_n) > 0$, the orientation determined by ω is the one defined in the statement of the proposition. ☐

Example 10.9. Considering $\mathbb{S}^n$ as a hypersurface in $\mathbb{R}^{n+1}$, the vector field $N = x^i \partial/\partial x^i$ along $\mathbb{S}^n$ is easily seen to be transverse. (In fact, this vector field is orthogonal to $\mathbb{S}^n$ with respect to the Euclidean metric.) Thus it induces an orientation on $\mathbb{S}^n$. This shows that all spheres are orientable. (The orientation on $\mathbb{S}^0$ given by this construction is the one that assigns the orientation $+1$ to the point $+1 \in \mathbb{S}^0$ and -1 to $-1 \in \mathbb{S}^0$.)

Manifolds with Boundary

An important application of this construction is to define a canonical orientation on the boundary of any oriented manifold with boundary. First, we note that an orientation of a smooth manifold with boundary can be defined exactly as in the case of a smooth manifold, with "chart" replaced by "generalized chart" as necessary.

One situation that arises frequently is the following. If M is a smooth n-manifold, a compact, embedded n-dimensional submanifold with boundary $D \subset M$ is called a *regular domain* in M. An orientation on M immediately yields an orientation on D, for example by restricting an orientation n-form to D. Examples are the closed unit ball in $\mathbb{R}^n$ and the closed upper hemisphere in $\mathbb{S}^n$, each of which inherits an orientation from its containing manifold.

If M is a smooth manifold with boundary, ∂M is easily seen to be an embedded hypersurface in M. Recall that any point $p \in M$ is in the domain of a generalized chart (U, φ), which means that φ is a diffeomorphism from U onto an open subset $\widetilde{U} \subset \mathbb{H}^n$. Since ∂M is locally characterized by $x^n = 0$ in such charts, generalized charts play a role for ∂M analogous to slice charts for ordinary embedded submanifolds.

Let $p \in \partial M$. A vector $N \in T_pM$ is said to be *inward-pointing* if $N \notin T_p\partial M$ and for some $\varepsilon > 0$ there exists a smooth curve segment $\gamma\colon [0, \varepsilon] \to M$ such that $\gamma(0) = p$ and $\gamma'(0) = N$. It is said to be *outward-pointing* if $-N$ is inward-pointing. The following lemma gives another characterization of inward-pointing vectors, which is usually much easier to check.

Lemma 10.10. *Suppose M is a smooth manifold with boundary and $p \in \partial M$. A vector $N \in T_pM$ is inward-pointing if and only if N has strictly positive x^n-component in every generalized chart $(x^1, \dots, x^n)$.*

Exercise 10.7. Prove Lemma 10.10.

A vector field along ∂M (defined just as for ordinary hypersurfaces) is said to be inward-pointing or outward-pointing if its value at each point has that property.

Lemma 10.11. *If M is any smooth manifold with boundary, there is a smooth outward-pointing vector field along ∂M.*

Proof. Cover a neighborhood of ∂M by generalized coordinate charts $\{(U_\alpha, \varphi_\alpha)\}$. In each such chart U_α, $N_\alpha = -\partial/\partial x^n|_{\partial M \cap U_\alpha}$ is a smooth vector field along $\partial M \cap U_\alpha$, which is outward-pointing by Lemma 10.10. Let $\{\psi_\alpha\}$ be a partition of unity subordinate to the cover $\{U_\alpha \cap \partial M\}$ of ∂M, and define a global vector field N along ∂M by

$$N = \sum_\alpha \psi_\alpha N_\alpha.$$

Clearly N is a smooth vector field along ∂M. To show that it is outward-pointing, let $(y^1, \dots, y^n)$ be any generalized coordinates in a neighborhood of $p \in \partial M$. Because each N_α is outward-pointing, it satisfies $dy^n(N_\alpha) < 0$.

Therefore, the y^n-component of N at p satisfies

$$dy^n(N_p) = \sum_\alpha \psi_\alpha(p) dy^n(N_\alpha|_p).$$

This sum is strictly negative, because each term is nonpositive and at least one term is negative. □

Proposition 10.12 (The Induced Orientation on a Boundary). *Let M be an oriented smooth manifold with boundary. The orientation on ∂M determined by any outward-pointing vector field along ∂M is independent of the choice of vector field.*

Remark. As a consequence of this proposition, there is a unique orientation on ∂M determined by the orientation of M. We call this orientation the *induced orientation* or the *Stokes orientation* on ∂M. (The second term is chosen because of the role this orientation will play in Stokes's theorem, to be described later in this chapter.)

Proof. Let Ω be an orientation form for M, and let $(x^1, \ldots, x^n)$ be generalized coordinates for M in a neighborhood of $p \in M$. Replacing x^1 by $-x^1$ if necessary, we may assume they are oriented coordinates, which implies that $\Omega = f\, dx^1 \wedge \cdots \wedge dx^n$ (locally) for some strictly positive function f. Suppose N is an outward-pointing vector field along ∂M. The orientation of ∂M determined by N is given by the orientation form $(N \lrcorner \Omega)|_{\partial M}$.

Because $x^n = 0$ along ∂M, the restriction $dx^n|_{\partial M}$ is equal to zero (Problem 5-16). Therefore, using the antiderivation property of i_N,

$$\begin{aligned}(N \lrcorner \Omega)|_{\partial M} &= f \sum_{i=1}^n (-1)^{i-1} dx^i(N) dx^i|_{\partial M} \wedge \cdots \wedge \widehat{dx^i}|_{\partial M} \wedge \cdots \wedge dx^n|_{\partial M} \\ &= (-1)^{n-1} f dx^n(N) dx^i|_{\partial M} \wedge \cdots \wedge dx^{n-1}|_{\partial M}.\end{aligned}$$

Because $dx^n(N) = N^n < 0$, this is a positive multiple of $(-1)^n dx^i|_{\partial M} \wedge \cdots \wedge dx^{n-1}|_{\partial M}$. If $\widetilde{N}$ is any other outward-pointing vector field, the same computation shows that $(\widetilde{N} \lrcorner \Omega)|_{\partial M}$ is a positive multiple of the same $(n-1)$-form, and thus a positive multiple of $(N \lrcorner \Omega)|_{\partial M}$. This proves that N and $\widetilde{N}$ determine the same orientation on ∂M. □

Example 10.13. This proposition gives a simpler proof that $\mathbb{S}^n$ is orientable, because it is the boundary of the closed unit ball.

Example 10.14. Let us determine the induced orientation on $\partial \mathbb{H}^n$ when $\mathbb{H}^n$ itself has the standard orientation inherited from $\mathbb{R}^n$. We can identify $\partial \mathbb{H}^n$ with $\mathbb{R}^n$ under the correspondence $(x^1, \ldots, x^{n-1}, 0) \leftrightarrow (x^1, \ldots, x^{n-1})$. Since the vector field $-\partial/\partial x^n$ is outward-pointing along $\partial \mathbb{H}^n$, the standard coordinate frame for $\mathbb{R}^{n-1}$ is positively oriented for $\partial \mathbb{H}^n$ if and only if $[-\partial/\partial x^n, \partial/\partial x^1, \dots, \partial/\partial x^{n-1}]$ is the standard orientation for $\mathbb{R}^n$. This orientation satisfies

$$\begin{aligned}[-\partial/\partial x^n, \partial/\partial x^1, \dots, \partial/\partial x^{n-1}] &= -[\partial/\partial x^n, \partial/\partial x^1, \dots, \partial/\partial x^{n-1}] \\ &= (-1)^n [\partial/\partial x^1, \dots, \partial/\partial x^{n-1}, \partial/\partial x^n].\end{aligned}$$

Thus the induced orientation on $\partial\mathbb{H}^n$ is equal to the standard orientation on $\mathbb{R}^{n-1}$ when n is even, but it is *opposite* to the standard orientation when n is odd. In particular, the standard coordinates on $\partial\mathbb{H}^n \approx \mathbb{R}^{n-1}$ are positively oriented if and only if n is even. (This fact will play an important role in the proof of Stokes's theorem below.)

Integration of Differential Forms

In this section, we will define in an invariant way the integrals of differential forms over manifolds. You should be sure you are familiar with the basic properties of multiple integrals in $\mathbb{R}^n$, as summarized in the Appendix.

We begin by considering differential forms on subsets of $\mathbb{R}^n$. For the time being, let us restrict attention to the case $n \geq 1$. Let $D \subset \mathbb{R}^n$ be a compact domain of integration, and let ω be a smooth n-form on D. (Remember, this means that ω has a smooth extension to some open set containing D.) Any such form can be written as

$$\omega = f\, dx^1 \wedge \cdots \wedge dx^n$$

for a function $f \in C^\infty(D)$. We define the *integral of ω over D* to be

$$\int_D \omega = \int_D f\, dV.$$

This can be written more suggestively as

$$\int_D f\, dx^1 \wedge \cdots \wedge dx^n = \int_D f\, dx^1 \cdots dx^n.$$

In simple terms, to compute the integral of a form such as $f\, dx^1 \wedge \cdots \wedge dx^n$, we just "erase the wedges"!

Somewhat more generally, let U be an open set in $\mathbb{R}^n$. We would like to define the integral of any compactly supported n-form ω over U. However, since neither U nor $\operatorname{supp}\omega$ may be a domain of integration in general, we need the following lemma.

Lemma 10.15. *Suppose $K \subset U \subset \mathbb{R}^n$, where U is an open set and K is compact. Then there is a compact domain of integration D such that $K \subset D \subset U$.*

Proof. For each $p \in K$, there is an open ball containing p whose closure is contained in U. By compactness, finitely many such open balls $U_1, \dots, U_m$ cover K. Since the boundary of an open ball is a codimension-1 submanifold, it has measure zero by Theorem 6.6, and so each ball is a domain of integration. The set $D = \overline{U}_1 \cup \cdots \cup \overline{U}_m$ is the required domain of integration. □

Now if $U \subset \mathbb{R}^n$ is open and ω is a compactly supported n-form on U, we define

$$\int_U \omega = \int_D \omega,$$

where D is any domain of integration such that $\operatorname{supp}\omega \subset D \subset U$. It is an easy matter to verify that this definition does not depend on the choice of D. Similarly, if V is an open subset of the upper half-space $\mathbb{H}^n$ and ω is a compactly supported n-form on V, we define

$$\int_V \omega = \int_{D \cap \mathbb{H}^n} \omega,$$

where D is chosen in the same way.

It is worth remarking that it is possible to extend the definition to integrals of noncompactly supported forms, and integrals of such forms play an important role in many applications. However, in such cases the resulting multiple integrals are improper, so one must pay close attention to convergence issues. For the purposes we have in mind, the compactly supported case will be more than sufficient.

The motivation for this definition is expressed in the following proposition.

Proposition 10.16. *Let D and E be domains of integration in $\mathbb{R}^n$, and let ω be an n-form on E. If $G\colon D \to E$ is a smooth map whose restriction to* $\operatorname{Int} D$ *is an orientation-preserving or orientation-reversing diffeomorphism onto* $\operatorname{Int} E$*, then*

$$\int_E \omega = \begin{cases} \displaystyle\int_D G^*\omega & \textit{if } G \textit{ is orientation-preserving,} \\ \displaystyle -\int_D G^*\omega & \textit{if } G \textit{ is orientation-reversing.} \end{cases}$$

Proof. Let us use $(y^1, \dots, y^n)$ to denote standard coordinates on E, and $(x^1, \dots, x^n)$ to denote those on D. Suppose first that G is orientation-preserving. Writing $\omega = f\, dy^1 \wedge \cdots \wedge dy^n$, the change of variables formula together with the formula of Lemma 9.11 for pullbacks of n-forms yield

$$\begin{aligned}\int_E \omega &= \int_E f\,dV \\ &= \int_D (f\circ G)\left|\det\left(\frac{\partial G^i}{\partial x^j}\right)\right| dV \\ &= \int_D (f\circ G)\det\left(\frac{\partial G^i}{\partial x^j}\right) dV \\ &= \int_D (f\circ G)\det\left(\frac{\partial G^i}{\partial x^j}\right) dx^1\wedge\cdots\wedge dx^n \\ &= \int_D G^*\omega.\end{aligned}$$

If G is orientation-reversing, the same computation holds except that a negative sign is introduced when the absolute value signs are removed. □

Corollary 10.17. *Suppose U, V are open subsets of $\mathbb{R}^n$, $G\colon U \to V$ is an orientation-preserving diffeomorphism, and ω is a compactly supported n-form on V. Then*

$$\int_V \omega = \int_U G^*\omega.$$

Proof. Let $E \subset V$ be a domain of integration containing $\operatorname{supp}\omega$. Since smooth maps take sets of measure zero to sets of measure zero, $D = G^{-1}(E) \subset U$ is a domain of integration containing $\operatorname{supp} G^*\omega$. Therefore, the result follows from the preceding proposition. □

Using this result, it is easy to make invariant sense of the integral of a differential form over an oriented manifold. Let M be a smooth, oriented n-manifold and ω an n-form on M. Suppose first that ω is compactly supported in the domain of a single oriented coordinate chart (U,φ). We define the *integral of ω over M* to be

$$\int_M \omega = \int_{\varphi(U)} (\varphi^{-1})^*\omega.$$

Since $(\varphi^{-1})^*\omega$ is a compactly supported n-form on the open subset $\varphi(U) \subset \mathbb{R}^n$, its integral is defined as discussed above.

Proposition 10.18. *With ω as above, $\int_M \omega$ does not depend on the choice of oriented coordinate chart whose domain contains* $\operatorname{supp}\omega$*.*

Proof. Suppose $(\widetilde{U},\widetilde{\varphi})$ is another oriented chart such that $\operatorname{supp}\omega \subset \widetilde{U}$. Because $\widetilde{\varphi}\circ\varphi^{-1}$ is an orientation-preserving diffeomorphism from $\varphi(U\cap\widetilde{U})$

to $\widetilde{\varphi}(U \cap \widetilde{U})$, Corollary 10.17 implies that

$$\begin{aligned}
\int_{\widetilde{\varphi}(\widetilde{U})} (\widetilde{\varphi}^{-1})^*\omega &= \int_{\widetilde{\varphi}(U\cap\widetilde{U})} (\widetilde{\varphi}^{-1})^*\omega \\
&= \int_{\varphi(U\cap\widetilde{U})} (\widetilde{\varphi}\circ\varphi^{-1})^*(\widetilde{\varphi}^{-1})^*\omega \\
&= \int_{\varphi(U\cap\widetilde{U})} (\varphi^{-1})^*(\widetilde{\varphi})^*(\widetilde{\varphi}^{-1})^*\omega \\
&= \int_{\varphi(U)} (\varphi^{-1})^*\omega.
\end{aligned}$$

Thus the two definitions of $\int_M \omega$ agree. □

If M is an oriented smooth n-manifold with boundary, and ω is an n-form on M that is compactly supported in the domain of a generalized chart, the definition of $\int_M \omega$ and the statement and proof of Proposition 10.18 go through unchanged, provided we interpret all the coordinate charts as generalized charts, and compute the integrals over open subsets of $\mathbb{H}^n$ in the way we described above.

To integrate over an entire manifold, we simply apply this same definition together with a partition of unity. Suppose M is an oriented smooth n-manifold (possibly with boundary) and ω is a compactly supported n-form on M. Let $\{(U_i, \varphi_i)\}$ be a finite cover of $\operatorname{supp}\omega$ by oriented coordinate charts (generalized charts if M has a nonempty boundary), and let $\{\psi_i\}$ be a subordinate partition of unity. We define the *integral of ω over M* to be

$$\int_M \omega = \sum_i \int_M \psi_i\omega. \tag{10.4}$$

Since for each i, the n-form $\psi_i\omega$ is compactly supported in U_i, each of the terms in this sum is well-defined according to our discussion above. To show that the integral is well-defined, therefore, we need only examine the dependence on the charts and the partition of unity.

Lemma 10.19. *The definition of $\int_M \omega$ given above does not depend on the choice of oriented charts or partition of unity.*

Proof. Suppose $\{(\widetilde{U}_j, \widetilde{\varphi}_j)\}$ is another finite collection of oriented charts whose domains cover $\operatorname{supp}\omega$, and $\{\widetilde{\psi}_j\}$ is a subordinate partition of unity. For each i, we compute

$$\begin{aligned}
\int_M \psi_i\omega &= \int_M \Big(\sum_j \widetilde{\psi}_j\Big)\psi_i\omega \\
&= \sum_j \int_M \widetilde{\psi}_j\psi_i\omega.
\end{aligned}$$

Summing over i, we obtain

$$\sum_i \int_M \psi_i \omega = \sum_{i,j} \int_M \widetilde{\psi}_j \psi_i \omega.$$

Observe that each term in this last sum is the integral of a form compactly supported in a single (generalized) chart (U_i, for example), so by Proposition 10.18 each term is well-defined, regardless of which coordinate map we use to compute it. The same argument, starting with $\int_M \widetilde{\psi}_j \omega$, shows that

$$\sum_j \int_M \widetilde{\psi}_j \omega = \sum_{i,j} \int_M \widetilde{\psi}_j \psi_i \omega.$$

Thus both definitions yield the same value for $\int_M \omega$. $\square$

As usual, we have a special definition in the 0-dimensional case. The integral of a compactly supported 0-form (i.e., a function) f over an oriented 0-manifold M is defined to be the sum

$$\int_M f = \sum_{p \in M} \pm f(p),$$

where we take the positive sign at points where the orientation is positive and the negative sign at points where it is negative. The assumption that f is compactly supported implies that there are only finitely many nonzero terms in this sum.

If $N \subset M$ is an oriented immersed k-dimensional submanifold (with or without boundary), and ω is a k-form on M, we interpret $\int_N \omega$ to mean $\int_N (\omega|_N)$. In particular, if M is a compact oriented manifold with boundary, then ∂M is a compact embedded $(n-1)$-manifold (without boundary). Thus if ω is an $(n-1)$-form on M, we can interpret $\int_{\partial M} \omega$ unambiguously as the integral of $\omega|_{\partial M}$ over ∂M, where ∂M is always understood to have the induced orientation.

Proposition 10.20 (Properties of Integrals of Forms). *Suppose M and N are oriented smooth n-manifolds with or without boundaries, and ω, η are compactly supported n-forms on M.*

(a) Linearity: *If $a, b \in \mathbb{R}$, then*

$$\int_M a\omega + b\eta = a \int_M \omega + b \int_M \eta.$$

(b) Orientation Reversal: *If $\overline{M}$ denotes M with the opposite orientation, then*

$$\int_{\overline{M}} \omega = -\int_M \omega.$$

(*c*) Positivity: *If ω is an orientation form for M, then $\int_M \omega > 0$.*

(*d*) Diffeomorphism Invariance: *If $F\colon N \to M$ is an orientation-preserving diffeomorphism, then $\int_M \omega = \int_N F^*\omega$.*

Proof. Parts (a) and (b) are left as an exercise. Suppose that ω is an orientation form for M. This means that for any oriented chart (U, φ), $(\varphi^*)^{-1}\omega$ is a positive function times $dx^1 \wedge \cdots \wedge dx^n$. Thus each term in the sum (10.4) defining $\int_M \omega$ is nonnegative, and at least one term is strictly positive, thus proving (c).

To prove (d), it suffices to assume that ω is compactly supported in a single coordinate chart, because any n-form on M can be written as a finite sum of such forms by means of a partition of unity. Thus suppose (U, φ) is an oriented coordinate chart on M whose domain contains the support of ω. It is easy to check that $(F^{-1}(U), \varphi \circ F)$ is an oriented coordinate chart on N whose domain contains the support of $F^*\omega$, and the result then follows immediately from Corollary 10.17. □

Exercise 10.8. Prove parts (a) and (b) of the preceding proposition.

Although the definition of the integral of a form based on partitions of unity is very convenient for theoretical purposes, it is useless for doing actual computations. It is generally quite difficult to write down a partition of unity explicitly, and even when one can be written down, one would have to be exceptionally lucky to be able to compute the resulting integrals (think of trying to integrate $e^{-1/x}$). For computational purposes, it is much more convenient to "chop up" the manifold into a finite number of pieces whose boundaries are sets of measure zero, and compute the integral on each one separately. One way to do this is described below.

A subset $D \subset M$ is called a *domain of integration* if $\overline{D}$ is compact and ∂D has measure zero (in the sense described in Chapter 6). For example, any regular domain (i.e., compact embedded n-submanifold with boundary) in an n-manifold is a domain of integration.

Proposition 10.21. *Let M be a compact, oriented, smooth n-manifold with or without boundary, and let ω be an n-form on M. Suppose $E_1, \dots, E_k$ are compact domains of integration in M; $D_1, \dots, D_k$ are compact domains of integration in $\mathbb{R}^n$; and for $i = 1, \dots, k$, $F_i\colon D_i \to M$ are smooth maps satisfying*

(*i*) *$F_i(D_i) = E_i$, and $F_i|_{\operatorname{Int} D_i}$ is an orientation-preserving diffeomorphism from $\operatorname{Int} D_i$ onto $\operatorname{Int} E_i$.*

(*ii*) *$M = E_1 \cup \cdots \cup E_k$.*

(*iii*) *For each $i \neq j$, E_i and E_j intersect only on their boundaries.*

Then

$$\int_M \omega = \sum_i \int_{D_i} F_i^* \omega.$$

Proof. As in the preceding proof, it suffices to assume that ω is compactly supported in the domain of a single oriented chart (U, φ). In fact, by starting with a cover of M by sufficiently nice charts, we may assume that ∂U has measure zero, and that φ extends to a diffeomorphism from $\overline{U}$ to a compact domain of integration $K \subset \mathbb{H}^n$.

For each i, let

$$A_i = \overline{U} \cap E_i \subset M.$$

Then A_i is a compact subset of M whose boundary has measure zero, since $\partial A_i \subset \partial U \cup F_i(\partial D_i)$. Define compact subsets $B_i, C_i \subset \mathbb{R}^n$ by

$$\begin{aligned} B_i &= F_i^{-1}(A_i), \\ C_i &= \varphi(A_i). \end{aligned}$$

Since smooth maps take sets of measure zero to sets of measure zero, both B_i and C_i are domains of integration, and $\varphi \circ F_i$ maps B_i to C_i and restricts to a diffeomorphism from $\operatorname{Int} B_i$ to $\operatorname{Int} C_i$. Therefore Proposition 10.16 implies that

$$\int_{C_i} (\varphi^{-1})^* \omega = \int_{B_i} F_i^* \omega.$$

Summing over i, and noting that the interiors of the various sets A_i are disjoint, we obtain

$$\begin{aligned} \int_M \omega &= \int_K (\varphi^{-1})^* \omega \\ &= \sum_i \int_{C_i} (\varphi^{-1})^* \omega \\ &= \sum_i \int_{B_i} F_i^* \omega \\ &= \sum_i \int_{D_i} F_i^* \omega. \end{aligned}$$

□

Example 10.22. Let us use this technique to compute the integral of a 2-form over $\mathbb{S}^2$, oriented by means of the outward-pointing vector field $N = x\,\partial/\partial x + y\,\partial/\partial y + z\,\partial/\partial z$. Let ω be the following 2-form:

$$\omega = x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy.$$

If we let D be the rectangle $[0,\pi]\times[0,2\pi]$ and $F\colon D\to\mathbb{S}^2$ be the spherical coordinate map

$$F(\varphi,\theta)=(\sin\varphi\cos\theta,\sin\varphi\sin\theta,\cos\varphi),$$

then the single map $F\colon D\to\mathbb{S}^2$ satisfies the hypotheses of Proposition 10.21, provided that it is orientation-preserving. Assuming this for the moment, we note that

$$\begin{aligned}
F^*dx&=\cos\varphi\cos\theta\,d\varphi-\sin\varphi\sin\theta\,d\theta,\\
F^*dy&=\cos\varphi\sin\theta\,d\varphi+\sin\varphi\cos\theta\,d\theta,\\
F^*dz&=-\sin\varphi\,d\varphi.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{\mathbb{S}^2}\omega&=\int_D F^*\omega\\
&=\int_D -\sin^3\varphi\cos^2\theta\,d\theta\wedge d\varphi+\sin^3\varphi\sin^2\theta\,d\varphi\wedge d\theta\\
&\qquad+\cos^2\varphi\sin\varphi\cos^2\theta\,d\varphi\wedge d\theta-\cos^2\varphi\sin\varphi\sin^2\theta\,d\theta\wedge d\varphi\\
&=\int_D\sin\varphi\,d\varphi\wedge d\theta\\
&=\int_0^{2\pi}\int_0^{\pi}\sin\varphi\,d\varphi\,d\theta\\
&=4\pi.
\end{aligned}$$

To check that F is orientation-preserving, we need to show that $(F_*\partial/\partial\varphi,F_*\partial/\partial\theta)$ is an oriented basis for $\mathbb{S}^2$ at each point, which means by definition that $(N,F_*\partial/\partial\varphi,F_*\partial/\partial\theta)$ is an oriented basis for $\mathbb{R}^3$. Calculating at an arbitrary point $(x,y,z)=F(\varphi,\theta)$, we find

$$\begin{aligned}
N&=\sin\varphi\cos\theta\frac{\partial}{\partial x}+\sin\varphi\sin\theta\frac{\partial}{\partial y}+\cos\varphi\frac{\partial}{\partial z};\\
F_*\frac{\partial}{\partial\varphi}&=\cos\varphi\cos\theta\frac{\partial}{\partial x}+\cos\varphi\sin\theta\frac{\partial}{\partial y}-\sin\varphi\frac{\partial}{\partial z};\\
F_*\frac{\partial}{\partial\theta}&=-\sin\varphi\sin\theta\frac{\partial}{\partial x}+\sin\varphi\cos\theta\frac{\partial}{\partial y}.
\end{aligned}$$

The transition matrix is therefore

$$\begin{pmatrix}\sin\varphi\cos\theta&\sin\varphi\sin\theta&\cos\varphi\\ \cos\varphi\cos\theta&\cos\varphi\sin\theta&-\sin\varphi\\ -\sin\varphi\sin\theta&\sin\varphi\cos\theta&0\end{pmatrix},$$

which has determinant $\sin\varphi>0$.

Stokes's Theorem

In this section we will state and prove the central result in the theory of integration on manifolds: Stokes's theorem for manifolds. This is a far-reaching generalization of the fundamental theorem of calculus and of the classical theorems of vector calculus.

Theorem 10.23 (Stokes's Theorem). *Let M be an oriented n-dimensional manifold with boundary, and let ω be a compactly supported $(n-1)$-form on M. Then*

$$\int_M d\omega = \int_{\partial M} \omega. \tag{10.5}$$

The statement of this theorem is concise and elegant, but it requires a bit of interpretation. First, as usual, ∂M is understood to have the induced (Stokes) orientation, and ω is understood to be restricted to ∂M on the right-hand side. If $\partial M = \varnothing$, then the right-hand side is to be interpreted as zero. When M is 1-dimensional, the right-hand integral is really just a finite sum.

With these understandings, we proceed with the proof of the theorem. You should check as you read through the proof that it works correctly when $n = 1$.

Proof. We begin by considering a very special case: Suppose M is the upper half space $\mathbb{H}^n$ itself. Then the fact that ω has compact support means that there is a number $R > 0$ such that $\operatorname{supp}\omega$ is contained in the rectangle $A = [-R, R]\times\cdots\times[-R, R]\times[0, R]$. We can write ω in standard coordinates as

$$\omega = \sum_{i=1}^{n} \omega_i \, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n,$$

where the hat means that dx^i is omitted. Therefore,

$$\begin{aligned}
d\omega &= \sum_{i=1}^{n} d\omega_i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\
&= \sum_{i,j=1}^{n} \frac{\partial \omega_i}{\partial x^j} dx^j \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\
&= \sum_{i=1}^{n} (-1)^{i-1} \frac{\partial \omega_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n.
\end{aligned}$$

Thus we compute

$$\begin{aligned}\int_{\mathbb{H}^n} d\omega &= \sum_{i=1}^{n}(-1)^{i-1}\int_A \frac{\partial \omega_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n \\ &= \sum_{i=1}^{n}(-1)^{i-1}\int_0^R \int_{-R}^R \cdots \int_{-R}^R \frac{\partial \omega_i}{\partial x^i}(x) dx^1 \cdots dx^n.\end{aligned}$$

We can rearrange the order of integration in each term so as to do the x^i integration first. By the fundamental theorem of calculus, the terms for which $i \neq n$ reduce to

$$\begin{aligned}&\sum_{i=1}^{n-1}(-1)^{i-1}\int_0^R \int_{-R}^R \cdots \int_{-R}^R \frac{\partial \omega_i}{\partial x^i}(x) dx^1 \cdots dx^n \\ &\quad = \sum_{i=1}^{n-1}(-1)^{i-1}\int_0^R \int_{-R}^R \cdots \int_{-R}^R \frac{\partial \omega_i}{\partial x^i}(x) dx^i \, dx^1 \cdots \widehat{dx^i} \cdots dx^n \\ &\quad = \sum_{i=1}^{n-1}(-1)^{i-1}\int_0^R \int_{-R}^R \cdots \int_{-R}^R \omega_i(x)\Big|_{x^i=-R}^{x^i=R} dx^1 \cdots \widehat{dx^i} \cdots dx^n \\ &\quad = 0,\end{aligned}$$

because we have chosen R large enough that $\omega = 0$ when $x^i = \pm R$. The only term that might not be zero is the one for which $i = n$. For that term we have

$$\begin{aligned}\int_{\mathbb{H}^n} d\omega &= (-1)^{n-1}\int_{-R}^R \cdots \int_{-R}^R \int_0^R \frac{\partial \omega_n}{\partial x^n}(x) dx^n \, dx^1 \cdots dx^{n-1} \\ &= (-1)^{n-1}\int_{-R}^R \cdots \int_{-R}^R \omega_i(x)\Big|_{x^n=0}^{x^n=R} dx^1 \cdots dx^{n-1} \qquad (10.6) \\ &= (-1)^{n}\int_{-R}^R \cdots \int_{-R}^R \omega_i(x^1, \ldots, x^{n-1}, 0)\, dx^1 \cdots dx^{n-1},\end{aligned}$$

because $\omega_n = 0$ when $x^n = R$.

To compare this to the other side of (10.5), we compute as follows:

$$\int_{\partial \mathbb{H}^n} \omega = \sum_i \int_{A \cap \partial \mathbb{H}^n} \omega_i(x^1, \ldots, x^{n-1}, 0)\, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n.$$

Because x^n vanishes on $\partial \mathbb{H}^n$, the restriction of dx^n to the boundary is identically zero. Thus the only term above that is nonzero is the one for which $i = n$, which becomes

$$\int_{\partial \mathbb{H}^n} \omega = \int_{A \cap \partial \mathbb{H}^n} \omega_n(x^1, \ldots, x^{n-1}, 0)\, dx^1 \wedge \cdots \wedge dx^{n-1}.$$

Taking into account the fact that the coordinates $(x^1, \dots, x^{n-1})$ are positively oriented for $\partial\mathbb{H}^n$ when n is even and negatively oriented when n is odd (Example 10.14), this becomes

$$\int_{\partial\mathbb{H}^n} \omega = (-1)^n \int_{-R}^{R} \cdots \int_{-R}^{R} \omega_n(x^1, \dots, x^{n-1}, 0)\, dx^1 \cdots dx^{n-1},$$

which is equal to (10.6).

Next let M be an arbitrary manifold with boundary, but consider an $(n-1)$-form ω that is compactly supported in the domain of a single (generalized) chart (U, φ). Assuming without loss of generality that φ is an oriented chart, the definition yields

$$\int_M d\omega = \int_{\mathbb{H}^n} (\varphi^{-1})^* d\omega = \int_{\mathbb{H}^n} d((\varphi^{-1})^*\omega),$$

since $(\varphi^{-1})^* d\omega$ is compactly supported on $\mathbb{H}^n$. By the computation above, this is equal to

$$\int_{\partial\mathbb{H}^n} (\varphi^{-1})^*\omega, \tag{10.7}$$

where $\partial\mathbb{H}^n$ is given the induced orientation. Since φ_* takes outward-pointing vectors on ∂M to outward-pointing vectors on $\mathbb{H}^n$ (by Lemma 10.10), it follows that $\varphi|_{U\cap\partial M}$ is an orientation-preserving diffeomorphism onto $\varphi(U)\cap\partial\mathbb{H}^n$, and thus (10.7) is equal to $\int_{\partial M}\omega$. This proves the theorem in this case.

Finally, let ω be an arbitrary compactly supported $(n-1)$-form. Choosing a cover of $\operatorname{supp}\omega$ by finitely many oriented (generalized) coordinate charts $\{(U_i, \varphi_i)\}$, and choosing a subordinate partition of unity $\{\psi_i\}$, we can apply the preceding argument to $\psi_i\omega$ for each i and obtain

$$\begin{aligned}
\int_{\partial M} \omega &= \sum_i \int_{\partial M} \psi_i\omega \\
&= \sum_i \int_M d(\psi_i\omega) \\
&= \sum_i \int_M d\psi_i \wedge \omega + \psi_i d\omega \\
&= \int_M d\Bigl(\sum_i \psi_i\Bigr) \wedge \omega + \int_M \Bigl(\sum_i \psi_i\Bigr) d\omega \\
&= 0 + \int_M d\omega,
\end{aligned}$$

because $\sum_i \psi_i \equiv 1$. □

Two special cases of Stokes's theorem are worthy of special note. The proofs are immediate.

Corollary 10.24. *Suppose M is a compact manifold without boundary. Then the integral of every exact form over M is zero:*

$$\int_M d\omega = 0 \quad \text{if } \partial M = \varnothing.$$

Corollary 10.25. *Suppose M is a compact smooth manifold with boundary. If ω is a closed form on M, then the integral of ω over ∂M is zero:*

$$\int_{\partial M} \omega = 0 \quad \text{if } d\omega = 0 \text{ on } M.$$

Example 10.26. Let N be a smooth manifold and suppose $\gamma\colon [a,b] \to N$ is an embedding, so that $M = \gamma[a,b]$ is an embedded 1-submanifold with boundary in N. If we give M the orientation such that γ is orientation-preserving, then for any smooth function $f \in C^\infty(N)$, Stokes's theorem says

$$\int_{[a,b]} \gamma^*\omega = \int_M df = f(\gamma(b)) - f(\gamma(b)).$$

Thus Stokes's theorem reduces to the fundamental theorem for line integrals (Theorem 4.20) in this case. In particular, when $\gamma\colon [a,b] \to \mathbb{R}$ is the inclusion map, then Stokes's theorem is just the ordinary fundamental theorem of calculus.

Another application of Stokes's theorem is to prove the classical result known as Green's theorem.

Theorem 10.27 (Green's Theorem). *Suppose D is a regular domain in $\mathbb{R}^2$, and P, Q are smooth real-valued functions on D. Then*

$$\int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy = \int_{\partial D} P\,dx + Q\,dy.$$

Proof. This is just Stokes's theorem applied to the 1-form $P\,dx + Q\,dy$. ☐

We will see other applications of Stokes's theorem later in this chapter.

Manifolds with Corners

In many applications of Stokes's theorem, it is necessary to deal with geometric objects such as triangles, squares, or cubes that are topological manifolds with boundary, but are not smooth manifolds with boundary because they have "corners." It is easy to generalize Stokes's theorem to this situation, and we do so in this section. We will use this generalization only in our discussion of de Rham cohomology in Chapter 11.

Let $\overline{\mathbb{R}^n_+}$ denote the closed positive "quadrant" of $\mathbb{R}^n$:

$$\overline{\mathbb{R}^n_+} = \{(x^1, \ldots, x^n) \in \mathbb{R}^n : x^1 \geq 0, \ldots, x^n \geq 0\}.$$

This space is the model for the type of corners we will be concerned with.

Exercise 10.9. Prove that $\overline{\mathbb{R}^n_+}$ is homeomorphic to the upper half-space $\mathbb{H}^n$.

Suppose M is a topological n-manifold with boundary. A *chart with corners* for M is a pair (U, φ), where U is an open subset of M, and φ is a homeomorphism from U to a (relatively) open set $\widetilde{U} \subset \overline{\mathbb{R}^n_+}$. Two charts with corners (U, φ), (V, ψ) are said to be smoothly compatible if the composite map $\varphi \circ \psi^{-1} \colon \psi(U \cap V) \to \varphi(U \cap V)$ is smooth. (As usual, this means that it admits a smooth extension to an open set in $\mathbb{R}^n$.)

A *smooth structure with corners* on a topological manifold with boundary is a maximal collection of smoothly compatible charts with corners whose domains cover M. A topological manifold with boundary together with a smooth structure with corners is called a *smooth manifold with corners.*

If M is a smooth manifold with corners, any chart with corners (U, φ) in the given smooth structure with corners is called a *smooth chart with corners* for M.

Example 10.28. Any closed rectangle in $\mathbb{R}^n$ is a smooth n-manifold with corners.

Because of the result of Exercise 10.9, charts with corners are indistinguishable topologically from generalized charts in the sense of topological manifolds with boundary. Thus from the topological view there is no difference between manifolds with boundary and manifolds with corners. The difference is in the smooth structure, because the compatibility condition for charts with corners is different from that for generalized charts.

It is easy to check that the boundary of $\overline{\mathbb{R}^n_+}$ in $\mathbb{R}^n$ is the set of points at which at least one coordinate vanishes. The points in $\overline{\mathbb{R}^n_+}$ at which more than one coordinate vanishes are called its *corner points.*

Lemma 10.29. *Let M be a smooth n-manifold with corners, and let $p \in M$. If $\varphi(p)$ is a corner point for some smooth chart with corners (U, φ), then the same is true for every such chart whose domain contains p.*

Proof. Suppose (U, φ) and (V, ψ) are two charts with corners such that $\varphi(p)$ is a corner point but $\psi(p)$ is not. To simplify notation, let us assume without loss of generality that $\varphi(p)$ has coordinates $(x^1, \ldots, x^k, 0, \ldots, 0)$ with $k \leq n-2$. Then $\psi(V)$ contains an open subset of some $(n-1)$-dimensional linear subspace $S \subset \mathbb{R}^n$, with $\psi(p) \in S$. (If $\psi(p)$ is a boundary point, S can be taken to be the unique subspace of the form $x^i = 0$ that contains $\psi(p)$. If $\psi(p)$ is an interior point, any $(n-1)$-dimensional subspace containing $\psi(p)$ will do.)

Let $\alpha\colon S\cap\psi(V) \to \mathbb{R}^n$ be the restriction of $\varphi\circ\psi^{-1}$ to $S\cap\psi(V)$. Because $\varphi\circ\psi^{-1}$ is a diffeomorphism, α is a smooth immersion. Let $T = \alpha_* S \subset \mathbb{R}^n$. Because T is $(n-1)$-dimensional, it must contain a vector X such that one of the last two components X^{n-1} or X^n is nonzero (otherwise T would be contained in a codimension-2 subspace). Renumbering the coordinates and replacing X by $-X$ if necessary, we may assume that $X^n < 0$.

Now let $\gamma\colon (-\varepsilon,\varepsilon) \to S$ be a smooth curve such that $\gamma(0) = p$ and $\alpha_*\gamma'(0) = X$. Then $\alpha(\gamma(t))$ has negative x^n coordinate for small $t > 0$, which contradicts the fact that α takes its values in $\overline{\mathbb{R}^n_+}$. □

If M is a smooth manifold with corners, a point $p \in M$ is called a corner point if $\varphi(p)$ is a corner point in $\overline{\mathbb{R}}^n_+$ with respect to some (and hence every) smooth chart with corners (U,φ). It is clear that every smooth manifold with or without boundary is also a smooth manifold with corners (but with no corner points). Conversely, a smooth manifold with corners is a smooth manifold with boundary if and only if it has no corner points. The boundary of a smooth manifold with corners, however, is in general not a smooth manifold with corners (think of the boundary of a cube, for example). In fact, even the boundary of $\overline{\mathbb{R}^n_+}$ itself is not a smooth manifold with corners. It is, however, a union of finitely many such: $\partial\overline{\mathbb{R}^n_+} = H_1 \cup \cdots \cup H_n$, where

$$H_i = \{(x^1,\dots,x^n) \in \overline{\mathbb{R}^n_+} : x^i = 0\}$$

is an $(n-1)$-dimensional smooth manifold with corners contained in the subspace defined by $x^i = 0$.

The usual flora and fauna of smooth manifolds—smooth maps, partitions of unity, tangent vectors, covectors, tensors, differential forms, orientations, and integrals of differential forms—can be defined on smooth manifolds with corners in exactly the same way as we have done for smooth manifolds and smooth manifolds with boundary, using smooth charts with corners in place of smooth charts or generalized charts. The details are left to the reader.

In addition, for Stokes's theorem we will need to integrate a differential form over the boundary of a smooth manifold with corners. Since the boundary is not itself a smooth manifold with corners, this requires a special definition. Let M be an oriented smooth n-manifold with corners, and suppose ω is an $(n-1)$-form on ∂M that is compactly supported in the domain of a single oriented smooth chart with corners (U,φ). We define the integral of ω over ∂M by

$$\int_{\partial M} \omega = \sum_{i=1}^{n} \int_{H_i} (\varphi^{-1})^*\omega,$$

where each H_i is given the induced orientation as part of the boundary of the set where $x^i \geq 0$. In other words, we simply integrate ω in coordinates over the codimension-1 portion of the boundary. Finally, if ω is an arbitrary compactly supported $(n-1)$-form on M, we define the integral of ω over ∂M by piecing together with a partition of unity just as in the case of a manifold with boundary.

In practice, of course, one does not evaluate such integrals by using partitions of unity. Instead, one "chops up" the boundary into pieces that can be parametrized by compact Euclidean domains of integration, just as for ordinary manifolds with or without boundary. If M is a smooth manifold with corners, we say a subset $A \subset \partial M$ has measure zero in ∂M if for every smooth chart with corners (U, φ), each set $\varphi(A) \cap H_i$ has measure zero in H_i for $i = 1, \ldots, n$. A domain of integration in ∂M is a subset $E \subset \partial M$ whose boundary has measure zero in ∂M. The following proposition is an analogue of Proposition 10.21.

Proposition 10.30. *The statement of Proposition 10.21 is true if M is replaced by the boundary of a compact, oriented, smooth n-manifold with corners.*

Exercise 10.10. Show how the proof of Proposition 10.21 needs to be adapted to prove Proposition 10.30.

Example 10.31. Let $I \times I = [0,1] \times [0,1]$ be the unit square in $\mathbb{R}^2$, and suppose ω is a smooth 1-form on $\partial(I \times I)$. Then it is not hard to check that the maps $F_i\colon I \to I \times I$ given by

$$\begin{aligned} F_1(t) &= (t, 0), \\ F_2(t) &= (1, t), \\ F_3(t) &= (1-t, 1), \\ F_4(t) &= (0, 1-t) \end{aligned} \tag{10.8}$$

satisfy the hypotheses of Proposition 10.30. (These four curve segments in sequence traverse the boundary of $I \times I$ in the counterclockwise direction.) Therefore,

$$\int_{\partial(I \times I)} \omega = \int_{F_1} \omega + \int_{F_2} \omega + \int_{F_3} \omega + \int_{F_4} \omega. \tag{10.9}$$

Exercise 10.11. Verify the claims of the preceding example.

The next theorem is the main result of this section.

Theorem 10.32 (Stokes's Theorem on Manifolds with Corners). *Let M be a smooth n-manifold with corners, and let ω be a compactly supported $(n-1)$-form on M. Then*

$$\int_M d\omega = \int_{\partial M} \omega.$$

Proof. The proof is nearly identical to the proof of Stokes's theorem proper, so we will just indicate where changes need to be made. By means of smooth charts with corners and a partition of unity just as in that proof, we may reduce the theorem to the case in which $M = \overline{\mathbb{R}^n_+}$. In that case, calculating exactly as in the proof of Theorem 10.23, we obtain

$$\begin{aligned}
\int_{\overline{\mathbb{R}^n_+}} d\omega &= \sum_{i=1}^{n} (-1)^{i-1} \int_0^R \cdots \int_0^R \frac{\partial \omega_i}{\partial x^i}(x) dx^1 \cdots dx^n \\
&= \sum_{i=1}^{n} (-1)^{i-1} \int_0^R \cdots \int_0^R \frac{\partial \omega_i}{\partial x^i}(x) dx^i \, dx^1 \cdots \widehat{dx^i} \cdots dx^n \\
&= \sum_{i=1}^{n} (-1)^{i-1} \int_0^R \cdots \int_0^R \omega_i(x) \Big|_{x^i=0}^{x^i=R} dx^1 \cdots \widehat{dx^i} \cdots dx^n \\
&= \sum_{i=1}^{n} (-1)^{i} \int_0^R \cdots \int_0^R \omega_n(x^1, \dots, 0, \dots, x^n) \, dx^1 \cdots \widehat{dx^i} \cdots dx^n \\
&= \sum_{i=1}^{n} \int_{H_i} \omega \\
&= \int_{\partial \overline{\mathbb{R}^n_+}} \omega.
\end{aligned}$$

(The factor $(-1)^i$ disappeared because the induced orientation on H_i is $(-1)^i$ times that of the standard coordinates $(x^1, \dots, \widehat{x^i}, \dots, x^n)$.) This completes the proof. □

Here is an immediate application of this result, which we will use when we study de Rham cohomology in the next chapter. Recall that two curve segments $\gamma_0, \gamma_1 \colon [a,b] \to M$ are said to be *path homotopic* if they are homotopic relative to their endpoints, that is, homotopic via a homotopy $H \colon [a,b] \times I \to M$ such that $H(a,t) = \gamma_0(a) = \gamma_1(a)$ and $H(b,t) = \gamma_0(b) = \gamma_1(b)$ for all $t \in I$.

Theorem 10.33. *Suppose M is a smooth manifold, and $\gamma_0, \gamma_1 \colon [a,b] \to M$ are path homotopic piecewise smooth curve segments. For every closed 1-form ω on M,*

$$\int_{\gamma_0} \omega = \int_{\gamma_1} \omega.$$

Proof. By means of an affine reparametrization, we may as well assume for simplicity that $[a,b] = [0,1]$. Assume first that γ_0 and γ_1 are smooth. By Proposition 6.20, γ_0 and γ_1 are smoothly homotopic relative to $\{0,1\}$. Let $H \colon I \times I \to M$ be such a smooth homotopy. Since ω is closed, we have

$$\int_{I \times I} d(H^*\omega) = \int_{I \times I} H^* d\omega = 0.$$

On the other hand, $I \times I$ is a manifold with corners, so Stokes's theorem implies

$$0 = \int_{I \times I} d(H^*\omega) = \int_{\partial(I \times I)} H^*\omega.$$

By Example 10.31 together with the diffeomorphism invariance of line integrals (Exercise 4.9), therefore,

$$\begin{aligned} 0 &= \int_{\partial(I \times I)} H^*\omega \\ &= \int_{F_1} H^*\omega + \int_{F_2} H^*\omega + \int_{F_3} H^*\omega + \int_{F_4} H^*\omega \\ &= \int_{H \circ F_1} \omega + \int_{H \circ F_2} \omega + \int_{H \circ F_3} \omega + \int_{H \circ F_4} \omega, \end{aligned}$$

where F_1, F_2, F_3, F_4 are defined by (10.8). The fact that H is relative to $\{0, 1\}$ means that $H \circ F_2$ and $H \circ F_4$ are constant maps, and therefore the second and fourth terms above are zero. The theorem then follows from the facts that $H \circ F_1 = \gamma_0$ and $H \circ F_3$ is a backward reparametrization of γ_1.

Next we consider the general case of piecewise smooth curves. We cannot simply apply the preceding result on each subinterval where γ_0 and γ_1 are smooth, because the restricted curves may not start and end at the same points. Instead, we will prove the following more general claim: *Let $\gamma_0, \gamma_1 \colon I \to M$ be piecewise smooth curve segments (not necessarily with the same endpoints), and suppose $H \colon I \times I \to M$ is any homotopy between them. Define curve segments $\sigma_0, \sigma_1 \colon I \to M$ by*

$$\begin{aligned} \sigma_0(t) &= H(0, t), \\ \sigma_1(t) &= H(1, t), \end{aligned}$$

and let $\widetilde{\sigma}_0, \widetilde{\sigma}_1$ be any smooth curve segments that are path homotopic to σ_0, σ_1 respectively. Then

$$\int_{\gamma_1} \omega - \int_{\gamma_0} \omega = \int_{\sigma_1} \omega - \int_{\sigma_0} \omega. \tag{10.10}$$

When specialized to the case in which γ_0 and γ_1 are path homotopic, this implies the theorem, because σ_0 and σ_1 are constant maps in that case.

Since γ_0 and γ_1 are piecewise smooth, there are only finitely many points $(a_1, \dots, a_m)$ in $(0, 1)$ at which either γ_0 or γ_1 is not smooth. We will prove the claim by induction on the number m of such points. When $m = 0$, both curves are smooth, and by Proposition 6.20 we may replace the given homotopy H by a smooth homotopy $\widetilde{H}$. Recall from the proof of Proposition 6.20 that the smooth homotopy $\widetilde{H}$ can actually be taken to be homotopic to H relative to $I\times\{0\}\cup I\times\{1\}$. Thus for $i=0,1$, the curve $\widetilde{\sigma}_i(t)=\widetilde{H}(i,t)$ is a smooth curve segment that is path homotopic to σ_i. In this setting, (10.10) just reduces to (10.9). Note that the integrals over $\widetilde{\sigma}_0$ and $\widetilde{\sigma}_1$ do not depend on which smooth curves path homotopic to σ_0 and σ_1 are chosen, by the smooth case of the theorem proved above.

Now let γ_0,γ_1 be homotopic piecewise smooth curves with m nonsmooth points $(a_1,\dots,a_m)$, and suppose the claim is true for curves with fewer than m such points. For $i=0,1$, let γ_i' be the restriction of γ_i to $[0,a_m]$, and let γ_i'' be its restriction to $[a_m,1]$. Let $\sigma\colon I\to M$ be the curve segment

$$\sigma(t)=H(a_m,t),$$

and let $\widetilde{\sigma}$ by any smooth curve segment that is path homotopic to σ. Then, since γ_i' and γ_i'' have fewer than m nonsmooth points, the inductive hypothesis implies

$$\begin{aligned}\int_{\gamma_1}\omega-\int_{\gamma_0}\omega&=\left(\int_{\gamma_1'}\omega-\int_{\gamma_0'}\omega\right)+\left(\int_{\gamma_1''}\omega-\int_{\gamma_0''}\omega\right)\\&=\left(\int_{\widetilde{\sigma}}\omega-\int_{\widetilde{\sigma}_0}\omega\right)+\left(\int_{\widetilde{\sigma}_1}\omega-\int_{\widetilde{\sigma}}\omega\right)\\&=\int_{\widetilde{\sigma}_1}\omega-\int_{\widetilde{\sigma}_0}\omega.\end{aligned}$$

This completes the proof. □

Integration on Riemannian Manifolds

In this section, we explore how the theory of orientations and integration can be specialized to Riemannian manifolds. Thinking of $\mathbb{R}^2$ and $\mathbb{R}^3$ as Riemannian manifolds, this will eventually lead us to the classical theorems of vector calculus as consequences of Stokes's theorem.

The Riemannian Volume Form

We begin with orientations. Let (M,g) be an oriented Riemannian manifold. We know from Proposition 8.16 that there is a smooth orthonormal frame $(E_1,\dots,E_n)$ in a neighborhood of each point of M. By replacing E_1 by $-E_1$ if necessary, we can find an *oriented* orthonormal frame in a neighborhood of each point.

Proposition 10.34. *Suppose (M,g) is an oriented Riemannian n-manifold. There is a unique orientation form $\Omega\in\mathcal{A}^n(M)$ such that*

$$\Omega_p(E_1,\dots,E_n)=1 \tag{10.11}$$

for every $p\in M$ and every oriented orthonormal basis (E_i) for T_pM.

Remark. The n-form whose existence and uniqueness are guaranteed by this proposition is called the *Riemannian volume form*, or sometimes the *Riemannian volume element*. Because of the role it plays in integration on Riemannian manifolds, as we will see shortly, it is often denoted by dV_g (or dA_g or ds_g in the 2-dimensional or 1-dimensional case, respectively). Be warned, however, that this notation is *not* meant to imply that the volume form is the exterior derivative of an $(n-1)$-form; in fact, as we will see in Chapter 11, this is never the case on a compact manifold. You should just interpret dV_g as a notational convenience.

Proof. Suppose first that such a form Ω exists. If $(E_1, \dots, E_n)$ is any local oriented orthonormal frame and $(\varepsilon^1, \dots, \varepsilon^n)$ is the dual coframe, we can write $\Omega = f\,\varepsilon^1 \wedge \cdots \wedge \varepsilon^n$ locally. The condition (10.11) then reduces to $f = 1$, so

$$\Omega = \varepsilon^1 \wedge \cdots \wedge \varepsilon^n. \tag{10.12}$$

This proves that such a form is uniquely determined.

To prove existence, we would like to *define* Ω in a neighborhood of each point by (10.12). If $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ is another oriented orthonormal frame, with dual coframe $(\widetilde{\varepsilon}^1, \dots, \widetilde{\varepsilon}^n)$, let

$$\widetilde{\Omega} = \widetilde{\varepsilon}^1 \wedge \cdots \wedge \widetilde{\varepsilon}^n.$$

We can write

$$\widetilde{E}_i = A_i^j E_j$$

for some matrix (A_i^j) of smooth functions. The fact that both frames are orthonormal means that $(A_i^j(p)) \in \mathrm{O}(n)$ for each p, so $\det(A_i^j) = \pm 1$, and the fact that the two frames are consistently oriented forces the positive sign. We compute

$$\begin{aligned}
\Omega(\widetilde{E}_1, \dots, \widetilde{E}_n) &= \det(\varepsilon^j(\widetilde{E}_i)) \\
&= \det(A_i^j) \\
&= 1 \\
&= \widetilde{\Omega}(\widetilde{E}_1, \dots, \widetilde{E}_n).
\end{aligned}$$

Thus $\Omega = \widetilde{\Omega}$, so defining Ω in a neighborhood of each point by (10.12) with respect to some oriented orthonormal frame yields a global n-form. The resulting form is clearly smooth and satisfies (10.11) for every oriented orthonormal basis. □

Although the expression for the Riemannian volume form with respect to an oriented orthonormal frame is particularly simple, it is also useful to have an expression for it in coordinates.

Lemma 10.35. *Let (M, g) be an oriented Riemannian manifold. If (x^i) is any oriented coordinate chart, then the Riemannian volume form has the local coordinate expression*

$$dV_g = \sqrt{\det(g_{ij})}\, dx^1 \wedge \cdots \wedge dx^n,$$

where g_{ij} are the components of g in these coordinates.

Proof. Let (x^i) be oriented coordinates near $p \in M$. Then locally $\Omega = f\, dx^1 \wedge \cdots \wedge dx^n$ for some positive coefficient function f. To compute f, let (E_i) be any oriented orthonormal frame defined on a neighborhood of p, and let (ε^i) be the dual coframe. If we write the coordinate frame in terms of the orthonormal frame as

$$\frac{\partial}{\partial x^i} = A_i^j E_j,$$

then we can compute

$$\begin{aligned} f &= \Omega\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) \\ &= \varepsilon^1 \wedge \cdots \wedge \varepsilon^n \left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) \\ &= \det\left(\varepsilon^j\left(\frac{\partial}{\partial x^i}\right)\right) \\ &= \det(A_i^j). \end{aligned}$$

On the other hand, observe that

$$\begin{aligned} g_{ij} &= \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle \\ &= \langle A_i^k E_k, A_j^l E_l \rangle \\ &= A_i^k A_j^l \langle E_k, E_l \rangle \\ &= \sum_k A_i^k A_j^k. \end{aligned}$$

This last expression is the (i, j)-entry of the matrix product $A^T A$, where $A = (A_i^j)$. Thus

$$\det(g_{ij}) = \det(A^T A) = \det A^T \det A = (\det A)^2,$$

from which it follows that $f = \det A = \pm\sqrt{\det(g_{ij})}$. Since both frames $(\partial/\partial x^i)$ and (E_j) are oriented, the sign must be positive. □

We noted earlier that real-valued functions cannot be integrated in a coordinate-independent way on an arbitrary manifold. However, with the additional structures of a Riemannian metric and an orientation, we can recover the notion of the integral of a function.

Suppose (M, g) is an oriented Riemannian manifold (with or without boundary), and let dV_g denote its Riemannian volume form. If f is a compactly supported smooth function on M, then $f\,dV_g$ is an n-form, so we can define the *integral of f over M* to be $\int_M f\,dV_g$. (This, of course, is the reason we chose the notation dV_g for the Riemannian volume form.) If M itself is compact, we define the *volume* of M by

$$\operatorname{Vol}(M) = \int_M dV_g.$$

Lemma 10.36. *Let (M, g) be an oriented Riemannian manifold. If f is a compactly supported smooth function on M and $f \geq 0$, then $\int_M f\,dV_g \geq 0$, with equality if and only if $f \equiv 0$.*

Proof. Clearly $\int_M f\,dV_g = 0$ if f is identically zero. If $f \geq 0$ and f is positive somewhere, then $f\,dV_g$ is an orientation form on the open subset $U \subset M$ where $f > 0$, so the result follows from Proposition 10.20(c). □

Hypersurfaces in Riemannian Manifolds

Let (M, g) be an oriented Riemannian manifold, and suppose $S \subset M$ is a submanifold. A vector field N along S is said to be *normal* to S if $N_p \perp T_pS$ for each $p \in S$. If S is a hypersurface, then any unit normal vector field along S is clearly transverse to S, so it determines an orientation of S by Proposition 10.8. The next proposition gives a very simple formula for the volume form of the induced metric on S with respect to this orientation.

Proposition 10.37. *Let (M, g) be an oriented Riemannian manifold, let $S \subset M$ be an immersed hypersurface, and let $\widetilde{g}$ denote the induced metric on S. Suppose N is a smooth unit normal vector field along S. With respect to the orientation of S determined by N, the volume form of $(S, \widetilde{g})$ is given by*

$$dV_{\widetilde{g}} = (N \lrcorner\, dV_g)|_S.$$

Proof. By Proposition 10.8, the $(n-1)$-form $N \lrcorner dV_g$ is an orientation form for S. To prove that it is the volume form for the induced Riemannian metric, we need only show that it gives the value 1 whenever it is applied to an oriented orthonormal frame for S. Thus let $(E_1, \dots, E_{n-1})$ be such a frame. At each point $p \in S$, the basis $(N_p, E_1|_p, \dots, E_{n-1}|_p)$ is orthonormal, and is oriented for T_pM (this is the definition of the orientation determined by N). Thus

$$(N \lrcorner\, dV_g)|_S(E_1, \dots, E_{n-1}) = dV_g(N, E_1, \dots, E_{n-1}) = 1,$$

which proves the result. □

The following lemma will be useful in our proofs of the classical theorems of vector analysis below.

Lemma 10.38. *With notation as in Proposition 10.37, if X is any vector field along S, we have*

$$X \lrcorner \, dV_g|_S = \langle X, N \rangle \, dV_{\widetilde{g}}. \tag{10.13}$$

Proof. Define two vector fields $X^\top$ and $X^\perp$ along S by

$$\begin{aligned} X^\perp &= \langle X, N \rangle N, \\ X^\top &= X - X^\perp. \end{aligned}$$

Then $X = X^\perp + X^\top$, where $X^\perp$ is normal to S and $X^\top$ is tangent to it. Using this decomposition,

$$X \lrcorner \, dV_g = X^\perp \lrcorner \, dV_g + X^\top \lrcorner \, dV_g.$$

Using Corollary 10.40, the first term simplifies to

$$(X^\perp \lrcorner \, dV_g)|_{\widetilde{M}} = \langle X, N \rangle (N \lrcorner \, dV_g)|_{\widetilde{M}} = \langle X, N \rangle dV_{\widetilde{g}}.$$

Thus (10.13) will be proved if we can show that $(X^\top \lrcorner \, dV_g)|_{\widetilde{M}} = 0$. If $X_1, \dots, X_{n-1}$ are any vectors tangent to $\widetilde{M}$, then

$$(X^\top \lrcorner \, dV_g)(X_1, \dots, X_{n-1}) = dV_g(X^\top, X_1, \dots, X_{n-1}) = 0,$$

because any n vectors in an $(n-1)$-dimensional vector space are linearly dependent. □

The result of Proposition 10.37 takes on particular importance in the case of a Riemannian manifold with boundary, because of the following proposition.

Proposition 10.39. *Suppose M is any Riemannian manifold with boundary. There is a unique smooth outward-pointing unit normal vector field N along ∂M.*

Proof. First we prove uniqueness. At any point $p \in \partial M$, the vector space $(T_p\partial M)^\perp \subset T_qM$ is 1-dimensional, so there are exactly two unit vectors at p that are normal to ∂M. Since any unit normal vector N is obviously transverse to ∂M, it must have nonzero x^n-component. Thus exactly one of the two choices of unit normal has negative x^n-component, which is equivalent to being outward-pointing.

To prove existence, we will show that there exists a smooth outward unit normal field in a neighborhood of each point. By the uniqueness result above, these vector fields all agree where they overlap, so the resulting vector field is globally defined.

Let $p \in \partial M$. By Proposition 8.24, there exists an adapted orthonormal frame for M in a neighborhood U of p: This is a local frame $(E_1, \dots, E_n)$ such that $(E_1, \dots, E_{n-1})$ restricts to an orthonormal frame for ∂M. This implies that E_n is a (smooth) unit normal vector field along ∂M in a neighborhood of p. It is obviously transverse to ∂M. If we assume (by shrinking U if necessary) that U is connected, then E_n must be either inward-pointing or outward-pointing on all of $\partial M \cap U$. Replacing E_n by $-E_n$ if necessary, we obtain a smooth outward-pointing unit normal vector field defined near p. This completes the proof. □

The next corollary is immediate.

Corollary 10.40. *If (M, g) is an oriented Riemannian manifold with boundary and $\widetilde{g}$ is the induced Riemannian metric on ∂M, then the volume form of $\widetilde{g}$ is*

$$dV_{\widetilde{g}} = (N \lrcorner\, dV_g)|_{\partial M},$$

where N is the unit outward normal vector field along ∂M.

Next we will show how Stokes's theorem reduces to familiar results in various special cases.

The Divergence Theorem

Let (M, g) be an oriented Riemannian manifold. Multiplication by the Riemannian volume form defines a map $*\colon C^\infty(M) \to \mathcal{A}^n(M)$:

$$*f = f\, dV_g.$$

If (E_i) is any local oriented orthonormal frame, then f can be recovered from $*f$ locally by

$$f = (*f)(E_1, \dots, E_n).$$

Thus $*$ is an isomorphism.

Define the *divergence operator* $\operatorname{div}\colon \mathcal{T}(M) \to C^\infty(M)$ by

$$\operatorname{div} X = *^{-1} d(X \lrcorner\, dV_g),$$

or equivalently,

$$d(X \lrcorner\, dV_g) = (\operatorname{div} X) dV_g.$$

In the special case of a domain with smooth boundary in $\mathbb{R}^3$, the following theorem is due to Gauss and is often referred to as *Gauss's theorem.*

Theorem 10.41 (The Divergence Theorem). *Let M be a compact, oriented Riemannian manifold with boundary. For any smooth vector field X on M,*

$$\int_M (\operatorname{div} X)\, dV_g = \int_{\partial M} \langle X, N\rangle\, dV_{\widetilde{g}},$$

where N is the outward-pointing unit normal vector field along ∂M and $\widetilde{g}$ is the induced Riemannian metric on ∂M.

Proof. By Stokes's theorem,

$$\begin{aligned}\int_M (\operatorname{div} X)\, dV_g &= \int_M d(X \lrcorner\, dV_g)\\ &= \int_{\partial M} X \lrcorner\, dV_g.\end{aligned}$$

The theorem then follows from Lemma 10.38. □

Surface Integrals

The original theorem that bears the name of Stokes concerned "surface integrals" of vector fields over surfaces in $\mathbb{R}^3$. Using the differential-forms version of Stokes's theorem, this can be generalized to surfaces in Riemannian 3-manifolds. (For reasons that will be explained later, the restriction to dimension 3 cannot be removed.)

Let (M, g) be an oriented Riemannian 3-manifold, and let $\beta\colon TM \to \Lambda^2 M$ denote the bundle map defined by $\beta(X) = X \lrcorner\, dV_g$. It is easily seen to be a bundle isomorphism by checking its values on the elements of any orthonormal frame.

Define an operator $\operatorname{curl}\colon \mathcal{T}(M) \to \mathcal{T}(M)$ by

$$\operatorname{curl} X = \beta^{-1} d(X^\flat),$$

or equivalently,

$$\operatorname{curl} X \lrcorner\, dV_g = d(X^\flat). \tag{10.14}$$

The following commutative diagram summarizes the relationships among the gradient, divergence, curl, and exterior derivative operators:

$$\begin{tikzcd} C^\infty(M) \arrow[r, "\operatorname{grad}"] \arrow[d, "\mathrm{Id}"] & \mathcal{T}(M) \arrow[r, "\operatorname{curl}"] \arrow[d, "\flat"] & \mathcal{T}(M) \arrow[r, "\operatorname{div}"] \arrow[d, "\beta"] & C^\infty(M) \arrow[d, "*"] \\ C^\infty(M) \arrow[r, "d"] & \mathcal{A}^1(M) \arrow[r, "d"] & \mathcal{A}^2(M) \arrow[r, "d"] & \mathcal{A}^3(M), \end{tikzcd} \tag{10.15}$$

Problem 10-22 shows that the composition of any two horizontal arrows in this diagram is zero.

Now suppose $S \subset M$ is a compact, embedded, 2-dimensional submanifold with or without boundary in M, and N is a smooth unit normal vector field along S. Let dA denote the induced Riemannian volume form on S with respect to the induced metric $g|_S$ and the orientation determined by N, so that $dA = (N \lrcorner dV_g)|_S$ by Proposition 10.37. For any smooth vector field X defined on M, the *surface integral* of X over S (with respect to the given choice of normal field) is defined as

$$\int_S \langle X, N \rangle \, dA.$$

The next result, in the special case in which $M = \mathbb{R}^3$, is the original theorem proved by Stokes.

Theorem 10.42 (Stokes's Theorem for Surface Integrals). *Suppose S is a compact, embedded, 2-dimensional submanifold with boundary in a Riemannian 3-manifold M, and suppose N is a smooth unit normal vector field along S. For any smooth vector field X on M,*

$$\int_S \langle \operatorname{curl} X, N \rangle \, dA = \int_{\partial S} \langle X, T \rangle \, ds,$$

where ds is the Riemannian volume form and T is the unique positively oriented unit tangent vector field on ∂S (with respect to the metric and orientation induced from S).

Proof. The general version of Stokes's theorem applied to the 1-form $X^\flat$ yields

$$\int_S d(X^\flat) = \int_{\partial S} X^\flat.$$

Thus the theorem will follow from the following two equations:

$$d(X^\flat)|_S = \langle \operatorname{curl} X, N \rangle \, dA, \tag{10.16}$$

$$X^\flat|_{\partial S} = \langle X, T \rangle \, ds. \tag{10.17}$$

Equation (10.16) is just the defining equation (10.14) for the curl combined with the result of Lemma 10.38. To prove (10.17), we note that $X^\flat|_{\partial S}$ is a 1-form on a 1-manifold, and thus must be equal to $f\,ds$ for some smooth function f on ∂S. To evaluate f, we note that $ds(T) = 1$, and so the definition of $X^\flat$ yields

$$f = f\,ds(T) = X^\flat(T) = \langle X, T \rangle.$$

This proves (10.17) and thus the theorem. □

The curl operator is defined only in dimension 3 because it is only in that case that $\Lambda^2 M$ is isomorphic to TM (via the map $\beta\colon X \mapsto X \lrcorner dV_g$). In fact, it was largely the desire to generalize the curl and the classical version of Stokes's theorem to higher dimensions that led to the entire theory of differential forms.

Problems

10-1. Prove that every smooth 1-manifold is orientable.

10-2. Suppose M is a smooth manifold that is the union of two open subsets, each of which is diffeomorphic to an open subset of $\mathbb{R}^n$, and whose intersection is connected. Show that M is orientable. Use this to give another proof that $\mathbb{S}^n$ is orientable.

10-3. Suppose $\pi\colon \widetilde{M} \to M$ is a smooth covering map and M is orientable. Show that $\widetilde{M}$ is also orientable.

10-4. Suppose M is a connected, oriented smooth manifold and Γ is a discrete group acting freely and properly on M. We say the action is orientation-preserving if for each $\gamma \in \Gamma$, the diffeomorphism $x \mapsto \gamma \cdot x$ is orientation-preserving. Show that M/Γ is orientable if and only if Γ is orientation-preserving.

10-5. Let E be the total space of the Möbius bundle, which is the quotient of $\mathbb{R}^2$ by the $\mathbb{Z}$-action $n \cdot (x, y) = (x + n, (-1)^n y)$ (see Problem 7-22). The *Möbius band* is the subset $M \subset E$ that is the image under the quotient map of the set $\{(x, y) \in \mathbb{R}^2 : |y| \le 1\}$. Show that neither E nor M is orientable.

10-6. Suppose M is a connected nonorientable smooth manifold. Define a set $\widetilde{M}$ by

$$\widetilde{M} = \coprod_{p \in M} \{\text{orientations of } T_pM\},$$

and let $\pi\colon \widetilde{M} \to M$ be the obvious map that sends an orientation of T_pM to p. Show that $\widetilde{M}$ has a unique smooth manifold structure such that π is a local diffeomorphism. With this structure, show that $\widetilde{M}$ is connected and orientable, and π is a 2-sheeted smooth covering map. [The covering $\pi\colon \widetilde{M} \to M$ is called the *orientation covering* of M.]

10-7. Let $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{R}^4$ denote the 2-torus, defined by $w^2 + x^2 = y^2 + z^2 = 1$, with the orientation determined by its product structure (see Exercise 10.4). Compute $\int_{\mathbb{T}^2} \omega$, where ω is the following 2-form on $\mathbb{R}^4$:

$$\omega = wy\, dx \wedge dz.$$

10-8. For each of the following 2-forms ω on $\mathbb{R}^3$, compute $\int_{\mathbb{S}^2} \Omega$, where $\mathbb{S}^2$ is oriented by its outward unit normal.

(a) $\omega = x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy$.

(b) $\omega = xz\,dy \wedge dz + yz\,dz \wedge dx + x^2\,dx \wedge dy$.

10-9. Let D denote the surface of revolution in $\mathbb{R}^3$ obtained by revolving the circle $(x-2)^2+z^2 = 1$ around the z-axis, with its induced Riemannian metric, and with the orientation induced by the outward unit normal.

(a) Compute the surface area of D.

(b) Compute the integral over D of the function $f(x,y,z) = z+1$.

(c) Compute the integral over D of the 2-form ω of Problem 10-8(b).

10-10. Let ω be the $(n-1)$-form on $\mathbb{R}^n \smallsetminus \{0\}$ defined by

$$\omega = |x|^{-n} \sum_{i=1}^{n} (-1)^{i-1} x^i\,dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n. \tag{10.18}$$

(a) Show that $\omega|_{\mathbb{S}^{n-1}}$ is the Riemannian volume element of $\mathbb{S}^{n-1}$ with respect to the round metric.

(b) Show that ω is closed but not exact.

10-11. Suppose M is an oriented Riemannian manifold, and $S \subset M$ is an oriented hypersurface (with or without boundary). Show that there is a unique smooth unit normal vector field along S that determines the given orientation of S.

10-12. Suppose S is an oriented embedded 2-manifold with boundary in $\mathbb{R}^3$, and let $C = \partial S$ with the induced orientation. By Problem 10-11, there is a unique smooth unit normal vector field N on S that determines the orientation. Let T be the oriented unit tangent vector field on C; let V be the unique vector field tangent to S along C that is outward-pointing; and let W be the restriction of N to C. Show that (T_p, V_p, W_p) is an oriented basis for $\mathbb{R}^3$ at each $p \in C$.

10-13. Let (M,g) be an oriented Riemannian n-manifold. With respect to any local coordinates (x^i), show that

$$\operatorname{div}\left(X^i \frac{\partial}{\partial x^i}\right) = \frac{1}{\sqrt{\det(g_{ij})}} \frac{\partial}{\partial x^i}\left(X^i \sqrt{\det(g_{ij})}\right),$$

where g_{ij} are the components of g. Conclude that on $\mathbb{R}^n$ with the Euclidean metric,

$$\operatorname{div}\left(X^i \frac{\partial}{\partial x^i}\right) = \sum_{i=1}^{n} \frac{\partial X^i}{\partial x^i}.$$

10-14. Show that the divergence operator on an oriented Riemannian manifold does not depend on the choice of orientation, and conclude that it is invariantly defined on all Riemannian manifolds.

10-15. Let (M, g) be a compact, oriented Riemannian manifold with boundary, let $\widetilde{g}$ denote the induced Riemannian metric on ∂M, and let N be the outward unit normal vector field along ∂M.

(a) Show that the divergence operator satisfies the following product rule for $f \in C^\infty(M)$, $X \in \mathfrak{T}(M)$:

$$\operatorname{div}(fX) = f \operatorname{div} X + \langle \operatorname{grad} f, X \rangle.$$

(b) Prove the following "integration by parts" formula:

$$\int_M \langle \operatorname{grad} f, X \rangle \, dV_g = \int_{\partial M} f \langle X, N \rangle \, dV_{\widetilde{g}} - \int_M (f \operatorname{div} X) \, dV_g.$$

10-16. Let (M, g) be an oriented Riemannian manifold. The operator $\Delta \colon C^\infty(M) \to C^\infty(M)$ defined by $\Delta u = \operatorname{div}(\operatorname{grad} u)$ is called the *Laplace operator*, and Δu is called the *Laplacian* of u. A function $u \in C^\infty(M)$ is said to be *harmonic* if $\Delta u = 0$.

(a) If M is compact, prove *Green's identities*:

$$\int_M u \Delta v \, dV_g + \int_M \langle \operatorname{grad} u, \operatorname{grad} v \rangle \, dV_g = \int_{\partial M} u \, Nv \, dV_{\widetilde{g}}. \tag{10.19}$$

$$\int_M (u \Delta v - v \Delta u) \, dV_g = \int_{\partial M} (u \, Nv - v \, Nu) dV_{\widetilde{g}}, \tag{10.20}$$

where N and $\widetilde{g}$ are as in Problem 10-15.

(b) If M is connected and $\partial M = \varnothing$, show that the only harmonic functions on M are the constants.

(c) If M is connected, $\partial M \neq \varnothing$, and u, v are harmonic functions on M whose restrictions to ∂M agree, show that $u \equiv v$.

10-17. Let (M, g) be a compact, connected, oriented Riemannian manifold, and let Δ be its Laplace operator. A real number λ is called an *eigenvalue* of Δ if there exists a smooth function u on M, not identically zero, such that $\Delta u = \lambda u$. In this case, u is called an *eigenfunction* corresponding to λ.

(a) Prove that 0 is an eigenvalue of Δ, and that all other eigenvalues are strictly negative.

(b) If u and v are eigenfunctions corresponding to distinct eigenvalues, show that $\int_M uv\, dV_g = 0$.

10-18. Let (M, g) be an oriented Riemannian n-manifold. This problem outlines an important generalization of the operator $*\colon C^\infty(M) \to \mathcal{A}^n(M)$ defined in this chapter.

(a) For each $k = 1, \ldots, n$, show that there is a unique inner product on $\Lambda^k(T_pM)$ with the following property: If (E_i) is any orthonormal basis for T_pM and (ε^i) is the dual basis, then $\{\varepsilon^I : I \text{ is increasing}\}$ is an orthonormal basis for $\Lambda^k(T_pM)$.

(b) For each $k = 0, \ldots, n$, show that there is a unique bundle map $*\colon \Lambda^k M \to \Lambda^{n-k} M$ satisfying

$$\omega \wedge *\eta = \langle \omega, \eta \rangle dV_g.$$

This map is called the *Hodge star operator*. [Hint: First prove uniqueness, and then define $*$ on elementary covectors defined with respect to an orthonormal basis.]

(c) Show that $*\colon \Lambda^0(M) \to \Lambda^n(M)$ is given by $*c = c\, dV_g$.

(d) Show that $* * \omega = (-1)^{k(n-k)}\omega$ if ω is a k-form.

10-19. Let (M, g) be a Riemannian manifold and $X \in \mathcal{T}(M)$.

(a) Show that

$$X \lrcorner\, dV_g = *X^\flat.$$

(b) Show that

$$\operatorname{div} X = *d * X^\flat.$$

(c) If M is 3-dimensional, show that

$$\operatorname{curl} X = (*dX^\flat)^\#.$$

10-20. Prove that $\mathbb{P}^n$ is orientable if and only if n is odd. [Hint: suppose η is a nonvanishing n-form on $\mathbb{P}^n$. Let $\pi : \mathbb{S}^n \to \mathbb{P}^n$ denote the universal covering map, and $\alpha : \mathbb{S}^n \to \mathbb{S}^n$ the antipodal map $\alpha(x) = -x$. Then compute $\alpha^*\pi^*\eta$ two ways.]

10-21. If ω is a symplectic form on a $2n$-manifold, show that $\omega \wedge \cdots \wedge \omega$ (the n-fold wedge product of ω with itself) is a nonvanishing $2n$-form on M, and thus every symplectic manifold is orientable.

10-22. Show that $\operatorname{curl} \circ \operatorname{grad} \equiv 0$ and $\operatorname{div} \circ \operatorname{curl} \equiv 0$ on any Riemannian 3-manifold.

10-23. On $\mathbb{R}^3$ with the Euclidean metric, show that

$$\operatorname{curl}\left(P\frac{\partial}{\partial x}+Q\frac{\partial}{\partial y}+R\frac{\partial}{\partial z}\right)$$
$$=\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\frac{\partial}{\partial x}+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\frac{\partial}{\partial y}+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\frac{\partial}{\partial z}.$$

10-24. Show that any finite product $M_1 \times \cdots \times M_k$ of smooth manifolds with corners is again a smooth manifold with corners. Give a counterexample to show that a finite product of smooth manifolds with boundary need not be a smooth manifold with boundary.

10-25. Suppose M is a smooth manifold with corners, and let $\mathcal{C}$ denote the set of corner points of M. Show that $M \smallsetminus \mathcal{C}$ is a smooth manifold with boundary.

11
De Rham Cohomology

In Chapter 9 we defined closed and exact forms: A differential form ω is closed if $d\omega = 0$, and exact if it is of the form $d\eta$. Because $d^2 = 0$, every exact form is closed. In this chapter, we explore the implications of the converse question: Is every closed form exact? The answer, in general, is no: In Example 4.23 we saw an example of a closed 1-form on $\mathbb{R}^n \smallsetminus \{0\}$ that was closed but not exact. In that example, the failure of exactness seemed to be a consequence of the "hole" in the center of the domain. For higher-degree forms, the answer to the question depends on subtle topological properties of the manifold, connected with the existence of "holes" of higher dimensions. Making this dependence quantitative leads to a new set of invariants of smooth manifolds, called the de Rham cohomology groups, which are the subject of this chapter.

There are many situations in which knowledge of which closed forms are exact has important consequences. For example, Stokes's theorem implies that if ω is exact, then the integral of ω over any compact submanifold without boundary is zero. Proposition 4.22 showed that a 1-form is conservative if and only if it is exact.

We begin by defining the de Rham cohomology groups and proving some of their basic properties, including diffeomorphism invariance. Then we prove that they are in fact *homotopy* invariants, which implies in particular that they are topological invariants. Using elementary methods, we compute some de Rham groups, including the zero-dimensional groups of all manifolds, the one-dimensional groups of simply connected manifolds, the top-dimensional groups of compact manifolds, and all of the de Rham groups of star-shaped open subsets of $\mathbb{R}^n$. Then we prove a general theorem that expresses the de Rham groups of a manifold in terms of those of its open subsets, called the Mayer–Vietoris theorem, and use it to compute all the de Rham groups of spheres.

At the end of the chapter, we turn our attention to the de Rham theorem, which expresses the equivalence of the de Rham groups with another set of groups defined purely topologically, the singular cohomology groups. To set the stage, we first give a brief summary of singular homology and cohomology theory, and prove that singular homology can be computed by restricting attention only to smooth simplices. In the last section, we prove the de Rham theorem.

The de Rham Cohomology Groups

In Chapter 4, we studied the closed 1-form

$$\omega = \frac{x\,dy - y\,dx}{x^2 + y^2}, \tag{11.1}$$

and showed that it is not exact on $\mathbb{R}^2 \smallsetminus \{0\}$, but it is exact on some smaller domains such as the right half-plane $H = \{(x,y) : x > 0\}$, where it is equal to $d\theta$ (see Example 4.26).

As we will see in this chapter, this behavior is typical: closed p-forms are always *locally* exact, so the question of whether a given closed form is exact depends on the global shape of the domain, not on the local properties of the form.

Let M be a smooth manifold. Because $d\colon \mathcal{A}^p(M) \to \mathcal{A}^{p+1}(M)$ is linear, its kernel and image are linear subspaces. We define

$$\begin{aligned}
\mathcal{Z}^p(M) &= \operatorname{Ker}[d\colon \mathcal{A}^p(M) \to \mathcal{A}^{p+1}(M)] = \{\text{closed } p\text{-forms}\},\\
\mathcal{B}^p(M) &= \operatorname{Im}[d\colon \mathcal{A}^{p-1}(M) \to \mathcal{A}^p(M)] = \{\text{exact } p\text{-forms}\}.
\end{aligned}$$

By convention, we consider $\mathcal{A}^p(M)$ to be the zero vector space when $p < 0$ or $p > n = \dim M$, so that that, for example, $\mathcal{B}^0(M) = 0$ and $\mathcal{Z}^n(M) = \mathcal{A}^n(M)$.

The fact that every exact form is closed implies that $\mathcal{B}^p(M) \subset \mathcal{Z}^p(M)$. Thus it makes sense to define the pth *de Rham cohomology group* (or just *de Rham group*) of M to be the quotient vector space

$$H^p_{dR}(M) = \frac{\mathcal{Z}^p(M)}{\mathcal{B}^p(M)}.$$

(It is, in particular, a group under vector addition. Perhaps "de Rham cohomology space" would be a more appropriate term, but because most other cohomology theories produce only groups it is traditional to use the term group in this context as well, bearing in mind that these "groups" are actually real vector spaces.) For any closed form ω on M, we let $[\omega]$ denote the equivalence class of ω in this quotient space, called the *cohomology class* of ω. Clearly $H^p_{dR}(M) = 0$ for $p > \dim M$, because $\mathcal{A}^p(M) = 0$ in that case. If $[\omega] = [\omega']$ (that is, if ω and ω' differ by an exact form), we say that ω and ω' are *cohomologous*.

The first thing we will show is that the de Rham groups are diffeomorphism invariants.

Proposition 11.1 (Induced Cohomology Maps). *For any smooth map $G\colon M \to N$, the pullback $G^*\colon \mathcal{A}^p(N) \to \mathcal{A}^p(M)$ carries $\mathcal{Z}^p(N)$ into $\mathcal{Z}^p(M)$ and $\mathcal{B}^p(N)$ into $\mathcal{B}^p(M)$. It thus descends to a linear map, still denoted by G^*, from $H^p_{dR}(N)$ to $H^p_{dR}(M)$, called the induced cohomology map. It has the following properties:*

(*a*) *If $F\colon N \to P$ is another smooth map, then*
$$(F \circ G)^* = G^* \circ F^*\colon H^p_{dR}(P) \to H^p_{dR}(M).$$

(*b*) *If Id_M denotes the identity map of M, then $(\mathrm{Id}_M)^*$ is the identity map of $H^p_{dR}(M)$.*

Proof. If ω is closed, then
$$d(G^*\omega) = G^*(d\omega) = 0,$$
so $G^*\omega$ is also closed. If $\omega = d\eta$ is exact, then
$$G^*\omega = G^*(d\eta) = d(G^*\eta),$$
which is also exact. Therefore, G^* maps $\mathcal{Z}^p(N)$ into $\mathcal{Z}^p(M)$ and $\mathcal{B}^p(N)$ into $\mathcal{B}^p(M)$. The induced cohomology map $G^*\colon H^p_{dR}(N) \to H^p_{dR}(M)$ is defined in the obvious way: For a closed p-form ω, let
$$G^*[\omega] = [G^*\omega].$$
If $\omega' = \omega + d\eta$, then $[G^*\omega'] = [G^*\omega + d(G^*\eta)] = [G^*\omega]$, so this map is well-defined. Properties (a) and (b) follow immediately from the analogous properties for the pullback map on forms. □

The next two corollaries are immediate.

Corollary 11.2 (Functoriality). *For each integer $p \geq 0$, the assignment $M \mapsto H^p_{dR}(M)$, $F \mapsto F^*$ is a contravariant functor from the category of smooth manifolds and smooth maps to the category of real vector spaces and linear maps.*

Corollary 11.3 (Diffeomorphism Invariance). *Diffeomorphic manifolds have isomorphic de Rham cohomology groups.*

Homotopy Invariance

In this section, we will present a profound generalization of Corollary 11.3, one surprising consequence of which will be that the de Rham cohomology groups are actually *topological* invariants. In fact, they are something much more: homotopy invariants.

Recall that a continuous map $F\colon M \to N$ between topological spaces is said to be a *homotopy equivalence* if there is a continuous map $G\colon N \to M$ such that $F \circ G \simeq \mathrm{Id}_N$ and $G \circ F \simeq \mathrm{Id}_M$. Such a map G is called a *homotopy inverse* for F. If there exists a homotopy equivalence between M and N, the two spaces are said to be *homotopy equivalent*. For example, the inclusion map $\iota\colon \mathbb{S}^{n-1} \hookrightarrow \mathbb{R}^n \smallsetminus \{0\}$ is a homotopy equivalence with homotopy inverse $r(x) = x/|x|$, because $r \circ \iota = \mathrm{Id}_{\mathbb{S}^{n-1}}$ and the identity map of $\mathbb{R}^n \smallsetminus \{0\}$ is homotopic to $\iota \circ r$ via the straight-line homotopy $H(x,t) = x + (1-t)x/|x|$.

The underlying fact that will allow us to prove the homotopy invariance of de Rham cohomology is that homotopic smooth maps induce the same cohomology map. To motivate the proof, suppose $F, G\colon M \to N$ are smoothly homotopic maps, and let us think about what needs to be shown. Given a closed p-form ω on N, we need somehow to produce a $(p-1)$-form η on M such that

$$d\eta = F^*\omega - G^*\omega. \tag{11.2}$$

One might hope to construct η in a systematic way, resulting in a map h from closed p-forms on M to $(p-1)$-forms on N that satisfies

$$d(h\omega) = F^*\omega - G^*\omega. \tag{11.3}$$

Instead of defining $h\omega$ only when ω is closed, it turns out to be far simpler to define for each p a map h from the space of *all* p-forms on N to the space of $(p-1)$-forms on M. Such maps cannot satisfy (11.3), but instead we will find maps that satisfy

$$d(h\omega) + h(d\omega) = F^*\omega - G^*\omega. \tag{11.4}$$

This implies (11.3) when ω is closed.

In general, if $F, G\colon M \to N$ are smooth maps, a collection of linear maps $h\colon \mathcal{A}^p(N) \to \mathcal{A}^{p-1}(M)$ such that (11.4) is satisfied for each p is called a *homotopy operator* between F^* and G^*. (The term *cochain homotopy* is used more frequently in the algebraic topology literature.) The key to our proof of homotopy invariance will be to construct a homotopy operator first in the following special case. For each $t \in [0,1]$, let $i_t\colon M \to M \times I$ be the embedding

$$i_t(x) = (x,t).$$

Clearly i_0 is homotopic to i_1. (The homotopy is the identity map of $M \times I$!)

Lemma 11.4 (Existence of a Homotopy Operator). *For any smooth manifold M, there exists a homotopy operator between the maps i_0 and i_1 defined above.*

Proof. For each p, we need to define a linear map $h\colon \mathcal{A}^p(M\times I) \to \mathcal{A}^{p-1}(M)$ such that

$$h(d\omega) + d(h\omega) = i_1^*\omega - i_0^*\omega. \tag{11.5}$$

We define h by the formula

$$h\omega = \int_0^1 \left(\frac{\partial}{\partial t} \lrcorner\, \omega\right) dt,$$

where t is the coordinate on I. More explicitly, $h\omega$ is the $(p-1)$-form on M whose action on vectors $X_1, \dots, X_{p-1} \in T_qM$ is

$$\begin{aligned}(h\omega)_q(X_1,\dots,X_{p-1}) &= \int_0^1 \left(\frac{\partial}{\partial t} \lrcorner\, \omega_{(q,t)}\right)(X_1,\dots,X_{p-1})\, dt \\ &= \int_0^1 \omega_{(q,t)}(\partial/\partial t, X_1,\dots,X_{p-1})\, dt.\end{aligned}$$

To show that h satisfies (11.5), we consider separately the cases in which $\omega = f(x,t)\, dt \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}}$ and $\omega = f(x,t)\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}$; since h is linear and every p-form on $M \times I$ can be written as a sum of such forms, this suffices.

Case I: $\omega = f(x,t)\, dt \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}}$. In this case, because $dt \wedge dt = 0$,

$$\begin{aligned}d(h\omega) &= d\left(\left(\int_0^1 f(x,t)\, dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}}\right) \\ &= \frac{\partial}{\partial x^j}\left(\int_0^1 f(x,t)\, dt\right) dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}} \\ &= \left(\int_0^1 \frac{\partial f}{\partial x^j}(x,t)\, dt\right) dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}}.\end{aligned}$$

On the other hand,

$$\begin{aligned}h(d\omega) &= h\left(\frac{\partial f}{\partial x^j} dx^j \wedge dt \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}}\right) \\ &= \int_0^1 \frac{\partial f}{\partial x^j}(x,t) \frac{\partial}{\partial t} \lrcorner\, (dx^j \wedge dt \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}})\, dt \\ &= -\left(\int_0^1 \frac{\partial f}{\partial x^j}(x,t)\, dt\right) dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}} \\ &= -d(h\omega).\end{aligned}$$

Thus the left-hand side of (11.5) is zero in this case. The right-hand side is zero as well, because $i_0^* dt = i_1^* dt = 0$ (since $t \circ i_0$ and $t \circ i_1$ are constant functions).

Case II: $\omega = f\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}$. Now $\partial/\partial t \lrcorner \omega = 0$, which implies that $d(h\omega) = 0$. On the other hand, by the fundamental theorem of calculus,

$$\begin{aligned} h(d\omega) &= h\left(\frac{\partial f}{\partial t} dt \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p} + \text{ terms without } dt\right) \\ &= \left(\int_0^1 \frac{\partial f}{\partial t}(x,t)dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_p} \\ &= (f(x,1) - f(x,0)) dx^{i_1} \wedge \cdots \wedge dx^{i_p} \\ &= i_1^*\omega - i_0^*\omega, \end{aligned}$$

which proves (11.5) in this case. □

Proposition 11.5. *Let $F, G\colon M \to N$ be homotopic smooth maps. For every p, the induced cohomology maps $F^*, G^*\colon H^p_{dR}(N) \to H^p_{dR}(M)$ are equal.*

Proof. By Proposition 6.20, there is a smooth homotopy $H\colon M \times I \to N$ from F to G. This means that $H \circ i_0 = F$ and $H \circ i_1 = G$, where $i_0, i_1\colon M \to M \times I$ are defined as above. Let $\widetilde{h}$ be the composite map $\widetilde{h} = h \circ H^*\colon \mathcal{A}^p(N) \to \mathcal{A}^{p-1}(M)$:

$$\mathcal{A}^p(N) \xrightarrow{H^*} \mathcal{A}^p(M \times I) \xrightarrow{h} \mathcal{A}^{p-1}(M),$$

where h is the homotopy operator constructed in Lemma 11.4.

For any $\omega \in \mathcal{A}^p(N)$, we compute

$$\begin{aligned} \widetilde{h}(d\omega) + d(\widetilde{h}\omega) &= h(H^* d\omega) + d(hH^*\omega) \\ &= hd(H^*\omega) + dh(H^*\omega) \\ &= i_1^* H^*\omega - i_0^* H^*\omega \\ &= (H \circ i_1)^*\omega - (H \circ i_0)^*\omega \\ &= G^*\omega - F^*\omega. \end{aligned}$$

Thus if ω is closed,

$$\begin{aligned} G^*[\omega] - F^*[\omega] &= [G^*\omega - F^*\omega] \\ &= [\widetilde{h}(d\omega) + d(\widetilde{h}\omega)] \\ &= 0, \end{aligned}$$

where the last line follows from $d\omega = 0$ and the fact that the cohomology class of any exact form is zero. □

The next theorem is the main result of this section.

Theorem 11.6 (Homotopy Invariance of de Rham Cohomology). *If M and N are homotopy equivalent smooth manifolds, then $H^p_{dR}(M) \cong H^p_{dR}(N)$ for each p.*

Proof. Suppose $F\colon M \to N$ is a homotopy equivalence, with homotopy inverse $G\colon N \to M$. By Theorem 6.19, there are smooth maps $\widetilde{F}\colon M \to N$ homotopic to F and $\widetilde{G}\colon N \to M$ homotopic to G. Because homotopy is preserved by composition, it follows that $\widetilde{F} \circ \widetilde{G} \simeq F \circ G \simeq \mathrm{Id}_N$ and $\widetilde{G} \circ \widetilde{F} \simeq G \circ F \simeq \mathrm{Id}_M$, so $\widetilde{F}$ and $\widetilde{G}$ are homotopy inverses of each other.

Now Proposition 11.5 shows that, on cohomology,

$$\widetilde{F}^* \circ \widetilde{G}^* = (\widetilde{G} \circ \widetilde{F})^* = (\mathrm{Id}_M)^* = \mathrm{Id}_{H^p_{dR}(M)}.$$

The same argument shows that $\widetilde{G}^* \circ \widetilde{F}^*$ is also the identity, so $\widetilde{F}^*\colon H^p_{dR}(N) \to H^p_{dR}(M)$ is an isomorphism. □

Corollary 11.7. *The de Rham cohomology groups are topological invariants: If M and N are homeomorphic smooth manifolds, then their de Rham groups are isomorphic.*

This result is remarkable, because the definition of the de Rham groups of M is intimately tied up with its smooth structure, and we had no reason to expect that different differentiable structures on the same topological manifold should give rise to the same de Rham groups.

Computations

The direct computation of the de Rham groups is not easy in general. However, in this section, we will compute them in several special cases.

We begin with disjoint unions.

Proposition 11.8 (Cohomology of Disjoint Unions). *Let $\{M_j\}$ be a countable collection of smooth manifolds, and let $M = \coprod_j M_j$. For each p, the inclusion maps $\iota_j\colon M_j \to M$ induce an isomorphism from $H^p_{dR}(M)$ to the direct product space $\prod_j H^p_{dR}(M_j)$.*

Proof. The pullback maps $\iota_j^*\colon \mathcal{A}^p(M) \to \mathcal{A}^p(M_j)$ already induce an isomorphism from $\mathcal{A}^p(M)$ to $\prod_j \mathcal{A}^p(M_j)$, namely

$$\omega \mapsto (\iota_1^*\omega, \iota_2^*\omega, \dots) = (\omega|_{M_1}, \omega|_{M_2}, \dots).$$

This map is injective because any p-form whose restriction to each M_j is zero must itself be zero, and it is surjective because giving an arbitrary p-form on each M_j defines one on M. □

Because of this proposition, each de Rham group of a disconnected manifold is just the direct product of the corresponding groups of its components. Thus we can concentrate henceforth on computing the de Rham groups of connected manifolds.

Our next computation gives an explicit characterization of zero-dimensional cohomology.

Proposition 11.9 (Zero-Dimensional Cohomology). *If M is a connected manifold, $H^0_{dR}(M)$ is equal to the space of constant functions and is therefore one-dimensional.*

Proof. Because there are no (-1)-forms, $\mathcal{B}^0(M) = 0$. A closed 0-form is a smooth function f such that $df = 0$, and since M is connected this is true if and only if f is constant. Thus $H^0_{dR}(M) = \mathcal{Z}^0(M) = \{\text{constants}\}$. □

Corollary 11.10 (Cohomology of Zero-Manifolds). *If M is a 0-dimensional manifold, the dimension of $H^0_{dR}(M)$ is equal to the cardinality of M, and all other de Rham cohomology groups vanish.*

Proof. By Propositions 11.8 and 11.9, $H^0_{dR}(M)$ is isomorphic to the direct product of one copy of $\mathbb{R}$ for each component of M, which is to say each point. □

Next we examine the de Rham cohomology of Euclidean space, and more generally of its star-shaped open subsets. (Recall that a subset $V \subset \mathbb{R}^n$ is said to be star-shaped with respect to a point $q \in V$ if for every $x \in V$, the line segment from q to x is entirely contained in V.) In Proposition 4.27, we showed that every closed 1-form on a star-shaped open subset of $\mathbb{R}^n$ is exact. The next theorem is a generalization of that result.

Theorem 11.11 (The Poincaré Lemma). *Let U be a star-shaped open subset of $\mathbb{R}^n$. Then $H^p_{dR}(U) = 0$ for $p \geq 1$.*

Proof. Suppose $U \subset \mathbb{R}^n$ is star-shaped with respect to q. The key feature of star-shaped sets is that they are *contractible*, which means the identity map of U is homotopic to the constant map sending U to q, by the obvious straight-line homotopy:

$$H(x,t) = q + t(x-q).$$

Thus the inclusion of $\{q\}$ into U is a homotopy equivalence. The Poincaré lemma then follows from the homotopy invariance of H^p_{dR} together with the obvious fact that $H^p_{dR}(\{q\}) = 0$ for $p > 0$ because $\{q\}$ is a 0-manifold. □

Exercise 11.1. If $U \subset \mathbb{R}^n$ is open and star-shaped with respect to 0 and $\omega = \sum' \omega_I \, dx^I$ is a closed p-form on U, show by tracing through the proof of the Poincaré lemma that the $(p-1)$-form η given explicitly by the formula

$$\eta = \sum_I{}' \sum_{q=1}^{p} (-1)^{q-1} \left(\int_0^1 t^{p-1} \omega_I(tx)\, dt \right) x^{i_q}\, dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_q}} \wedge \cdots \wedge dx^{i_p}$$

satisfies $d\eta = \omega$. When ω is a 1-form, show that η is equal to the potential function f defined in Proposition 4.27.

The next two results are easy corollaries of the Poincaré lemma.

Corollary 11.12 (Cohomology of Euclidean Space). *For all $p \geq 1$, $H^p_{dR}(\mathbb{R}^n) = 0$.*

Proof. Euclidean space $\mathbb{R}^n$ is star-shaped. □

Corollary 11.13 (Local Exactness of Closed Forms). *Let M be a smooth manifold, and let ω be a closed p-form on M, $p \geq 1$. For any $q \in M$, there is a neighborhood U of q on which ω is exact.*

Proof. Every $q \in M$ has a neighborhood diffeomorphic to an open ball in $\mathbb{R}^n$, which is star-shaped. The result follows from the diffeomorphism invariance of de Rham cohomology. □

One of the most interesting special cases is that of simply connected manifolds, for which we can compute the first cohomology explicitly.

Theorem 11.14. *If M is a simply connected smooth manifold, then $H^1_{dR}(M) = 0$.*

Proof. Let ω be a closed 1-form on M. We need to show that ω is exact. By Theorem 4.22, this is true if and only ω is conservative, that is, if and only if all line integrals of ω depend only on endpoints. Since any two piecewise smooth curve segments with the same endpoints are path homotopic, the result follows from Theorem 10.33. □

Finally, we turn our attention to the top-dimensional cohomology of compact manifolds. We begin with the orientable case. Suppose M is an orientable compact smooth manifold. There is a natural linear map $I\colon \mathcal{A}^n(M) \to \mathbb{R}$ given by integration over M:

$$I(\omega) = \int_M \omega.$$

Because the integral of any exact form is zero, I descends to a linear map, still denoted by the same symbol, from $H^n_{dR}(M)$ to $\mathbb{R}$. (Note that every n-form on an n-manifold is closed.)

Theorem 11.15 (Top Cohomology, Orientable Case). *For any compact, connected, orientable, smooth n-manifold M, the map $I\colon H^n_{dR}(M) \to \mathbb{R}$ is an isomorphism. Thus $H^n_{dR}(M)$ is one-dimensional, spanned by the cohomology class of any orientation form.*

Proof. The 0-dimensional case is an immediate consequence of Corollary 11.10, so we may assume that $n \geq 1$. Let Ω be an orientation form for M, and set $c = \int_M \Omega$. By Proposition 10.20(c), $c > 0$. Therefore, for any $a \in \mathbb{R}$, $I[a\omega] = ac$, so $I\colon H^n_{dR}(M) \to \mathbb{R}$ is surjective. To complete the proof, we need only show that it is injective. In other words, we have to show the following: If ω is any n-form satisfying $\int_M \omega = 0$, then ω is exact.

Let $\{U_1, \dots, U_m\}$ be a finite cover of M by open sets that are diffeomorphic to $\mathbb{R}^n$, and let $M_k = U_1 \cup \cdots \cup U_k$ for $k = 1, \dots, m$. Since M is connected, by reordering the sets if necessary, we may assume that $M_k \cap U_{k+1} \neq \varnothing$ for each k. We will prove the following claim by induction on k: *If ω is a compactly supported n-form on M_k that satisfies $\int_{M_k} \omega = 0$, then there exists a compactly supported $(n-1)$-form η on M_k such that $d\eta = \omega$.* When $k = m$, this is the statement we are seeking to prove, because every form on a compact manifold is compactly supported.

For $k = 1$, since $M_1 = U_1$ is diffeomorphic to $\mathbb{R}^n$, the claim reduces to a statement about compactly supported forms on $\mathbb{R}^n$. The proof of this statement is somewhat technical, so we postpone it to the end of this section (Lemma 11.19). Assuming this for now, we continue with the induction.

Assume the claim is true for some $k \geq 1$, and suppose ω is a compactly supported n-form on $M_{k+1} = M_k \cup U_{k+1}$ that satisfies $\int_{M_{k+1}} \omega = 0$. Choose an auxiliary n-form $\Omega \in \mathcal{A}^n(M_{k+1})$ that is compactly supported in $M_k \cap U_{k+1}$ and satisfies $\int_{M_{k+1}} \Omega = 1$. (Such a form is easily constructed by using a bump function in coordinates.) Let $\{\varphi, \psi\}$ be a partition of unity for M_{k+1} subordinate to the cover $\{M_k, U_{k+1}\}$.

Let $c = \int_{M_{k+1}} \varphi\omega$. Observe that $\varphi\omega - c\Omega$ is compactly supported in M_k, and its integral is equal to zero by our choice of c. Therefore, by the induction hypothesis, there is a compactly supported $(n-1)$-form α on M_k such that $d\alpha = \varphi\omega - c\Omega$. Similarly, $\psi\omega + c\Omega$ is compactly supported in U_{k+1}, and its integral is

$$\begin{aligned}
\int_{U_{k+1}} (\psi\omega + c\Omega) &= \int_{M_{k+1}} (1-\varphi)\omega + c\int_{M_{k+1}} \Omega \\
&= \int_{M_{k+1}} \omega - \int_{M_{k+1}} \varphi\omega + c \\
&= 0.
\end{aligned}$$

Thus by Lemma 11.19, there exists another $(n-1)$-form β, compactly supported in U_{k+1}, such that $d\beta = \psi\omega + c\Omega$. Both α and β can be extended by zero to smooth compactly supported forms on M_{k+1}. We compute

$$d(\alpha+\beta) = (\varphi\omega - c\Omega) + (\psi\omega + c\Omega) = (\varphi+\psi)\omega = \omega,$$

which completes the inductive step. $\square$

We now have enough information to compute the de Rham cohomology of the circle and the punctured plane completely.

Corollary 11.16 (Cohomology of the Circle). *The de Rham cohomology groups of the circle are as follows: $H^0_{dR}(\mathbb{S}^1)$ and $H^1_{dR}(\mathbb{S}^1)$ are both 1-dimensional, spanned by the function 1 and the cohomology class of the form ω defined by* (11.1)*, respectively.*

Proof. Because the restriction of ω to $\mathbb{S}^1$ never vanishes, it is an orientation form. $\square$

Corollary 11.17 (Cohomology of the Punctured Plane). *Let $M = \mathbb{R}^2 \smallsetminus \{0\}$. Then $H^0_{dR}(M)$ and $H^1_{dR}(M)$ are both 1-dimensional, spanned by the function 1 and the cohomology class of the form ω defined by* (11.1)*, respectively.*

Proof. Because the inclusion map $\iota\colon \mathbb{S}^1 \hookrightarrow M$ is a homotopy equivalence, $\iota^*\colon H^1_{dR}(M) \to H^1_{dR}(\mathbb{S}^1)$ is an isomorphism. The result then follows from the preceding corollary. $\square$

Next we consider the nonorientable case.

Theorem 11.18 (Top Cohomology, Nonorientable Case). *Let M be a compact, connected, nonorientable, smooth n-manifold. Then $H^n_{dR}(M) = 0$.*

Proof. We have to show that every n-form on M is exact. By the result of Problem 10-6, there is an orientable manifold $\widetilde{M}$ and a 2-sheeted smooth covering map $\pi\colon \widetilde{M} \to M$. Let $\alpha\colon \widetilde{M} \to M$ be the map that interchanges the two points in each fiber of π. If $U \subset M$ is any evenly covered open set, then α just interchanges the two components of $\pi^{-1}(U)$, so α is a smooth covering transformation; in fact, it is the unique nontrivial covering transformation of π. Now, α cannot be orientation-preserving—if it were, the entire covering group $\{\mathrm{Id}_{\widetilde{M}}, \alpha\}$ would be orientation-preserving, and then M would be orientable by the result of Problem 10-4. By connectedness of $\widetilde{M}$ and the fact that α is a diffeomorphism, it follows that α is orientation-reversing.

Suppose ω is any n-form on M, and let $\Omega = \pi^*\omega \in \mathcal{A}^n(\widetilde{M})$. Then $\pi \circ \alpha = \pi$ implies

$$\alpha^*\Omega = \alpha^*\pi^*\omega = (\pi\circ\alpha)^*\omega = \pi^*\omega = \Omega.$$

Because α is orientation-reversing, therefore, we conclude from Proposition 10.16 that

$$\int_{\widetilde{M}} \Omega = -\int_{\widetilde{M}} \alpha^*\Omega = -\int_{\widetilde{M}} \Omega.$$

This implies that $\int_{\widetilde{M}} \Omega = 0$, so by Theorem 11.15, there exists $\eta \in \mathcal{A}^{n-1}(\widetilde{M})$ such that $d\eta = \Omega$. Let $\widetilde{\eta} = \frac{1}{2}(\eta + \alpha^*\eta)$. Using the fact that $\alpha \circ \alpha = \mathrm{Id}_{\widetilde{M}}$, we compute

$$\alpha^*\widetilde{\eta} = \tfrac{1}{2}(\alpha^*\eta + (\alpha \circ \alpha)^*\eta) = \widetilde{\eta}$$

and

$$\begin{aligned} d\widetilde{\eta} &= \tfrac{1}{2}(d\eta + d\alpha^*\eta) \\ &= \tfrac{1}{2}(d\eta + \alpha^* d\eta) \\ &= \tfrac{1}{2}(\Omega + \alpha^*\Omega) \\ &= \Omega. \end{aligned}$$

Now let $U \subset M$ be any evenly covered open set. There are exactly two smooth local sections $\sigma_1, \sigma_2 \colon U \to \widetilde{M}$ over U, which are related by $\sigma_2 = \alpha \circ \sigma_1$. Observe that

$$\sigma_2^*\widetilde{\eta} = (\alpha \circ \sigma_1)^*\widetilde{\eta} = \sigma_1^*\alpha^*\widetilde{\eta} = \sigma_1^*\widetilde{\eta}.$$

Therefore, we can define a global $(n-1)$-form γ on M by setting $\gamma = \sigma^*\widetilde{\eta}$ for any local section σ. To determine its exterior derivative, choose a smooth local section σ in a neighborhood of any point, and compute

$$d\gamma = d\sigma^*\widetilde{\eta} = \sigma^* d\widetilde{\eta} = \sigma^*\Omega = \sigma^*\pi^*\omega = (\pi \circ \sigma)^*\omega = \omega,$$

because $\pi \circ \sigma = \mathrm{Id}_U$. □

Finally, here is the technical lemma that is needed to complete the proof of Theorem 11.15. It can be thought of as a refinement of the Poincaré lemma for compactly supported n-forms.

Lemma 11.19. *Let $n \geq 1$, and suppose ω is a compactly supported n-form on $\mathbb{R}^n$ such that $\int_{\mathbb{R}^n} \omega = 0$. Then there exists a compactly supported $(n-1)$-form η on $\mathbb{R}^n$ such that $d\eta = \omega$.*

Remark. Of course, we know that ω is exact by the Poincaré lemma, so the novelty here is the claim that we can find a *compactly supported* $(n-1)$-form η such that $d\eta = \omega$.

Proof. We will carry out the proof by induction on n. For $n = 1$, we can write $\omega = f\,dx$ for some smooth compactly supported function f. Choose $R > 0$ such that $\operatorname{supp} f \subset [-R, R]$, and define $F \colon \mathbb{R} \to \mathbb{R}$ by

$$F(x) = \int_{-R}^{x} f(t)\,dt.$$

Then clearly $dF = F'(x)\,dx = f\,dx = \omega$. When $x < -R$, $F(x) = 0$ by our choice of R. When $x > R$, the fact that $\int_{\mathbb{R}} \omega = 0$ translates to

$$0 = \int_{-R}^{R} f(t)\,dt = \int_{-R}^{x} f(t)\,dt = F(x),$$

so in fact $\operatorname{supp} F \subset [-R, R]$. This completes the proof for the case $n = 1$.

Now let $n \geq 1$, and suppose the lemma is true on $\mathbb{R}^n$. Let us consider $\mathbb{R}^{n+1}$ as the product space $\mathbb{R} \times \mathbb{R}^n$, with coordinates $(y, x) = (y, x^1, \ldots, x^n)$. Let $\Omega = dx^1 \wedge \cdots \wedge dx^n$, considered as an n-form on $\mathbb{R} \times \mathbb{R}^n$.

Suppose ω is any compactly supported $(n+1)$-form on $\mathbb{R} \times \mathbb{R}^n$ such that

$$\int_{\mathbb{R} \times \mathbb{R}^n} \omega = 0.$$

Then ω can be written

$$\omega = f\,dy \wedge \Omega$$

for some compactly supported smooth function f. Choose $R > 0$ such that $\operatorname{supp} f$ is contained in the set $\{(y, x) : |y| \leq R \text{ and } |x| \leq R\}$.

Let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be any bump function supported in $[-R, R]$ and satisfying $\int_{\mathbb{R}} \varphi(y)\,dy = 1$. Define smooth functions $e, E, F, \widetilde{F}\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ as follows:

$$\begin{aligned}
e(y, x) &= \varphi(y);\\
E(y, x) &= \int_{-R}^{y} \varphi(t)\,dt;\\
F(y, x) &= \int_{-R}^{y} f(t, x)\,dt;\\
\widetilde{F}(y, x) &= \int_{-R}^{R} f(t, x)\,dt = F(R, x).
\end{aligned}$$

These functions have the properties

$$\frac{\partial e}{\partial x^j} = \frac{\partial E}{\partial x^j} = 0; \quad \frac{\partial E}{\partial y} = e; \quad \frac{\partial F}{\partial y} = f; \quad \frac{\partial \widetilde{F}}{\partial y} = 0.$$

Let Σ denote the n-form $\iota^*(\widetilde{F}\,\Omega)$ on $\mathbb{R}^n$, where $\iota\colon \mathbb{R}^n \hookrightarrow \mathbb{R} \times \mathbb{R}^n$ is the embedding $\iota(x) = (0, x)$. Its integral satisfies

$$\begin{aligned}
\int_{\mathbb{R}^n} \Sigma &= \int_{\mathbb{R}^n} \widetilde{F}(0, x)\,dx^1 \cdots dx^n\\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}} f(y, x)\,dy\,dx^1 \cdots dx^n\\
&= \int_{\mathbb{R} \times \mathbb{R}^n} \omega\\
&= 0.
\end{aligned}$$

Therefore, by the inductive hypothesis, there exists a compactly supported $(n-1)$-form σ on $\mathbb{R}^n$ such that $d\sigma = \Sigma$.

Now define an n-form η on $\mathbb{R} \times \mathbb{R}^n$ by

$$\eta = (F - \widetilde{F}E)\Omega - e\, dy \wedge \pi^*\sigma,$$

where $\pi\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is the projection. When $y < R$, we have $F \equiv \widetilde{F} \equiv e \equiv 0$, so η vanishes there. When $y > R$, then $e \equiv 0$, $E \equiv 1$, and $F \equiv \widetilde{F}$, so $\eta = 0$ there also. Finally, when $|x|$ is sufficiently large, then F, $\widetilde{F}$, and σ are all zero, so $\eta = 0$ there as well. Thus η is compactly supported.

To show that $d\eta = \omega$, we compute

$$d\eta = dF \wedge \Omega - \widetilde{F}\, dE \wedge \Omega - E\, d\widetilde{F} \wedge \Omega - de \wedge dy \wedge \pi^*\sigma + e\, dy \wedge d(\pi^*\sigma).$$

We consider each of these five terms separately. The first term is

$$\begin{aligned} dF \wedge \Omega &= \left(\frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial x^j} dx^j \right) \wedge \Omega \\ &= f\, dy \wedge \Omega \\ &= \omega, \end{aligned}$$

because $dx^j \wedge \Omega = 0$. For the second term we have

$$-\widetilde{F}\, dE \wedge \Omega = -\widetilde{F} \frac{\partial E}{\partial y} dy \wedge \Omega = -\widetilde{F} e\, dy \wedge \Omega.$$

Because $\partial \widetilde{F}/\partial y = 0$, the third term reduces to

$$-E\, d\widetilde{F} \wedge \Omega = -E \frac{\partial \widetilde{F}}{\partial x^j} dx^j \wedge \Omega = 0.$$

The fourth term is likewise zero because $de \wedge dy = \varphi'(y) dy \wedge dy = 0$. For the fifth term, we observe that $\pi^*\iota^*\widetilde{F} = \widetilde{F}$ because $\widetilde{F}$ is independent of y, and $\pi^*\iota^*\Omega = \Omega$ by direct computation, and therefore $\pi^*\Sigma = \widetilde{F}\Omega$. Therefore,

$$\begin{aligned} e\, dy \wedge d(\pi^*\sigma) &= e\, dy \wedge \pi^*(d\sigma) \\ &= e\, dy \wedge \pi^*(\Sigma) \\ &= \widetilde{F} e\, dy \wedge \Omega. \end{aligned}$$

Thus the second and fifth terms cancel, and we are left with $d\eta = \omega$. □

For some purposes, it is useful to define a generalization of the de Rham cohomology groups using only compactly supported forms. Let $\mathcal{A}_c^p(M)$ denote the space of compactly-supported p-forms on M. The pth *de Rham cohomology group of M with compact support* is the quotient space

$$H_c^p(M) = \frac{\operatorname{Ker}[d\colon \mathcal{A}_c^p(M) \to \mathcal{A}_c^{p+1}(M)]}{\operatorname{Im}[d\colon \mathcal{A}_c^{p-1}(M) \to \mathcal{A}_c^p(M)]}.$$

Of course, when M is compact, this just reduces to ordinary de Rham cohomology. But for noncompact manifolds, the two groups can be different, as the next exercise shows.

Exercise 11.2. Using Lemma 11.19, show that $H_c^n(\mathbb{R}^n)$ is 1-dimensional.

We will not use compactly supported cohomology in this book, but it plays an important role in algebraic topology.

The Mayer–Vietoris Theorem

In this section, we prove a very general theorem that can be used to compute the de Rham cohomology groups of many spaces, by expressing them as unions of open submanifolds with simpler cohomology.

For this purpose, we need to introduce some simple algebraic concepts. More details about the ideas introduced here can be found in [Lee00, Chapter 13] or in any textbook on algebraic topology.

Let $\mathcal{R}$ be a commutative ring, and let A^* be any sequence of $\mathcal{R}$-modules and linear maps:

$$\cdots \to A^{p-1} \xrightarrow{d} A^p \xrightarrow{d} A^{p+1} \to \cdots .$$

(In all of our applications, the ring will be either $\mathbb{Z}$, in which case we are looking at abelian groups and homomorphisms, or $\mathbb{R}$, in which case we have vector spaces and linear map. The terminology of modules is just a convenient way to combine the two cases.)

Such a sequence is said to be a *complex* if the composition of any two successive applications of d is the zero map:

$$d \circ d = 0\colon A^p \to A^{p+2} \quad \text{for each } p.$$

It is called an *exact sequence* if the image of each d is equal to the kernel of the next:

$$\operatorname{Im}[d\colon A^{p-1} \to A^p] = \operatorname{Ker}[d\colon A^p \to A^{p+1}].$$

Clearly every exact sequence is a complex, but the converse need not be true. If A^* is a complex, then the image of each map d is contained in the kernel of the next, so we define the pth *cohomology group* of A^* to be the quotient module

$$H^p(A^*) = \frac{\operatorname{Ker}[d\colon A^p \to A^{p+1}]}{\operatorname{Im}[d\colon A^{p-1} \to A^p]}.$$

It can be thought of as a quantitative measure of the failure of exactness at A^p. (In algebraic topology, a complex as we have defined it is usually called a *cochain complex*, while a *chain complex* is defined similarly except that the maps go in the direction of decreasing indices:

$$\cdots \to A_{p+1} \xrightarrow{\partial} A_p \xrightarrow{\partial} A_{p-1} \to \cdots .$$

In that case, the term *homology* is used in place of cohomology.)

If A^* and B^* are complexes, a *cochain map* from A^* to B^*, denoted by $F\colon A^* \to B^*$, is a collection of linear maps $F\colon A^p \to B^p$ (it is easiest to use the same symbol for all of the maps) such that the following diagram commutes for each p:

$$\begin{tikzcd}
\cdots \arrow[r] & A^p \arrow[r, "d"] \arrow[d, "F"'] & A^{p+1} \arrow[r] \arrow[d, "F"] & \cdots \\
\cdots \arrow[r] & B^p \arrow[r, "d"'] & B^{p+1} \arrow[r] & \cdots .
\end{tikzcd}$$

The fact that $F \circ d = d \circ F$ means that any cochain map induces a linear map on cohomology $F^*\colon H^p(A^*) \to H^p(B^*)$ for each p, just as in the case of de Rham cohomology.

A *short exact sequence of complexes* consists of three complexes A^*, B^*, C^*, together with cochain maps

$$0 \to A^* \xrightarrow{F} B^* \xrightarrow{G} C^* \to 0$$

such that each sequence

$$0 \to A^p \xrightarrow{F} B^p \xrightarrow{G} C^p \to 0$$

is exact. This means F is injective, G is surjective, and $\operatorname{Im} F = \operatorname{Ker} G$.

Lemma 11.20 (The Zigzag Lemma). *Given a short exact sequence of complexes as above, for each p there is a linear map*

$$\delta\colon H^p(C^*) \to H^{p+1}(A^*),$$

called the connecting homomorphism, such that the following sequence is exact:

$$\cdots \xrightarrow{\delta} H^p(A^*) \xrightarrow{F^*} H^p(B^*) \xrightarrow{G^*} H^p(C^*) \xrightarrow{\delta} H^{p+1}(A^*) \xrightarrow{F^*} \cdots . \qquad (11.6)$$

Proof. We will sketch only the main idea; you can either carry out the details yourself or look them up.

The hypothesis means that the following diagram commutes and has exact horizontal rows:

$$\begin{tikzcd}
0 \arrow[r] & A^p \arrow[r, "F"] \arrow[d, "d"] & B^p \arrow[r, "G"] \arrow[d, "d"] & C^p \arrow[r] \arrow[d, "d"] & 0 \\
0 \arrow[r] & A^{p+1} \arrow[r, "F"] \arrow[d, "d"] & B^{p+1} \arrow[r, "G"] \arrow[d, "d"] & C^{p+1} \arrow[r] \arrow[d, "d"] & 0 \\
0 \arrow[r] & A^{p+2} \arrow[r, "F"] & B^{p+2} \arrow[r, "G"] & C^{p+2} \arrow[r] & 0.
\end{tikzcd}$$

Suppose $c^p \in C^p$ represents a cohomology class; this means that $dc^p = 0$. Since $G\colon B^p \to C^p$ is surjective, there is some element $b^p \in B^p$ such that $Gb^p = c^p$. Because the diagram commutes, $Gdb^p = dGb^p = dc^p = 0$, and therefore $db^p \in \operatorname{Ker} G = \operatorname{Im} F$. Thus there exists $a^{p+1} \in A^{p+1}$ satisfying $Fa^{p+1} = db^p$. By commutativity of the diagram again, $Fda^{p+1} = dFa^{p+1} = ddb^p = 0$. Since F is injective, this implies $da^{p+1} = 0$, so a^{p+1} represents a cohomology class in $H^{p+1}(A^*)$. The connecting homomorphism δ is defined by setting $\delta[c^p] = [a^{p+1}]$ for any such $a^{p+1} \in A^{p+1}$, that is, provided there exists $b^p \in B^p$ such that

$$\begin{aligned} Gb^p &= c^p, \\ Fa^{p+1} &= db^p. \end{aligned}$$

A number of facts have to be verified: that the cohomology class $[a^{p+1}]$ is well-defined, independently of the choices made along the way; that the resulting map δ is linear; and that the resulting sequence (11.6) is exact. Each of these verifications is a routine "diagram chase" like the one we used to define δ; the details are left as an exercise. □

Exercise 11.3. Complete (or look up) the proof of the zigzag lemma.

The situation in which we will apply this lemma is the following. Suppose M is a smooth manifold, and U, V are open subsets of M such that $M = U \cup V$. We have the following diagram of inclusions:

$$\begin{tikzcd}
& U \arrow[dr, "k"] & \\
U \cap V \arrow[ur, "i"] \arrow[dr, "j"'] & & M, \\
& V \arrow[ur, "l"'] &
\end{tikzcd} \tag{11.7}$$

which induce pullback maps on differential forms:

$$\begin{tikzcd} & \mathcal{A}^p(U) \arrow[dr, "i^*"] & \\ \mathcal{A}^p(M) \arrow[ur, "k^*"] \arrow[dr, "l^*"'] & & \mathcal{A}^p(U \cap V), \\ & \mathcal{A}^p(V) \arrow[ur, "j^*"'] & \end{tikzcd}$$

as well as corresponding induced cohomology maps. Note that these pullback maps are really just restrictions: For example, $k^*\omega = \omega|_U$. We will consider the following sequence:

$$0 \to \mathcal{A}^p(M) \xrightarrow{k^* \oplus l^*} \mathcal{A}^p(U) \oplus \mathcal{A}^p(V) \xrightarrow{i^* - j^*} \mathcal{A}^p(U \cap V) \to 0, \tag{11.8}$$

where

$$\begin{aligned} (k^* \oplus l^*)\omega &= (k^*\omega, l^*\omega), \\ (i^* - j^*)(\omega, \eta) &= i^*\omega - j^*\eta. \end{aligned} \tag{11.9}$$

Because pullbacks commute with d, these maps descend to linear maps on the corresponding de Rham cohomology groups.

Theorem 11.21 (Mayer–Vietoris). *Let M be a smooth manifold, and let U, V be open subsets of M whose union is M. For each p, there is a linear map $\delta\colon H^p_{dR}(U \cap V) \to H^{p+1}_{dR}(M)$ such that the following sequence is exact:*

$$\begin{aligned} \cdots \xrightarrow{\delta} H^p_{dR}(M) \xrightarrow{k^* \oplus l^*} H^p_{dR}(U) \oplus H^p_{dR}(V) \xrightarrow{i^* - j^*} H^p_{dR}(U \cap V) \\ \xrightarrow{\delta} H^{p+1}_{dR}(M) \xrightarrow{k^* \oplus l^*} \cdots \end{aligned} \tag{11.10}$$

Remark. The sequence (11.10) is called the *Mayer–Vietoris sequence* for the open cover $\{U, V\}$.

Proof. The heart of the proof will be to show that the sequence (11.8) is exact for each p. Because pullback maps commute with the exterior derivative, (11.8) therefore defines a short exact sequence of chain maps, and the Mayer–Vietoris theorem follows immediately from the zigzag lemma.

We begin by proving exactness at $\mathcal{A}^p(M)$, which just means showing that $k^* \oplus l^*$ is injective. Suppose that $\sigma \in \mathcal{A}^p(M)$ satisfies $(k^* \oplus l^*)\sigma = (\sigma|_U, \sigma|_V) = (0, 0)$. This means that the restrictions of σ to U and V are both zero. Since $\{U, V\}$ is an open cover of M, this implies that σ is zero.

To prove exactness at $\mathcal{A}^p(U) \oplus \mathcal{A}^p(V)$, first observe that

$$(i^* - j^*) \circ (k^* \oplus l^*)(\sigma) = (i^* - j^*)(\sigma|_U, \sigma|_V) = \sigma_{U \cap V} - \sigma|_{U \cap V} = 0,$$

which shows that $\mathrm{Im}(k^* \oplus l^*) \subset \mathrm{Ker}(i^* - j^*)$. Conversely, suppose $(\eta, \eta') \in \mathcal{A}^p(U) \oplus \mathcal{A}^p(V)$ and $(i^* - j^*)(\eta, \eta') = 0$. This means $\eta|_{U\cap V} = \eta'|_{U\cap V}$, so there is a global smooth p-form σ on M defined by

$$\sigma = \begin{cases} \eta & \text{on } U, \\ \eta' & \text{on } V. \end{cases}$$

Clearly $(\eta, \eta') = (k^* \oplus l^*)\sigma$, so $\mathrm{Ker}(i^* - j^*) \subset \mathrm{Im}(k^* \oplus l^*)$.

Exactness at $\mathcal{A}^p(U \cap V)$ means that $i^* - j^*$ is surjective. This the only nontrivial part of the proof, and the only part that really uses any properties of smooth manifolds.

Let $\omega \in \mathcal{A}^p(U \cap V)$ be arbitrary. We need to show that there exist $\eta \in \mathcal{A}^p(U)$ and $\eta' \in \mathcal{A}^p(V)$ such that

$$\omega = (i^* - j^*)(\eta, \eta') = i^*\eta - j^*\eta' = \eta|_{U\cap V} - \eta'|_{U\cap V}.$$

Let $\{\varphi, \psi\}$ be a partition of unity subordinate to the open cover $\{U, V\}$, and define $\eta \in \mathcal{A}^p(V)$ by

$$\eta = \begin{cases} \psi\omega, & \text{on } U \cap V, \\ 0 & \text{on } U \smallsetminus \operatorname{supp}\psi. \end{cases} \tag{11.11}$$

On the set $(U \cap V) \smallsetminus \operatorname{supp}\psi$ where these definitions overlap, they both give zero, so this defines η as a smooth p-form on U. Similarly, define $\eta' \in \mathcal{A}^p(V)$ by

$$\eta' = \begin{cases} -\varphi\omega, & \text{on } U \cap V, \\ 0 & \text{on } V \smallsetminus \operatorname{supp}\varphi. \end{cases} \tag{11.12}$$

Then we have

$$\eta|_{U\cap V} - \eta'|_{U\cap V} = \psi\omega - (-\varphi\omega) = (\psi + \varphi)\omega = \omega,$$

which was to be proved. □

For later use, we record the following corollary to the proof, which explicitly characterizes the connecting homomorphism δ.

Corollary 11.22. *The connecting homomorphism $\delta\colon H^p_{dR}(U \cap V) \to H^{p+1}_{dR}(M)$ is defined as follows. Given $\omega \in \mathcal{Z}^p(U \cap V)$, there are p-forms $\eta \in \mathcal{A}^p(U)$ and $\eta' \in \mathcal{A}^p(V)$ such that $\omega = \eta|_{U\cap V} - \eta'|_{U\cap V}$, and then $\delta[\omega] = [d\eta]$, where $d\eta$ is extended by zero to all of M.*

Proof. A characterization of the connecting homomorphism was given in the proof of the zigzag lemma. Specializing this characterization to the situation of the short exact sequence (11.8), we find that $\delta[\omega] = [\sigma]$ provided there exist $(\eta, \eta') \in \mathcal{A}^p(U) \oplus \mathcal{A}^p(V)$ such that

$$\begin{aligned} i^*\eta - j^*\eta' &= \omega, \\ (k^*\sigma, l^*\sigma) &= (d\eta, d\eta'). \end{aligned} \tag{11.13}$$

Arguing just as in the proof of the Mayer–Vietoris theorem, if $\{\varphi, \psi\}$ is a partition of unity subordinate to $\{U, V\}$, then formulas (11.11) and (11.12) define smooth forms $\eta \in \mathcal{A}^p(U)$ and $\eta' \in \mathcal{A}^p(V)$ satisfying the first equation of (11.13). Let σ be the form on M obtained by extending $d\eta$ to be zero outside of $U \cap V$. Because ω is closed,

$$\sigma|_{U\cap V} = d\eta|_{U\cap V} = d(\omega + \eta')|_{U\cap V} = d\eta'|_{U\cap V},$$

and the second equation of (11.13) follows easily. □

Our first application of the Mayer–Vietoris theorem will be to compute all of the de Rham cohomology groups of spheres. In the last section of this chapter, we will use the theorem again as an essential ingredient in the proof of the de Rham theorem.

Theorem 11.23. *For $n \geq 1$, the de Rham cohomology groups of $\mathbb{S}^n$ are*

$$H^p_{dR}(\mathbb{S}^n) \cong \begin{cases} \mathbb{R} & \text{if } p = 0 \text{ or } p = n, \\ 0 & \text{if } 0 < p < n. \end{cases}$$

Proof. We already know $H^0_{dR}(\mathbb{S}^n)$ and $H^n_{dR}(\mathbb{S}^n)$ from the preceding section. For good measure, we give here another proof for $H^n_{dR}(\mathbb{S}^n)$.

Let N and S be the north and south poles in $\mathbb{S}^n$, respectively, and let $U = \mathbb{S}^n \smallsetminus \{S\}$, $V = \mathbb{S}^n \smallsetminus \{N\}$. By stereographic projection, both U and V are diffeomorphic to $\mathbb{R}^n$, and thus $U \cap V$ is diffeomorphic to $\mathbb{R}^n \smallsetminus \{0\}$.

Part of the Mayer–Vietoris sequence for $\{U, V\}$ reads

$$H^{p-1}_{dR}(U) \oplus H^{p-1}_{dR}(V) \to H^{p-1}_{dR}(U \cap V) \to H^p_{dR}(\mathbb{S}^n) \to H^p_{dR}(U) \oplus H^p_{dR}(V).$$

Because U and V are diffeomorphic to $\mathbb{R}^n$, the groups on both ends are trivial when $p > 1$, which implies that $H^p_{dR}(\mathbb{S}^n) \cong H^{p-1}_{dR}(U \cap V)$. Moreover, $U \cap V$ is diffeomorphic to $\mathbb{R}^n \smallsetminus \{0\}$ and therefore homotopy equivalent to $\mathbb{S}^{n-1}$, so in the end we conclude that

$$H^p_{dR}(\mathbb{S}^n) \cong H^{p-1}_{dR}(\mathbb{S}^{n-1}) \quad \text{for } p > 1.$$

We will prove the theorem by induction on n. The case $n = 1$ is taken care of by Corollary 11.16, so suppose $n \geq 2$ and assume the theorem is true for $\mathbb{S}^{n-1}$. Clearly $H^0_{dR}(\mathbb{S}^n) \cong \mathbb{R}$ by Proposition 11.9, and $H^1_{dR}(\mathbb{S}^n) = 0$

because $\mathbb{S}^n$ is simply connected. For $p > 1$, the inductive hypothesis then gives

$$H^p_{dR}(\mathbb{S}^n) \cong H^{p-1}_{dR}(\mathbb{S}^{n-1}) \cong \begin{cases} 0 & \text{if } p < n, \\ \mathbb{R} & \text{if } p = n. \end{cases}$$

This completes the proof. □

Singular Homology and Cohomology

The topological invariance of the de Rham groups suggests that there should be some purely topological way of computing them. There is indeed, and the connection between the de Rham groups and topology was first proved by de Rham himself in the 1930s. The theorem that bears his name is a major landmark in the development of smooth manifold theory. We will give a proof in the next section.

In the category of topological spaces, there are a number of ways of defining cohomology groups that measure the existence of "holes" in different dimensions, but that have nothing to do with differential forms or smooth structures. In this section, we describe the most straightforward one, called singular cohomology. Because a complete treatment of singular cohomology would be far beyond the scope of this book, we can only summarize the basic ideas here. For more details, you can consult a standard textbook on algebraic topology, such as [Bre93, Mun84, Spa89]. (See also [Lee00, Chapter 13] for a more concise treatment.)

Suppose $v_0, \dots, v_p$ are any $p+1$ points in some Euclidean space $\mathbb{R}^n$. They are said to be in *general position* if they are not contained in any $(p-1)$-dimensional affine subspace. A *geometric p-simplex* is a subset of $\mathbb{R}^n$ of the form

$$\left\{ \sum_{i=0} t_i v_i : 0 \le t_i \le 1 \text{ and } \sum_{i=0}^{k} t_i = 1 \right\},$$

for some $(p+1)$-tuple $(v_0, \dots, v_p)$ in general position. The points v_i are called the *vertices* of the simplex, and the geometric simplex with vertices $v_0, \dots, v_p$ is denoted by $\langle v_0, \dots, v_p \rangle$. It is a compact convex set, in fact the smallest convex set containing $\{v_0, \dots, v_p\}$. The *standard p-simplex* is the simplex $\Delta_p = \langle e_0, e_1, \dots, e_p \rangle \subset \mathbb{R}^p$, where $e_0 = 0$ and e_i is the ith standard basis vector. For example, $\Delta_0 = \{0\}$, $\Delta_1 = [0,1]$, and Δ_2 is the triangle with vertices $(0,0)$, $(1,0)$, and $(0,1)$.

Exercise 11.4. Show that a geometric p-simplex is a p-dimensional smooth manifold with corners smoothly embedded in $\mathbb{R}^n$.

Let M be a topological space. A continuous map $\sigma\colon \Delta_p \to M$ is called a *singular p-simplex* in M. The *singular chain group* of M in dimension p, denoted by $C_p(M)$, is the free abelian group generated by all singular p-simplices in M. An element of this group, called a *singular p-chain*, is just a finite formal linear combination of singular p-simplices with integer coefficients.

One special case that arises frequently is that in which the space M is a convex subset of some Euclidean space $\mathbb{R}^m$. In that case, for any ordered $(p+1)$-tuple of points $(w_0, \dots, w_p)$ in M (not necessarily in general position), there is a unique affine map from $\mathbb{R}^p$ to $\mathbb{R}^m$ that takes e_i to w_i for $i = 0, \dots, p$. The restriction of this affine map to Δ_p is denoted by $\alpha(w_0, \dots, w_p)$, and is called an *affine singular simplex* in M.

For each $i = 0, \dots, p$, we define the ith *face map* in Δ_p to be the affine singular $(p-1)$-simplex $F_{i,p}\colon \Delta_{p-1} \to \Delta_p$ defined by

$$F_{i,p} = \alpha(e_0, \dots, \widehat{e_i}, \dots, e_p).$$

(As usual, the hat indicates that e_i is omitted.) It maps Δ_{p-1} homeomorphically onto the $(p-1)$-dimensional boundary face of Δ_p opposite e_i. The *boundary* of a singular p-simplex $\sigma\colon \Delta_p \to M$ is the singular $(p-1)$-chain $\partial\sigma$ defined by

$$\partial\sigma = \sum_{i=0}^{p} (-1)^i \sigma \circ F_{i,p}.$$

This extends uniquely to a group homomorphism $\partial\colon C_p(M) \to C_{p-1}(M)$, called the *singular boundary operator*. The basic fact about the boundary operator is the following lemma.

Lemma 11.24. *If c is any singular chain, then $\partial(\partial c) = 0$.*

Proof. The starting point is the fact that

$$F_{i,p} \circ F_{j,p-1} = F_{j,p} \circ F_{i-1,p-1} \tag{11.14}$$

when $i > j$, which can be verified by following what both compositions do to each of the vertices of Δ_{p-2}. Using this, the proof of the lemma is just a straightforward computation. □

A singular p-chain c is called a *cycle* if $\partial c = 0$, and a *boundary* if $c = \partial b$ for some singular $(p+1)$-chain b. Let $Z_p(M)$ denote the set of singular p-cycles in M, and $B_p(M)$ the set of singular p-boundaries. Because ∂ is a homomorphism, $Z_p(M)$ and $B_p(M)$ are subgroups of $C_p(M)$, and because $\partial \circ \partial = 0$, $B_p(M) \subset Z_p(M)$. The pth *singular homology group* of M is the quotient group

$$H_p(M) = \frac{Z_p(M)}{B_p(M)}.$$

To put it another way, the sequence of abelian groups and homomorphisms

$$\cdots \to C_{p+1}(M) \xrightarrow{\partial} C_p(M) \xrightarrow{\partial} C_{p-1}(M) \to \cdots$$

is a complex, called the *singular chain complex*, and $H_p(M)$ is the homology of this complex.

Any continuous map $F\colon M \to N$ induces a homomorphism $F_\#\colon C_p(M) \to C_p(N)$ on each singular chain group, defined by $F_\#(\sigma) = \sigma \circ F$ for any singular simplex σ and extended by linearity to chains. An easy computation shows that $F \circ \partial = \partial \circ F$, so F is a chain map, and therefore induces a homomorphism on singular homology, denoted by $F_*\colon H_p(M) \to H_p(N)$. It is immediate that $(G \circ F)_* = G_* \circ F_*$ and $(Id_M)_* = \mathrm{Id}_{H_p(M)}$, so singular homology defines a covariant functor from the category of topological spaces and continuous maps to the category of abelian groups and homomorphisms. In particular, homeomorphic spaces have isomorphic singular homology groups.

Proposition 11.25 (Properties of Singular Homology).

(*a*) *For any one-point space* $\{q\}$, $H_0(\{q\})$ *is the infinite cyclic group generated by the cohomology class of the unique singular* 0*-simplex mapping* Δ_0 *to* q, *and* $H_p(\{q\}) = 0$ *for all* $p \neq 0$.

(*b*) *Let* $\{M_j\}$ *be any collection of topological spaces, and let* $M = \coprod_j M_j$. *The inclusion maps* $\iota_j\colon M_j \hookrightarrow M$ *induce an isomorphism from* $\oplus_j H_p(M_j)$ *to* $H_p(M)$.

(*c*) *Homotopy equivalent spaces have isomorphic singular homology groups.*

Sketch of Proof. In a one-point space $\{q\}$, there is exactly one singular p-simplex for each p, namely the constant map. The result of part (a) follows from an analysis of the boundary maps. Part (b) is immediate because the maps ι_j already induce an isomorphism on the chain level: $\oplus_j C_p(M_j) \cong C_p(M)$.

The main step in the proof of homotopy invariance is the construction for any space M of a linear map $h\colon C_p(M) \to C_{p+1}(M \times I)$ satisfying

$$h \circ \partial + \partial \circ h = (i_1)_\# - (i_0)_\#, \tag{11.15}$$

where $\iota_j\colon M \to M \times I$ is the injection $\iota_j(x) = (x, j)$. From this it follows just as in the proof of Proposition 11.5 that homotopic maps induce the same homology homomorphism, and then in turn that homotopy equivalent spaces have isomorphic singular cohomology groups. □

In addition to the properties above, singular homology satisfies the following version of the Mayer–Vietoris theorem. Suppose M is a topological space and $U, V \subset M$ are open subsets whose union is M. The usual diagram (11.7) of inclusions induces homology homomorphisms:

$$\begin{tikzcd} & H_p(U) \arrow[dr, "k_*"] & \\ H_p(U \cap V) \arrow[ur, "i_*"] \arrow[dr, "j_*"'] & & H_p(M). \\ & H_p(V) \arrow[ur, "l_*"'] & \end{tikzcd} \tag{11.16}$$

Theorem 11.26 (Mayer–Vietoris for Singular Homology). *Let M be a topological space and let U, V be open subsets of M whose union is M. For each p there is a homomorphism $\partial_* \colon H_p(M) \to H_{p-1}(U \cap V)$ such that the following sequence is exact:*

$$\begin{aligned} \cdots \xrightarrow{\partial_*} H_p(U \cap V) \xrightarrow{\alpha} H_p(U) \oplus H_p(V) \xrightarrow{\beta} H_p(M) & \\ \xrightarrow{\partial_*} H_{p-1}(U \cap V) \xrightarrow{\alpha} \cdots, & \end{aligned} \tag{11.17}$$

where

$$\begin{aligned} \alpha[c] &= (i_*[c], -j_*[c]), \\ \beta([c], [c']) &= k_*[c] + l_*[c'], \end{aligned}$$

and $\partial_*[e] = [c]$ provided there exist $d \in H_p(U)$ and $d' \in H_p(V)$ such that $k_*d + l_*d'$ is homologous to e and

$$(i_*c, -j_*c) = (\partial d, \partial d').$$

Sketch of Proof. The basic idea, of course, is to construct a short exact sequence of complexes and use the zigzag lemma. The hardest part of the proof is showing that any homology class $[e] \in H_p(M)$ can be represented in the form $[k_*d - l_*d']$, where d and d' are chains in U and V, respectively. □

Note that the maps α and β in this Mayer–Vietoris sequence can be replaced by

$$\begin{aligned} \widetilde{\alpha}[c] &= (i_*[c], j_*[c]), \\ \widetilde{\beta}([c], [c']) &= k_*[c] - l_*[c'], \end{aligned}$$

and the same proof goes through. We have chosen the definition given in the statement of the theorem because it leads to a cohomology exact sequence that is compatible with the Mayer–Vietoris sequence for de Rham cohomology; see the proof of the de Rham theorem below.

Singular Cohomology

In addition to the singular homology groups, for any abelian group G one can define a closely related sequence of groups $H^p(M;G)$ called the *singular cohomology groups* with coefficients in G. The precise definition is unimportant for our purposes; we will only be concerned with the special case $G=\mathbb{R}$, in which case it can be shown that $H^p(M;\mathbb{R})$ is a real vector space that is isomorphic to the space $\operatorname{Hom}(H_p(M),\mathbb{R})$ of group homomorphisms from $H_p(M)$ into $\mathbb{R}$. (If you like, you can take this as a definition of $H^p(M,\mathbb{R})$.) Any continuous map $F\colon M\to N$ induces a linear map $F^*\colon H^p(N;\mathbb{R})\to H^p(M;\mathbb{R})$ by $(F^*\gamma)[c]=\gamma(F_*[c])$ for any $\gamma\in H^p(N;\mathbb{R})\cong\operatorname{Hom}(H_p(M),\mathbb{R})$ and any singular p-chain c. The functorial properties of F_* carry over to cohomology: $(G\circ F)^*=F^*\circ G^*$ and $(\mathrm{Id}_M)^*=\mathrm{Id}_{H^p(M;\mathbb{R})}$.

The following properties of the singular cohomology groups follow easily from the definitions and Proposition 11.25.

Proposition 11.27 (Properties of Singular Cohomology).

(*a*) *For any one-point space $\{q\}$, $H^p(\{q\};\mathbb{R})$ is trivial except when $p=0$, in which case it is one-dimensional.*

(*b*) *If $\{M_j\}$ is any collection of topological spaces and $M=\coprod_j M_j$, then $H^p(M;\mathbb{R})\cong\prod_j H^p(M_j;\mathbb{R})$.*

(*c*) *Homotopy equivalent spaces have isomorphic singular homology groups.*

The key fact about the singular cohomology groups that we will need is that they too satisfy a Mayer–Vietoris theorem.

Theorem 11.28 (Mayer–Vietoris for Singular Cohomology). *Suppose M, U, and V satisfy the hypotheses of Theorem 11.26. For each p there is a homomorphism $\partial^*\colon H^p(U\cap V;\mathbb{R})\to H^{p+1}(M;\mathbb{R})$ such that the following sequence is exact:*

$$\cdots\xrightarrow{\delta}H^p(M;\mathbb{R})\xrightarrow{k^*\oplus l^*}H^p(U;\mathbb{R})\oplus H^p(V;\mathbb{R})\xrightarrow{i^*-j^*}H^p(U\cap V;\mathbb{R})$$
$$\xrightarrow{\delta}H^{p+1}(M;\mathbb{R})\xrightarrow{k^*\oplus l^*}\cdots,\quad(11.18)$$

where the maps $k^*\oplus l^*$ and i^*-j^* are defined as in (11.9)*, and $\partial^*\gamma=\gamma\circ\partial_*$, with ∂_* as in Theorem 11.26.*

Sketch of Proof. For any homomorphism $F\colon A\to B$ between abelian groups, there is a *dual homomorphism* $F^*\colon\operatorname{Hom}(B,\mathbb{R})\to\operatorname{Hom}(A,\mathbb{R})$ given by $F^*\gamma=\gamma\circ F$. Applying this to the Mayer–Vietoris sequence (11.17) for singular homology, we obtain the cohomology sequence (11.18). The exactness of the resulting sequence is a consequence of the fact that the functor

$A \mapsto \operatorname{Hom}(A, \mathbb{R})$ is *exact*, meaning that it takes exact sequences to exact sequences. This in turn follows from the fact that $\mathbb{R}$ is an *injective group*: Whenever H is a subgroup of an abelian group G, every homomorphism from H into $\mathbb{R}$ extends to all of G. □

Smooth Singular Homology

The connection between singular and de Rham cohomology will be established by integrating differential forms over singular chains. More precisely, given a singular p-simplex σ in a manifold M and a p-form on M, we would like to pull ω back by σ and integrate the resulting form over Δ_p. However, there is an immediate problem with this approach, because forms can only be pulled back by *smooth* maps, while singular simplices are in general only continuous. (Actually, since only first derivatives of the map appear in the formula for the pullback, it would be sufficient to consider C^1 maps, but merely continuous ones definitely will not do.) In this section we overcome this problem by showing that singular homology can be computed equally well with smooth simplices.

If M is a smooth manifold, a *smooth p-simplex* in M is a smooth map $\sigma\colon \Delta_p \to M$. The subgroup of $C_p(M)$ generated by smooth simplices is denoted by $C_p^\infty(M)$ and called the *smooth chain group* in dimension p; elements of this group, which are finite formal linear combinations of smooth simplices, are called *smooth chains*. Because the boundary of a smooth simplex is a smooth chain, we can define the pth *smooth singular homology group* of M to be the quotient group

$$H_p^\infty(M) = \frac{\operatorname{Ker}[\partial\colon C_p^\infty(M) \to C_{p-1}^\infty(M)]}{\operatorname{Im}[\partial\colon C_{p+1}^\infty(M) \to C_p^\infty(M)]}.$$

The inclusion map $\iota\colon C_p^\infty(M) \hookrightarrow C_p(M)$ obviously commutes with the boundary operator, and so induces a map on homology: $\iota_*\colon H_p^\infty(M) \to H_p(M)$.

Theorem 11.29. *For any smooth manifold M, the map $\iota_*\colon H_p^\infty(M) \to H_p(M)$ induced by inclusion is an isomorphism.*

Proof. [*Author's note: This proof still needs to be written. The basic idea is to construct, with the help of the Whitney approximation theorem, a smoothing operator $s\colon C_p(M) \to C_p^\infty(M)$ such that $s \circ \partial = \partial \circ s$ and $s \circ \iota$ is the identity on $C_p^\infty(M)$, and a homotopy operator that shows that $\iota \circ s$ induces the identity map on $H_p(M)$.*] □

The de Rham Theorem

In this section we will state and prove the de Rham theorem. Before getting to the theorem itself, we need one more algebraic lemma. Its proof is another diagram chase like the proof of the zigzag lemma.

Lemma 11.30 (The Five Lemma). *Consider the following commutative diagram of modules and linear maps:*

$$\begin{tikzcd} A_1 \arrow[r, "\alpha_1"] \arrow[d, "f_1"] & A_2 \arrow[r, "\alpha_2"] \arrow[d, "f_2"] & A_3 \arrow[r, "\alpha_3"] \arrow[d, "f_3"] & A_4 \arrow[r, "\alpha_4"] \arrow[d, "f_4"] & A_5 \arrow[d, "f_5"] \\ B_1 \arrow[r, "\beta_1"] & B_2 \arrow[r, "\beta_2"] & B_3 \arrow[r, "\beta_3"] & B_4 \arrow[r, "\beta_4"] & B_5. \end{tikzcd}$$

If the horizontal rows are exact and f_1, f_2, f_4, and f_5 are isomorphisms, then f_3 is also an isomorphism.

Exercise 11.5. Prove (or look up) this lemma.

Suppose M is a smooth manifold, ω is a closed p-form on M, and σ is a smooth p-simplex in M. We define the *integral of ω over σ* to be

$$\int_\sigma \omega = \int_{\Delta_p} \sigma^*\omega.$$

This makes sense because Δ_p is a smooth p-submanifold with corners embedded in $\mathbb{R}^p$, and inherits the orientation of $\mathbb{R}^p$. (Or we could just consider Δ_p as a domain of integration in $\mathbb{R}^p$.) (When $p = 1$, this is the same as the line integral of ω over the smooth curve segment $\sigma\colon [0,1] \to M$.) If $c = \sum_{i=1}^k c_i\sigma_i$ is a smooth p-chain, the integral of ω over c is defined as

$$\int_c \omega = \sum_{i=1}^k c_i \int_{\sigma_i} \omega.$$

Theorem 11.31 (Stokes's Theorem for Chains). *If c is a smooth p-chain in a smooth manifold M, and ω is a $(p-1)$-form on M, then*

$$\int_{\partial c} \omega = \int_c d\omega.$$

Proof. It suffices to prove the theorem when c is just a smooth simplex σ. Since Δ_p is a manifold with corners, Stokes's theorem says

$$\int_\sigma d\omega = \int_{\Delta_p} \sigma^* d\omega = \int_{\Delta_p} d\sigma^*\omega = \int_{\partial\Delta_p} \sigma^*\omega.$$

The maps $\{F_{i,p} : 0 = 1, \dots, p\}$ are parametrizations of the boundary faces of Δ_p satisfying the conditions of Proposition 10.30, except possibly that they might not be orientation preserving. To check the orientations, note that $F_{i,p}$ is the restriction to $\Delta_p \cap \partial\mathbb{H}^p$ of the affine diffeomorphism sending $\langle e_0, \ldots, e_p\rangle$ to $\langle e_0, \ldots, \widehat{e_i}, \ldots, e_p, e_i\rangle$. This is easily seen to be orientation preserving if and only if $(e_0, \ldots, \widehat{e_i}, \ldots, e_p, e_i)$ is an even permutation of $(e_0, \ldots, e_p)$, which is the case if and only if $p-i$ is even. Since the standard coordinates on $\partial\mathbb{H}^p$ are positively oriented if and only if p is even, the upshot is that $F_{i,p}$ is orientation preserving for $\partial\Delta_p$ if and only if i is even. Thus, by Proposition 10.30.

$$\begin{aligned}
\int_{\partial\Delta_p} \sigma^*\omega &= \sum_{i=0}^{p} (-1)^i \int_{\Delta_{p-1}} F_{i,p}^* \sigma^*\omega \\
&= \sum_{i=0}^{p} (-1)^i \int_{\Delta_{p-1}} (\sigma \circ F_{i,p})^*\omega \\
&= \sum_{i=0}^{p} (-1)^i \int_{\sigma \circ F_{i,p}} \omega.
\end{aligned}$$

By definition of the singular boundary operator, this is equal to $\int_{\partial\sigma} \omega$. □

Using this theorem, we can define a natural linear map $\mathcal{I}\colon H^p_{dR}(M) \to H^p(M;\mathbb{R})$, called the *de Rham homomorphism*, as follows. For any $[\omega] \in H^p_{dR}(M)$ and $[c] \in H_p(M) \cong H^\infty_p(M)$, we define

$$\mathcal{I}[\omega][c] = \int_{\widetilde{c}} \omega, \tag{11.19}$$

where $\widetilde{c}$ is any smooth p-cycle representing the homology class $[c]$. This is well-defined, because if $\widetilde{c}$ is the boundary of a smooth $(p-1)$-chain $\widetilde{b}$, then

$$\int_{\widetilde{c}} \omega = \int_{\partial\widetilde{b}} \omega = \int_{\widetilde{b}} d\omega = 0,$$

while if $\omega = d\eta$ is exact, then

$$\int_{\widetilde{c}} \omega = \int_{\widetilde{c}} d\eta = \int_{\partial\widetilde{c}} \omega = 0.$$

(Note that $\partial\widetilde{c} = 0$ and $d\omega = 0$ because they represent a homology class and a cohomology class, respectively.) Clearly $\mathcal{I}[\omega][c+c'] = \mathcal{I}[\omega][c] + \mathcal{I}[\omega][c']$, and the resulting homomorphism $\mathcal{I}[\omega]\colon H_p(M) \to \mathbb{R}$ depends linearly on ω. Thus $\mathcal{I}[\omega]$ is a well-defined element of $\mathrm{Hom}(H_p(M), \mathbb{R})$, which we identify with $H^p(M;\mathbb{R})$.

Lemma 11.32. *If $F\colon M \to N$ is a smooth map, then the following diagram commutes:*

$$\begin{tikzcd} H^p_{dR}(N) \arrow[r, "F^*"] \arrow[d, "\mathcal{I}"'] & H^p_{dR}(M) \arrow[d, "\mathcal{I}"] \\ H^p(N;\mathbb{R}) \arrow[r, "F^*"'] & H^p(M;\mathbb{R}). \end{tikzcd}$$

Proof. Directly from the definitions, if σ is a smooth p-simplex in M and ω is a p-form on N,

$$\int_\sigma F^*\omega = \int_{\Delta_p} \sigma^* F^*\omega = \int_{\Delta_p} (F\circ\sigma)^*\omega = \int_{F\circ\sigma} \omega.$$

This is equivalent to $\mathcal{I}(F^*[\omega])[\sigma] = \mathcal{I}[\omega](F_*[\sigma])$. □

Lemma 11.33. *If M is a smooth manifold and U, V are open subsets of M whose union is M, then the following diagram commutes:*

$$\begin{tikzcd} H^{p-1}_{dR}(U\cap V) \arrow[r, "\delta"] \arrow[d, "\mathcal{I}"'] & H^p_{dR}(M) \arrow[d, "\mathcal{I}"] \\ H^{p-1}(U\cap V;\mathbb{R}) \arrow[r, "\partial^*"'] & H^p(M;\mathbb{R}), \end{tikzcd} \tag{11.20}$$

where δ and ∂^* are the connecting homomorphisms of the Mayer–Vietoris sequences for de Rham and singular cohomology, respectively.

Proof. Identifying $H^p(M;\mathbb{R})$ with $\operatorname{Hom}(H_p(M),\mathbb{R})$ as usual, commutativity of (11.20) reduces to the following equation for any $[\omega] \in H^{p-1}_{dR}(U\cap V)$ and any $[e] \in H_p(M)$:

$$\mathcal{I}(\delta[\omega])[e] = (\partial^*\mathcal{I}[\omega])[e] = \mathcal{I}[\omega](\partial_*[e]).$$

If σ is a $(p-1)$-form representing $\delta[\omega]$ and $[c]$ is a p-chain representing $\partial_*[e]$, this is the same as

$$\int_e \sigma = \int_c \omega.$$

By the characterizations of δ and ∂_* given in Corollary 11.22 and Theorem 11.26, we can choose $\sigma = d\eta$ (extended by zero to all of M), where $\eta \in \mathcal{A}^p(U)$ and $\eta' \in \mathcal{A}^p(U)$ are forms such that $\omega = \eta|_{U\cap V} - \eta'|_{U\cap V}$; and $c = \partial d$, where d, d' are smooth simplices in U and V, respectively, such that $d + d'$ represents the same homology class as e. Then, because $\partial d + \partial d' = \partial e = 0$

and $d\eta|_{U\cap V} - d\eta'|_{U\cap V} = d\omega = 0$, we have

$$\begin{aligned}\int_c \omega &= \int_{\partial d} \omega \\ &= \int_{\partial d} \eta - \int_{\partial d} \eta' \\ &= \int_{\partial d} \eta + \int_{\partial d'} \eta' \\ &= \int_d d\eta + \int_{d'} d\eta' \\ &= \int_d \sigma + \int_{d'} \sigma \\ &= \int_e \sigma.\end{aligned}$$

Thus the diagram commutes. □

Theorem 11.34 (de Rham). *For any smooth manifold M and any $p \geq 0$, the de Rham homomorphism $\mathcal{I}\colon H^p_{dR}(M) \to H^p(M;\mathbb{R})$ is an isomorphism.*

Proof. Let us say that a smooth manifold M is a *de Rham manifold* if the de Rham homomorphism $\mathcal{I}\colon H^p_{dR}(M) \to H^p(M;\mathbb{R})$ is an isomorphism for each p. Since the de Rham homomorphism commutes with the cohomology maps induced by smooth maps (Lemma 11.32), any manifold that is diffeomorphic to a de Rham manifold is also de Rham. The theorem will be proved once we show that every smooth manifold is de Rham.

If M is any smooth manifold, an open cover $\{U_i\}$ of M is called a *de Rham cover* if each open set U_i is a de Rham manifold, and every finite intersection $U_{i_1} \cap \cdots \cap U_{i_k}$ is de Rham. A de Rham cover that is also a basis for the topology of M is called a *de Rham basis* for M.

Step 1: *If $\{M_j\}$ is any countable collection of de Rham manifolds, then their disjoint union is de Rham.* By Propositions 11.8 and 11.27(c), for both de Rham and singular cohomology, the inclusions $\iota_j\colon M_j \hookrightarrow \coprod_j M_j$ induce isomorphisms between the cohomology groups of the disjoint union and the direct product of the cohomology groups of the manifolds M_j. By Lemma 11.32, $\mathcal{I}$ respects these isomorphisms.

Step 2: *Every convex open subset of $\mathbb{R}^n$ is de Rham.* Let U be such a subset. By the Poincaré lemma, $H^p_{dR}(U)$ is trivial when $p \neq 0$. Since U is homotopy equivalent to a one-point space, Proposition 11.27 implies that the singular cohomology groups of U are also trivial for $p \neq 0$. In the $p = 0$ case, $H^0_{dR}(U)$ is the one-dimensional space consisting of the constant functions, and $H^0(U;\mathbb{R}) = \operatorname{Hom}(H_0(U),\mathbb{R})$ is also one-dimensional because $H_0(U)$ is spanned by any singular 0-simplex. If $\sigma\colon \Delta_0 \to M$ is a 0-simplex (which is smooth because any map from a 0-manifold is smooth), and f is the constant function equal to 1, then

$$\mathcal{I}[f][\sigma] = \int_{\Delta_0} \sigma^* f = (f \circ \sigma)(0) = 1.$$

This shows that $\mathcal{I}\colon H^0_{dR}(U) \to H^0(U;\mathbb{R})$ is not the zero map, so it is an isomorphism.

Step 3: *If M has a finite de Rham cover, then M is de Rham.* Suppose $M = U_1 \cup \cdots \cup U_k$, where the open sets U_i and their finite intersections are de Rham. We will prove the result by induction on k. Suppose first that M has a de Rham cover consisting of two sets $\{U, V\}$. Putting together the Mayer–Vietoris sequences for de Rham and singular cohomology, we obtain the following commutative diagram in which the horizontal rows are exact and the vertical maps are all given by de Rham homomorphisms:

$$\begin{tikzcd}
H^{p-1}_{dR}(U) \oplus H^{p-1}_{dR}(V) \arrow[r] \arrow[d] & H^{p-1}_{dR}(U \cap V) \arrow[r] \arrow[d] & H^p_{dR}(M) \arrow[r] \arrow[d] & {} \\
H^{p-1}(U;\mathbb{R}) \oplus H^{p-1}(V;\mathbb{R}) \arrow[r] & H^{p-1}(U \cap V;\mathbb{R}) \arrow[r] & H^p(M;\mathbb{R}) \arrow[r] & {}
\end{tikzcd}$$

$$\begin{tikzcd}
H^p_{dR}(U) \oplus H^p_{dR}(V) \arrow[r] \arrow[d] & H^p_{dR}(U \cap V) \arrow[d] \\
H^p(U;\mathbb{R}) \oplus H^p(V;\mathbb{R}) \arrow[r] & H^p(U \cap V;\mathbb{R}).
\end{tikzcd}$$

The commutativity of the diagram is an immediate consequence of Lemmas 11.32 and 11.33. By hypothesis the first, second, fourth, and fifth vertical maps are all isomorphisms, so by the five lemma the middle map is an isomorphism, which proves that M is de Rham.

Now assume the claim is true for smooth manifolds admitting a de Rham cover with $k \geq 2$ sets, and suppose $\{U_1, \ldots, U_{k+1}\}$ is a de Rham cover of M. Put $U = U_1 \cup \cdots \cup U_k$ and $V = U_{k+1}$. The hypothesis implies that U and V are de Rham, and so is $U \cap V$ because it has a k-fold de Rham cover given by $\{U_1 \cap U_{k+1}, \ldots, U_k \cap U_{k+1}\}$. Thus $M = U \cup V$ is also de Rham by the argument above.

Step 4: *If M has a de Rham basis, then M is de Rham.* Suppose $\{U_\alpha\}$ is a de Rham basis for M. Let $f\colon M \to \mathbb{R}$ be a continuous proper function, such as the one constructed in the proof of the Whitney embedding theorem (Theorem 6.12). For each integer m, define subsets A_m and A'_m of M by

$$\begin{aligned}
A_m &= \{q \in M : m \leq f(q) \leq m+1\},\\
A'_m &= \{q \in M : m - \tfrac{1}{2} < f(q) < m + 1 + \tfrac{1}{2}\}.
\end{aligned}$$

For each point $q \in A_m$, there is a basis open set containing q and contained in A'_m. The collection of all such basis sets is an open cover of A_m. Since f is proper, A_m is compact, and therefore it is covered by finitely many of these basis sets. Let B_m be the union of this finite collection of sets. This is a finite de Rham cover of B_m, so by Step 3, B_m is de Rham.

Observe that $B_m \subset A'_m$, so B_m intersects $B_{\widetilde{m}}$ nontrivially only when $\widetilde{m} = m-1$, m, or $m+1$. Therefore, if we define

$$U = \bigcup_{m \text{ odd}} B_m,$$

then U is de Rham by Step 1, because it is the disjoint union of the de Rham manifolds B_m. Similarly,

$$V = \bigcup_{m \text{ even}} B_m$$

is de Rham. Finally, $M = U \cup V$ is de Rham by Step 3.

Step 5: *Any open subset of* $\mathbb{R}^n$ *is de Rham.* If $U \subset \mathbb{R}^n$ is such a subset, then U has a basis consisting of Euclidean balls. Because each ball is convex, it is de Rham, and because any finite intersection of balls is again convex, finite intersections are also de Rham. Thus U has a de Rham basis, so it is de Rham by Step 4.

Step 6: *Every smooth manifold is de Rham.* Any smooth manifold has a basis of coordinate domains. Since every coordinate domain is diffeomorphic to an open subset of $\mathbb{R}^n$, as are finite intersections of coordinate domains, this is a de Rham basis. The claim therefore follows from Step 4. □

This result expresses a deep connection between the topological and analytic properties of a smooth manifold, and plays a central role in differential geometry. If one has some information about the topology of a manifold M, the de Rham theorem can be used to draw conclusions about solutions to differential equations such as $d\eta = \omega$ on M. Conversely, if one can prove that such solutions do or do not exist, then one can draw conclusions about the topology.

Problems

11-1. Let M be a smooth manifold, and $\omega \in \mathcal{A}^p(M)$, $\eta \in \mathcal{A}^q(M)$. Show that the de Rham cohomology class of $\omega \wedge \eta$ depends only on the de Rham cohomology classes of ω and η, and thus there is a well-defined bilinear map $\cup\colon H^p_{dR}(M) \times H^q_{dR}(M) \to H^{p+q}_{dR}(M)$ given by

$$[\omega] \cup [\eta] = [\omega \wedge \eta].$$

(This bilinear map is called the *cup product*.)

11-2. Let M be an orientable smooth manifold and suppose ω is a closed p-form on M.

(a) Show that ω is exact if and only if the integral of ω over every smooth p-cycle is zero.

(b) Now suppose that $H_p(M)$ is generated by the homology classes of finitely many smooth p-cycles $\{c_1, \dots, c_m\}$. The numbers $P_1(\omega), \dots, P_m(\omega)$ defined by

$$P_i = \int_{c_i} \omega$$

are called the *periods* of ω with respect to this set of generators. Show that ω is exact if and only if all of its periods are zero.

11-3. Let M be a smooth n-manifold and suppose $S \subset M$ is an immersed, compact, oriented, p-dimensional submanifold. A *smooth triangulation* of S is a smooth p-cycle $c = \sum_i n_i \sigma_i$ in M with the following properties:

- Each σ_i is an orientation-preserving embedding of Δ_p into S.
- If $i \neq j$, then $\sigma_i(\operatorname{Int}\Delta_p) \cap \sigma_j(\operatorname{Int}\Delta_p) = \varnothing$.
- $S = \bigcup_i \sigma_i(\Delta_p)$.

(It can be shown that every smooth orientable submanifold admits a smooth triangulation, but we will not use that fact.) Two p-dimensional submanifolds $S, S' \subset M$ are said to be *homologous* if there exist smooth triangulations c for S and c' for S' such that $c - c'$ is a boundary.

(a) If c is a smooth triangulation of S and ω is any p-form on M, show that $\int_c \omega = \int_S \omega$.

(b) If ω is closed and S, S' are homologous, show that $\int_S \omega = \int_{S'} \omega$.

11-4. Suppose (M, g) is a Riemannian n-manifold. A p-form ω on M is called a *calibration* if ω is closed and $\omega_q(X_1, \dots, X_p) \le 1$ whenever $(X_1, \dots, X_p)$ are orthonormal vectors in some tangent space T_qM. A smooth submanifold $S \subset M$ is said to be *calibrated* if there is a calibration ω such that $\omega|_S$ is the volume form for the induced Riemannian metric on S. If $S \subset M$ is a smoothly triangulated calibrated submanifold, show that the volume of S (with respect to the induced Riemannian metric) is less than or equal to that of any other submanifold homologous to S. (Calibrations were invented in 1985 by Reese Harvey and Blaine Lawson [HL82]; they have become increasingly important in recent years because in many situations a calibration is the only known way of proving that a given submanifold is volume minimizing in its homology class.)

11-5. Let $D \subset \mathbb{R}^3$ be the torus of revolution obtained by revolving the circle $(x-2)^2 + z^2 = 1$ around the z-axis, with the induced Riemannian metric. Show that the inner circle $\{(x, y, z) : z = 0, x^2 + y^2 = 1\}$ is calibrated, and therefore has the shortest length in its homology class.

11-6. Let M be a compact, connected, orientable, smooth n-manifold, and let p be any point of M. Let V be a neighborhood of p diffeomorphic to $\mathbb{R}^n$ and let $U = M \smallsetminus \{p\}$.

(a) Show that the connecting homomorphism $\delta\colon H^{n-1}_{dR}(U \cap V) \to H^n_{dR}(M)$ is an isomorphism. [Hint: Consider the $(n-1)$-form ω on $U \cap V \approx \mathbb{R}^n \smallsetminus \{0\}$ defined in coordinates by (10.18) (Problem 10-10).]

(b) Use the Mayer–Vietoris sequence of $\{U, V\}$ to show that $H^n_{dR}(M \smallsetminus \{p\}) = 0$.

11-7. Let M be a compact, connected, smooth manifold of dimension $n \ge 3$. For any $p \in M$ and $0 \le k < n$, show that the map $H^k_{dR}(M) \to H^k_{dR}(M \smallsetminus \{p\})$ induced by inclusion $M \smallsetminus \{p\} \hookrightarrow M$ is an isomorphism. [Hint: Use a Mayer–Vietoris sequence together with the result of Problem 11-6. The cases $k = 1$ and $k = n-1$ will require special handling.]

11-8. Let M_1, M_2 be smooth, connected, orientable manifolds of dimension $n \ge 2$, and let $M_1 \# M_2$ denote their smooth connected sum (see Problem 5-20). Show that $H^k_{dR}(M_1 \# M_2) \cong H^k_{dR}(M_1) \oplus H^k_{dR}(M_2)$ for $0 < k < n$.

11-9. Suppose (M, ω) is a $2n$-dimensional symplectic manifold.

(a) Show that $\omega^n = \omega \wedge \cdots \wedge \omega$ (the n-fold wedge product of ω with itself) is not exact. [Hint: See Problem 10-21.]

(b) Show that $H^{2k}_{dR}(M) \neq 0$ for $k = 1, \ldots, n$.

(c) Show that the only sphere that admits a symplectic structure is $\mathbb{S}^2$.

12
Integral Curves and Flows

In this chapter, we begin to explore vector fields in more depth. The primary objects associated with vector fields are "integral curves," which are smooth curves whose tangent vector at each point is equal to the value of the vector field there. We will show in this chapter that a vector field on a manifold determines a unique integral curve through each point; the proof is an application of the existence and uniqueness theorem for solutions of ordinary differential equations.

The collection of all integral curves of a given vector field on a manifold determines a family of diffeomorphisms of (open subsets of) the manifold, called a "flow." Any smooth $\mathbb{R}$-action is a flow, for example; but we will see that there are flows that are not $\mathbb{R}$-actions because the diffeomorphisms may not be defined for all $t \in \mathbb{R}$ or all points in the manifold.

In subsequent chapters, we will begin to study some of the profound applications of these ideas.

Integral Curves

A smooth curve $\gamma\colon J \to M$ determines a tangent vector $\gamma'(t) \in T_{\gamma(t)}M$ at each point of the curve. In this section we describe a way to work backwards: Given a tangent vector at each point, we seek a curve that has those tangent vectors.

Let M be a smooth manifold and let V be a smooth vector field on M. An *integral curve* of V is a smooth curve $\gamma\colon J \to M$ defined on an open interval $J \subset \mathbb{R}$ such that

$$\gamma'(t) = V_{\gamma(t)} \quad \text{for all } t \in J.$$

In other words, the tangent vector to γ at each point is equal to the value of V at that point. If $0 \in J$, the point $p = \gamma(0)$ is called the *starting point* of γ. (The reason for the term "integral curve" will be explained shortly.)

Example 12.1 (Integral Curves).

(a) Let $V = \partial/\partial x$ be the first coordinate vector field on $\mathbb{R}^2$. It is easy to check that any curve of the form $\gamma(t) = (t + a, b)$ for constants a and b is an integral curve of V, satisfying $\gamma(0) = (a, b)$. Thus there is an integral curve passing through each point of the plane.

(b) Let $W = x\,\partial/\partial x + y\,\partial/\partial y$ on $\mathbb{R}^2$. If $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ is a smooth curve, written in standard coordinates as $\gamma(t) = (x(t), y(t))$, then the condition $\gamma'(t) = W_{\gamma(t)}$ for γ to be an integral curve translates to

$$\begin{aligned} x'(t) \left.\frac{\partial}{\partial x}\right|_{(x(t),y(t))} &+ y'(t) \left.\frac{\partial}{\partial y}\right|_{(x(t),y(t))} \\ &= x(t) \left.\frac{\partial}{\partial x}\right|_{(x(t),y(t))} + y(t) \left.\frac{\partial}{\partial y}\right|_{(x(t),y(t))}. \end{aligned}$$

Comparing the components of these vectors, this is equivalent to the pair of ordinary differential equations

$$\begin{aligned} x'(t) &= x(t), \\ y'(t) &= y(t). \end{aligned}$$

These equations have the solutions $x(t) = ae^t$ and $y(t) = be^t$ for arbitrary constants a and b, and thus each curve of the form $\gamma(t) = (ae^t, be^t)$ is an integral curve of W. Since $\gamma(0) = (a, b)$, we see once again that there is an integral curve passing through each point $(a, b) \in \mathbb{R}^2$

As the second example above illustrates, finding integral curves boils down to solving a system of ordinary differential equations in a coordinate chart. More generally, let $\gamma\colon J \to M$ be any smooth curve. Writing γ in local coordinates as $\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t))$, the condition $\gamma'(t) = V_{\gamma(t)}$ that γ be an integral curve of a smooth vector field V can be written in local coordinates on an open set U as

$$(\gamma^i)'(t) \left.\frac{\partial}{\partial x^i}\right|_{\gamma(t)} = V^i(\gamma(t)) \left.\frac{\partial}{\partial x^i}\right|_{\gamma(t)},$$

which reduces to the system of ordinary differential equations (ODEs)

$$\begin{aligned}
(\gamma^1)'(t) &= V^1(\gamma^1(t),\dots,\gamma^n(t)),\\
&\dots\\
(\gamma^n)'(t) &= V^n(\gamma^1(t),\dots,\gamma^n(t)),
\end{aligned}$$

where the component functions V^i are smooth on U. The fundamental fact about such systems, which we will state precisely and prove later in the chapter, is that there is a unique solution, at least for t in a small time interval $(-\varepsilon,\varepsilon)$, satisfying any initial condition of the form $(\gamma^1(0),\dots,\gamma^n(0)) = (a^1,\dots,a^n)$ for $(a^1,\dots,a^n) \in U$. (This the reason for the terminology "integral curves," because solving a system of ODEs is often referred to as "integrating" the system.) For now, we just note that this implies there is a unique integral curve, at least for a short time, starting at any point in the manifold. Moreover, we will see that up to reparametrization, there is a unique integral curve passing through each point.

The following simple lemma shows how an integral curve can be reparametrized to change its starting point.

Lemma 12.2 (Translation Lemma). *Let V be a smooth vector field on a smooth manifold M, let $J \subset \mathbb{R}$ be an open interval containing 0, and let $\gamma\colon J \to M$ be an integral curve of V. For any $a \in J$, let $\widetilde{J} = \{t \in \mathbb{R} : t + a \in J\}$. Then the curve $\widetilde{\gamma}\colon \widetilde{J} \to M$ defined by $\widetilde{\gamma}(t) = \gamma(t+a)$ is an integral curve of V starting at $\gamma(a)$.*

Proof. One way to see this is as a straightforward application of the chain rule in local coordinates. Somewhat more invariantly, we can examine the action of $\widetilde{\gamma}'(t)$ on a smooth function f defined in a neighborhood of a point $\widetilde{\gamma}(t_0)$. By the chain rule and the fact that γ is an integral curve,

$$\begin{aligned}
\widetilde{\gamma}'(t_0)f &= \left.\frac{d}{dt}\right|_{t=t_0} (f\circ\widetilde{\gamma})(t)\\
&= \left.\frac{d}{dt}\right|_{t=t_0} (f\circ\gamma)(t+a)\\
&= (f\circ\gamma)'(t_0+a)\\
&= \gamma'(t_0+a)f = V_{\gamma(t_0+a)}f = V_{\widetilde{\gamma}(t_0)}f.
\end{aligned}$$

Thus $\widetilde{\gamma}$ is an integral curve of V. □

Flows

There is another way to visualize the family of integral curves associated with a vector field. Let V be a vector field on a smooth manifold M, and suppose it has the property that for each point $p \in M$ there is a unique integral curve $\theta^{(p)} \colon \mathbb{R} \to M$ starting at p. (It may not always be the case that all of the integral curves are defined for all $t \in \mathbb{R}$, but for purposes of illustration let us assume for the time being that they are.) For each $t \in \mathbb{R}$, we can define a map θ_t from M to itself by sending each point $p \in M$ to the point obtained by following the curve starting at p for time t:

$$\theta_t(p) = \theta^{(p)}(t).$$

This defines a family of maps $\theta_t \colon M \to M$ for $t \in \mathbb{R}$. If $q = \theta^{(p)}(s)$, the translation lemma implies that the integral curve starting at q satisfies $\theta^{(q)}(t) = \theta^{(p)}(t+s)$. When we translate this into a statement about the maps θ_t, it becomes

$$\theta_t \circ \theta_s(p) = \theta_{t+s}(p).$$

Together with the equation $\theta_0(p) = \theta^{(p)}(0) = p$, which holds by definition, this implies that the map $\theta \colon \mathbb{R} \times M \to M$ is an action of the additive group $\mathbb{R}$ on M.

Motivated by these considerations, we define a *global flow* on M (sometimes also called a *one-parameter group action*) to be a smooth left action of $\mathbb{R}$ on M; that is, a smooth map $\theta \colon \mathbb{R} \times M \to M$ satisfying the following properties for all $s, t \in \mathbb{R}$ and all $p \in M$:

$$\begin{aligned} \theta(t, \theta(s,p)) &= \theta(t+s, p), \\ \theta(0,p) &= p. \end{aligned} \tag{12.1}$$

Given a global flow θ on M, we define two collections of maps as follows.

- For each $t \in \mathbb{R}$, define $\theta_t \colon M \to M$ by

  $$\theta_t(p) = \theta(t,p).$$

  The defining properties (12.1) are equivalent to the *group laws*:

  $$\begin{aligned} \theta_t \circ \theta_s &= \theta_{t+s}, \\ \theta_0 &= \mathrm{Id}_M . \end{aligned} \tag{12.2}$$

  As is the case for any smooth group action, each map $\theta_t \colon M \to M$ is a diffeomorphism.

- For each $p \in M$, define a smooth curve $\theta^{(p)} \colon \mathbb{R} \to M$ by

  $$\theta^{(p)}(t) = \theta(t,p).$$

  The image of this curve is just the orbit of p under the group action. Because any group action on a set partitions the set into disjoint orbits, it follows that M is the disjoint union of the images of these curves.

The next proposition shows that every global flow arises as the set of integral curves of some vector field.

Proposition 12.3. *Let $\theta\colon \mathbb{R}\times M \to M$ be a global flow. For each $p \in M$, define a tangent vector $V_p \in T_pM$ by*

$$V_p = \theta^{(p)\prime}(0) = \left.\frac{\partial}{\partial t}\right|_{t=0} \theta(t,p).$$

The assignment $p \mapsto V_p$ is a smooth vector field on M, and each curve $\theta^{(p)}$ is an integral curve of V.

The vector field V defined in this proposition is called the *infinitesimal generator* of θ, for reasons we will explain below.

Proof. To show that V is smooth, it suffices to show that Vf is smooth for any smooth function f defined on an open subset of M. For any such function, just note that

$$Vf(p) = V_pf = \theta^{(p)\prime}(0)f = \left.\frac{d}{dt}\right|_{t=0} f(\theta^{(p)}(t)) = \left.\frac{d}{dt}\right|_{t=0} f(\theta(t,p)).$$

Because $\theta(t,p)$ depends smoothly on (t,p), so does $f(\theta(t,p))$ by composition, and therefore so also does the derivative of $f(\theta(t,p))$ with respect to t. (You can interpret this derivative as the action of the smooth vector field $\partial/\partial t$ on the smooth function $f\circ\theta\colon \mathbb{R}\times M \to \mathbb{R}$.) If follows that $Vf(p)$ depends smoothly on p, so V is smooth.

To show that $\theta^{(p)}$ is an integral curve of V, we need to show that

$$\theta^{(p)\prime}(t) = V_{\theta^{(p)}(t)}$$

for all $p \in M$ and all $t \in \mathbb{R}$. Let $t_0 \in \mathbb{R}$ be arbitrary, and set $q = \theta^{(p)}(t_0) = \theta_{t_0}(p)$, so that what we have to show is $\theta^{(p)\prime}(t_0) = V_q$. By the group law, for all t,

$$\begin{aligned}\theta^{(q)}(t) &= \theta_t(q)\\ &= \theta_t(\theta_{t_0}(p))\\ &= \theta_{t+t_0}(p)\\ &= \theta^{(p)}(t+t_0).\end{aligned}$$

Therefore, for any smooth function f defined in a neighborhood of q,

$$\begin{aligned}V_qf &= \theta^{(q)\prime}(0)f\\ &= \left.\frac{d}{dt}\right|_{t=0} f(\theta^{(q)}(t))\\ &= \left.\frac{d}{dt}\right|_{t=0} f(\theta^{(p)}(t+t_0))\\ &= \theta^{(p)\prime}(t_0)f,\end{aligned}$$

which was to be shown. ☐

Another important property of the infinitesimal generator is that it is invariant under the flow, in the following sense. Let V be a smooth vector field on a smooth manifold M, and let $F\colon M \to M$ be a diffeomorphism. We say that V is *invariant* under F if $F_*V = V$. Unwinding the definition of the push-forward of a vector field, this means that for each $p \in M$, $F_*V_p = V_{F(p)}$.

Proposition 12.4. *Let θ be a global flow on M and let V be its infinitesimal generator. Then V is invariant under θ_t for each $t \in \mathbb{R}$.*

Proof. Let $p \in M$ and $t_0 \in \mathbb{R}$ be arbitrary, and set $q = \theta_{t_0}(p)$. We need to show that

$$(\theta_{t_0})_* V_p = V_q.$$

Applying the left-hand side to a smooth function f defined in a neighborhood of q and using the definition of V, we obtain

$$\begin{aligned}
(\theta_{t_0 *} V_p) f &= V_p(f \circ \theta_{t_0}) \\
&= \frac{d}{dt}\bigg|_{t=0} f \circ \theta_{t_0} \circ \theta^{(p)}(t) \\
&= \frac{d}{dt}\bigg|_{t=0} f(\theta_{t_0}(\theta_t(p))) \\
&= \frac{d}{dt}\bigg|_{t=0} f(\theta_{t_0+t}(p)) \\
&= \frac{d}{dt}\bigg|_{t=0} f(\theta^{(p)}(t_0+t)) \\
&= \theta^{(p)\prime}(t_0) f.
\end{aligned}$$

Since $\theta^{(p)}$ is an integral curve of V, $\theta^{(p)\prime}(t_0) = V_q$. ☐

We have seen that every global flow gives rise to a smooth vector field whose integral curves are precisely the curves defined by the flow. Conversely, we would like to be able to say that every smooth vector field is the infinitesimal generator of a global flow. However, it is easy to see that this cannot be the case, because there are vector fields whose integral curves are not defined for all $t \in \mathbb{R}$, as the following examples show.

Example 12.5. Let $M = \{(x, y) \in \mathbb{R}^2 : x < 0\}$, and let $V = \partial/\partial x$. Reasoning as in Example 12.1(a), we see that the integral curve of V starting at $(a, b) \in M$ is $\gamma(t) = (t + a, b)$. However, in this case, γ is defined only for $t < -a$.

Example 12.6. For a somewhat more subtle example, let M be all of $\mathbb{R}^2$ and let $W = x^2\partial/\partial x$. You can check easily that the unique integral curve of W starting at $(1,0)$ is

$$\gamma(t) = \left(\frac{1}{1-t}, 0\right).$$

This curve is defined only for $t < 1$.

For this reason, we make the following definitions. If M is a smooth manifold, a *flow domain* for M is an open subset $\mathcal{D} \subset \mathbb{R} \times M$ with the property that for each $p \in M$, the set $\mathcal{D}_p = \{t \in \mathbb{R} : (t,p) \in \mathcal{D}\}$ is an open interval containing 0. A *flow* on M is a smooth map $\theta\colon \mathcal{D} \to M$, where $\mathcal{D} \subset \mathbb{R} \times M$ is a flow domain, that satisfies

$$\begin{aligned} \theta(0,p) &= p \text{ for all } p \in M, \\ \theta(t,\theta(s,p)) &= \theta(t+s,p) \text{ whenever } s \in \mathcal{D}_p \text{ and } t \in \mathcal{D}_{\theta(s,p)}. \end{aligned}$$

We sometimes call θ a *local flow* to distinguish it from a global flow as defined earlier. The unwieldy term *local one-parameter group action* is also commonly used.

If θ is a flow, we define $\theta_t(p) = \theta^{(p)}(t) = \theta(t,p)$ whenever $(t,p) \in \mathcal{D}$, just as for a local flow. Similarly, the *infinitesimal generator* of θ is defined by $V_p = \theta^{(p)\prime}(0)$.

Lemma 12.7 (Properties of Flows). *Let $\mathcal{D}$ be a flow domain for M, and let $\theta\colon \mathcal{D} \to M$ be a flow.*

(*a*) *For each $t \in \mathbb{R}$, the set $M_t = \{p \in M : (t,p) \in \mathcal{D}\}$ is open in M, and $\theta_t\colon M_t \to M$ is a diffeomorphism onto an open subset of M.*

(*b*) *The following relation holds whenever the left-hand side is defined:*

$$\theta_t \circ \theta_s = \theta_{t+s}.$$

(*c*) *The infinitesimal generator V of θ is a smooth vector field.*

(*d*) *For each $p \in M$, $\theta^{(p)}\colon \mathcal{D}_p \to M$ is an integral curve of V starting at p.*

(*e*) *For each $t \in \mathbb{R}$, $\theta_{t*}V = V$ on the open set $\theta_t(M_t)$.*

Exercise 12.1. Prove Lemma 12.7.

Another important property of flows is the following.

Lemma 12.8. *Suppose θ is a flow on M with infinitesimal generator V, and $p \in M$. If $V_p = 0$, then $\theta^{(p)}$ is the constant curve $\theta^{(p)}(t) \equiv p$. If $V_p \neq 0$, then $\theta^{(p)}\colon \mathcal{D}_p \to M$ is an immersion.*

Proof. For simplicity, write $\gamma = \theta^{(p)}$. Let $t \in \mathcal{D}_p$ be arbitrary, and put $q = \gamma(t)$. Note that the push-forward $\gamma_* \colon T_t\mathbb{R} \to T_qM$ is zero if and only if $\gamma'(t) = 0$. Part (e) of Lemma 12.7 shows that $V_q = \theta_{t*}V_p$. Therefore $\gamma'(t) = V_q = 0$ if and only if $\gamma'(0) = V_p = 0$; in other words, if $\gamma'(t)$ vanishes for some $t \in \mathcal{D}_p$ it vanishes for all such t. Thus if $V_p = 0$, then γ is a smooth map whose push-forward at each point is zero, which implies that it is a constant map (because its domain is connected). On the other hand, if $V_p \neq 0$, then γ_* is nonzero, hence injective, at each point, so γ is an immersion. □

The Fundamental Theorem on Flows

In this section we will see that every smooth vector field gives rise to a flow, which is unique if we require it to be *maximal*, which means that it cannot be extended to any larger flow domain.

Theorem 12.9 (Fundamental Theorem on Flows). *Let V be a smooth vector field on a smooth manifold M. There is a unique maximal flow whose infinitesimal generator is V.*

The flow whose existence is asserted in this theorem is called the flow *generated by V*.

The term "infinitesimal generator" comes from the following picture. In a local coordinate chart, a reasonably good approximation to the flow can be obtained by composing very many small affine translations, with the direction and length of each successive motion determined by the value of the vector field at the point arrived at in the previous step. Long ago, mathematicians thought of a flow as being composed of infinitely many infinitesimally small linear steps.

As we saw earlier in this chapter, finding integral curves of V (and therefore finding the flow generated by V) boils down to solving a system of ordinary differential equations, at least locally. Thus before beginning the proof, let us state a basic theorem about solutions of ordinary differential equations. We will give the proof of this theorem in the last section of the chapter.

Theorem 12.10 (ODE Existence, Uniqueness, and Smoothness). *Let $U \subset \mathbb{R}^n$ be open, and let $V \colon U \to \mathbb{R}^n$ be a smooth map. For any $(t_0, x_0) \in \mathbb{R} \times U$ and any sufficiently small $\varepsilon > 0$, there exist an open set $U_0 \subset U$ containing x_0 and a smooth map $\theta \colon (t_0 - \varepsilon, t_0 + \varepsilon) \times U_0 \to U$ such that for each $x \in U_0$, the curve $\gamma(t) = \theta(t, x)$ is the unique solution on $(t_0 - \varepsilon, t_0 + \varepsilon)$ to the initial-value problem*

$$\begin{aligned} {\gamma^i}'(t) &= V^i(\gamma(t)), \\ \gamma^i(t_0) &= x^i. \end{aligned}$$

Using this result, we now prove the fundamental theorem on flows.

Proof of Theorem 12.9. We begin by noting that the existence assertion of the ODE theorem implies that there exists an integral curve starting at each point $p \in M$, because the equation for an integral curve is a system of ODEs in any local coordinates around p.

Now suppose $\gamma, \widetilde{\gamma}\colon J \to M$ are two integral curves defined on the same open interval J such that $\gamma(t_0) = \widetilde{\gamma}(t_0)$ for some $t_0 \in J$. Let $\mathcal{S}$ be the set of $t \in J$ such that $\gamma(t) = \widetilde{\gamma}(t)$. Clearly $\mathcal{S}$ is nonempty because $t_0 \in \mathcal{S}$ by hypothesis, and it is closed in J by continuity. On the other hand, suppose $t_1 \in \mathcal{S}$. Then in a coordinate neighborhood around the point $p = \gamma(t_1)$, γ and $\widetilde{\gamma}$ are both solutions to same ODE with the same initial condition $\gamma(t_1) = \widetilde{\gamma}(t_1) = p$. By the ODE theorem, there is a neighborhood $(t_1 - \varepsilon, t_1 + \varepsilon)$ of t_1 on which there is a unique solution to this initial-value problem. Thus $\gamma \equiv \widetilde{\gamma}$ on $(t_1 - \varepsilon, t_1 + \varepsilon)$, which implies that $\mathcal{S}$ is open in J. Since J is connected, $\mathcal{S} = J$, which implies that $\gamma = \widetilde{\gamma}$ on all of J. Thus any two integral curves that agree at one point agree on their common domain.

For each $p \in M$, let $\mathcal{D}_p$ be the union of all intervals $J \subset \mathbb{R}$ containing 0 on which an integral curve starting at p is defined. Define $\theta^{(p)}\colon \mathcal{D}_p \to M$ by letting $\theta^{(p)}(t) = \gamma(t)$, where γ is any integral curve starting at p and defined on an open interval containing 0 and t. Since all such integral curves agree at t by the argument above, $\theta^{(p)}$ is well defined, and is obviously the unique maximal integral curve starting at p.

Now, let $\mathcal{D}(V) = \{(t,p) \in \mathbb{R} \times M : t \in \mathcal{D}_p\}$, and define $\theta\colon \mathcal{D}(V) \to M$ by $\theta(t,p) = \theta^{(p)}(t)$. As usual, we also write $\theta_t(p) = \theta(t,p)$. Clearly $\theta_0 = \mathrm{Id}_M$ by definition. We will verify that θ satisfies the group law

$$\theta_t \circ \theta_s(p) = \theta_{t+s}(p) \tag{12.3}$$

whenever the left-hand side is defined. Fix any $s \in \mathcal{D}_p$, and write $q = \theta_s(p) = \theta^{(p)}(s)$. Define $\gamma(t) = \theta_{t+s}(p) = \theta^{(p)}(t+s)$ wherever the latter is defined. Then $\gamma(0) = q$, and the translation lemma shows that γ is an integral curve of V. Thus by the uniqueness assertion above γ must be equal to $\theta^{(q)}$ wherever both are defined, which shows that (12.3) holds when both sides are defined. Since both $\theta^{(p)}$ and $\theta^{(q)}$ are maximal, it follows that $t \in \mathcal{D}_q$ if and only if $t + s \in \mathcal{D}_p$; in particular, if the left-hand side of (12.3) is defined, then so is the right-hand side.

Next we will show that $\mathcal{D}(V)$ is open in $\mathbb{R} \times M$ and that $\theta\colon \mathcal{D}(V) \to M$ is smooth. This implies $\mathcal{D}(V)$ is a flow domain; since it is obviously maximal by definition, this will complete the proof.

Define a subset $W \subset \mathcal{D}(V)$ as the set of all $(t,p) \in \mathcal{D}(V)$ such that θ is defined and smooth on a product open set $J \times U \subset \mathbb{R} \times M$, where J is an open interval containing 0 and t. Clearly W is open in $\mathbb{R} \times M$ and the restriction of θ to W is smooth, so it suffices to show that $W = \mathcal{D}(V)$. Suppose this is not the case. Then there exists some point $(t_0, p_0) \in \mathcal{D}(V) \setminus W$. For simplicity, let us assume $t_0 > 0$; the argument for $t_0 < 0$ is similar.

Let $\tau = \sup\{t \in \mathbb{R} : (t, p_0) \in W\}$. By the ODE theorem (applied in coordinates around p_0), θ is defined and smooth in some neighborhood of $(0, p_0)$, so $\tau > 0$. Let $q_0 = \theta^{(p_0)}(\tau)$. By the ODE theorem again, there is some $\varepsilon > 0$ and a neighborhood U_0 of q_0 such that $\theta\colon (-\varepsilon, \varepsilon) \times U_0 \to M$ is defined and smooth. We will use the group law to show that θ extends smoothly to a neighborhood of (τ, p_0), which contradicts our choice of τ.

Choose some $t_1 < \tau$ such that $t_1 + \varepsilon > \tau$ and $\theta^{(p_0)}(t_1) \in U_0$. Since $t_1 < \tau$, $(t_1, p_0) \in W$, and so there is a product neighborhood $(-\delta, t_1 + \delta) \times U_1$ of (t_1, p_0) on which θ is defined and smooth. Because $\theta(t_1, p_0) \in U_0$, we can choose U_1 small enough that θ maps $\{t_1\} \times U_1$ into U_0. Because θ satisfies the group law, we have

$$\theta_t(p) = \theta_{t-t_1} \circ \theta_{t_1}(p)$$

whenever the right-hand side is defined. By our choice of t_1, $\theta_{t_1}(p)$ is defined for $p \in U_1$, and depends smoothly on p. Moreover, since $\theta_{t_1}(p) \in U_0$ for all such p, it follows that $\theta_{t-t_1} \circ \theta_{t_1}(p)$ is defined whenever $p \in U_1$ and $|t - t_1| < \varepsilon$, and depends smoothly on (t, p). This gives a smooth extension of θ to the product set $(-\delta, t_1 + \varepsilon) \times U_1$, which contradicts our choice of τ. This completes the proof that $W = \mathcal{D}(V)$. □

Complete Vector Fields

As we noted above, not every vector field generates a global flow. The ones that do are important enough to deserve a name. We say a vector field is *complete* if it generates a global flow, or equivalently if each of its integral curves is defined for all $t \in \mathbb{R}$.

It is not always easy to determine by looking at a vector field whether it is complete or not. If you can solve the ODE explicitly to find all of the integral curves, and they all exist for all time, then the vector field is complete. On the other hand, if you can find a single integral curve that cannot be extended to all of $\mathbb{R}$, as we did for the vector field of Example 12.6, then it is not complete. However, it is often impossible to solve the ODE explicitly, so it is useful to have some general criteria for determining when a vector field is complete.

In this section we will show that all vector fields on a compact manifold are complete. (Problem 12-1 gives a more general sufficient condition.) The proof will be based on the following lemma.

Lemma 12.11 (Escape Lemma). *Let M be a smooth manifold and let V be a vector field on M. If γ is an integral curve of V whose maximal domain is not all of $\mathbb{R}$, then the image of γ cannot lie in any compact subset of M.*

Proof. Suppose γ is defined on a maximal domain of the form (a, b), and assume that $b < \infty$. (The argument for the case $a > -\infty$ is similar.) We will show that if $\gamma[0,b)$ lies in a compact set, then γ can be extended past b, which is a contradiction.

Let $p = \gamma(0)$ and let θ denote the flow of V, so $\gamma = \theta^{(p)}$ by the uniqueness of integral curves. If $\{t_i\}$ is any sequence of times approaching b from below, then the sequence $\{\gamma(t_i)\}$ lies in a compact subset of M, and therefore has a subsequence converging to a point $q \in M$. Choose a neighborhood U of q and a positive number ε such that θ is defined on $(-\varepsilon, \varepsilon) \times U$. Pick some i large enough that $\gamma(t_i) \in U$ and $t_i > b - \varepsilon$, and define $\sigma\colon [0, t_i + \varepsilon) \to M$ by

$$\sigma(t) = \begin{cases} \gamma(t), & 0 \le t < b, \\ \theta_{t-t_i} \circ \theta_{t_i}(p), & t_i - \varepsilon < t < t_i + \varepsilon. \end{cases}$$

These two definitions agree where they overlap, because $\theta_{t-t_i} \circ \theta_{t_i}(p) = \theta_t(p) = \gamma(t)$ by the group law for θ. Thus σ is an integral curve extending γ, which contradicts the maximality of γ. Therefore, $\gamma[0,b)$ cannot lie in any compact set. □

Theorem 12.12. *Suppose M is a compact manifold. Then every vector field on M is complete.*

Proof. If M is compact, the escape lemma implies that no integral curve can have a maximal domain that is not all of $\mathbb{R}$, because the image of any integral curve is contained in the compact set M. □

Proof of the ODE Theorem

In this section we prove the ODE existence, uniqueness, and smoothness theorem (Theorem 12.10). Actually, it will be useful to prove the existence theorem under the somewhat weaker hypothesis that the vector field is only Lipschitz continuous (see the Appendix).

We will prove Theorem 12.10 in several parts: The uniqueness assertion follows from the next theorem, existence from Theorem 12.13, and smoothness from Theorem 12.16. Throughout this section, $U \subset \mathbb{R}^n$ will be an open set, and $V\colon U \to \mathbb{R}^n$ will be a Lipschitz continuous map. For any $t_0 \in \mathbb{R}$ and any $x \in U$ we will study the initial-value problem

$$\begin{aligned} \gamma^{i\prime}(t) &= V^i(\gamma(t)), \\ \gamma^i(t_0) &= x^i. \end{aligned} \tag{12.4}$$

Theorem 12.13 (Uniqueness of ODE Solutions). *Any two solutions to* (12.4) *are equal on their common domain.*

Proof. Suppose $\gamma, \widetilde{\gamma}\colon J \to U$ are two solutions to the ODE on the same open interval J, not necessarily with the same initial conditions. The Schwartz inequality and the Lipschitz estimate for V imply

$$\begin{aligned}\frac{d}{dt}|\widetilde{\gamma}(t) - \gamma(t)|^2 &= 2(\widetilde{\gamma}(t) - \gamma(t)) \cdot (V(\widetilde{\gamma}(t)) - V(\gamma(t)))\\ &\le 2|\widetilde{\gamma}(t) - \gamma(t))|\,|V(\widetilde{\gamma}(t)) - V(\gamma(t))|\\ &\le 2C|\widetilde{\gamma}(t) - \gamma(t)|^2.\end{aligned}$$

It follows easily that

$$\frac{d}{dt}(e^{-2Ct}|\widetilde{\gamma}(t) - \gamma(t)|^2) \le 0,$$

and so

$$e^{-2Ct}|\widetilde{\gamma}(t) - \gamma(t)|^2 \le e^{-2Ct_0}|\widetilde{\gamma}(t_0) - \gamma(t_0)|^2, \quad t \ge t_0.$$

Similarly, using the estimate

$$\frac{d}{dt}|\widetilde{\gamma}(t) - \gamma(t)|^2 \ge -2C|\widetilde{\gamma}(t) - \gamma(t)|^2,$$

we conclude that

$$e^{2Ct}|\widetilde{\gamma}(t) - \gamma(t)|^2 \le e^{2Ct_0}|\widetilde{\gamma}(t_0) - \gamma(t_0)|^2, \quad t \le t_0.$$

Putting these two estimates together, we obtain the following for all $t \in J$:

$$|\widetilde{\gamma}(t) - \gamma(t)| \le e^{C|t-t_0|}|\widetilde{\gamma}(t_0) - \gamma(t_0)|. \tag{12.5}$$

Thus $\gamma(t_0) = \widetilde{\gamma}(t_0)$ implies $\gamma \equiv \widetilde{\gamma}$ on all of J. □

Theorem 12.14 (Existence of ODE Solutions). *For each $t_0 \in \mathbb{R}$ and $x_0 \in U$, there exist an open interval J_0 containing t_0, an open set $U_0 \subset U$ containing x_0, and for each $x \in U_0$ a differentiable curve $\gamma\colon J_0 \to U$ satisfying the initial-value problem* (12.4).

Proof. If γ is any continuous curve in U, the fundamental theorem of calculus implies that γ is a solution to the initial-value problem (12.4) if and only if it satisfies the integral equation

$$\gamma(t) = x + \int_{t_0}^{t} V(\gamma(s))\, ds, \tag{12.6}$$

where the integral of the vector-valued function $V(\gamma(s))$ is obtained by integrating each component separately. For any such γ we define a new curve $A\gamma$ by

$$A\gamma(t) = x + \int_{t_0}^{t} V(\gamma(s))\, ds. \tag{12.7}$$

Then we are led to seek a fixed point for A in a suitable metric space of curves.

Let C be a Lipschitz constant for V. Given $t_0 \in \mathbb{R}$ and $x_0 \in U$, choose $r > 0$ such that $\overline{B}_r(x_0) \subset U$, and let M be the supremum of $|V(x)|$ on $\overline{B}_r(x_0)$. Set $J_0 = (t_0 - \varepsilon, t_0 + \varepsilon)$ and $U_0 = B_\delta(x_0)$, where ε and δ are chosen small enough that

$$\delta \le \frac{r}{2}, \quad \varepsilon < \min\left(\frac{r}{2M}, \frac{1}{C}\right).$$

For any $x \in U_0$, let $\mathcal{M}_x$ denote the set of all continuous maps $\gamma\colon \overline{J}_0 \to \overline{B}_r(x_0)$ satisfying $\gamma(0) = x$. We define a metric on $\mathcal{M}_x$ by

$$d(\gamma, \widetilde{\gamma}) = \sup_{t \in \overline{J}_0} |\gamma(t) - \widetilde{\gamma}(t)|.$$

Any sequence of maps in $\mathcal{M}_x$ that is Cauchy in this metric is uniformly convergent, and therefore has a continuous limit γ. Clearly, the conditions that γ take its values in $\overline{B}_r(x_0)$ and $\gamma(0) = x$ are preserved in the limit. Therefore, $\mathcal{M}_x$ is a complete metric space.

We wish to define a map $A\colon \mathcal{M}_x \to \mathcal{M}_x$ by formula (12.7). The first thing we need to verify is that A really does map $\mathcal{M}_x$ into itself. It is clear from the definition that $A\gamma(0) = x$ and $A\gamma$ is continuous (in fact, it is differentiable by the fundamental theorem of calculus). Thus we need only check that $A\gamma$ takes its values in $\overline{B}_r(x_0)$. If $\gamma \in \mathcal{M}_x$, then for any $t \in \overline{J}_0$,

$$\begin{aligned}
|A\gamma(t) - x_0| &= |x + \int_{t_0}^t V(\gamma(s))\,ds - x_0| \\
&\le |x - x_0| + \int_{t_0}^t |V(\gamma(s))|\,ds \\
&< \delta + M\varepsilon \le r
\end{aligned}$$

by our choice of δ and ε.

Next we check that A is a contraction. If $\gamma, \widetilde{\gamma} \in \mathcal{M}_x$, then using the Lipschitz condition on V, we obtain

$$\begin{aligned}
d(A\gamma, A\widetilde{\gamma}) &= \sup_{t \in \overline{J}_0} \left| \int_{t_0}^t V(\gamma(s))\,ds - \int_{t_0}^t V(\widetilde{\gamma}(s))\,ds \right| \\
&\le \sup_{t \in \overline{J}_0} \int_{t_0}^t |V(\gamma(s)) - V(\widetilde{\gamma}(s))|\,ds \\
&\le \sup_{t \in \overline{J}_0} \int_{t_0}^t C|\gamma(s) - \widetilde{\gamma}(s)|\,ds \\
&\le C\varepsilon d(\gamma, \widetilde{\gamma}).
\end{aligned}$$

Because we have chosen ε so that $C\varepsilon < 1$, this shows that A is a contraction. By the contraction lemma, A has a fixed point $\gamma \in \mathcal{M}_x$, which is a solution to (12.4). □

As a preliminary step in proving smoothness of the solution, we need the following continuity result.

Lemma 12.15 (Continuity of ODE Solutions). *Suppose J_0 is an open interval containing t_0, $U_0 \subset U$ is an open set, and $\theta\colon J_0 \times U_0 \to U$ is any map such that for each $x \in U_0$, $\gamma(t) = \theta(t, x)$ solves* (12.4)*. Then θ is continuous.*

Proof. It suffices to show that θ is continuous in a neighborhood of each point, so by shrinking J_0 and U_0 slightly we might as well assume that J_0 is precompact in $\mathbb{R}$ and U_0 is precompact in U.

First we note that θ is Lipschitz continuous in x, with a constant that is independent of t, because (12.5) implies

$$|\theta(t, \widetilde{x}) - \theta(t, x)| \le e^{CT}|\widetilde{x} - x| \quad \text{for all } x, \widetilde{x} \in U_0, \tag{12.8}$$

where $T = \sup_{\overline{J}_0} |t - t_0|$.

Now let $(t, x), (\widetilde{t}, \widetilde{x}) \in J_0 \times U_0$ be arbitrary. Using the fact that every solution to the initial-value problem satisfies the integral equation (12.6), we find

$$\begin{aligned}
|\theta(\widetilde{t}, \widetilde{x}) - \theta(t, x)| &\le |\widetilde{x} - x| + \left| \int_{t_0}^{\widetilde{t}} V(\theta(s, \widetilde{x}))\, ds - \int_{t_0}^{t} V(\theta(s, x))\, ds \right| \\
&\le |\widetilde{x} - x| + \int_{t_0}^{t} |V(\theta(s, \widetilde{x})) - V(\theta(s, x))|\, ds \\
&\qquad + \int_{t}^{\widetilde{t}} |V(\theta(s, \widetilde{x}))|\, ds \\
&\le |\widetilde{x} - x| + C \int_{t_0}^{t} |\theta(s, \widetilde{x}) - \theta(s, x)|\, ds + \int_{t}^{\widetilde{t}} M\, ds \\
&\le |\widetilde{x} - x| + CTe^{CT}|\widetilde{x} - x| + M|\widetilde{t} - t|,
\end{aligned}$$

where M is the supremum of $|V|$ on $\overline{U}_0$. It follows that θ is continuous. □

Theorem 12.16 (Smoothness of ODE Solutions). *Let θ be as in the preceding theorem. If V is smooth, then so is θ.*

Proof. We will prove the following claim by induction on k:

$$\textit{If } V \textit{ is of class } C^{k+1}, \textit{ then } \theta \textit{ is of class } C^k. \tag{12.9}$$

From this it follows that if V is smooth, then θ is of class C^k for every k, and thus is smooth.

The hardest part of the proof is the $k=1$ step. Expressed in terms of θ, the initial-value problem (12.4) reads

$$\begin{aligned} \frac{\partial}{\partial t}\theta^i(t,x) &= V^i(\theta(t,x)),\\ \theta^i(t_0,x) &= x^i. \end{aligned} \tag{12.10}$$

Let us pretend for a moment that everything in sight is smooth, and differentiate both of these equations with respect to x^j. Since mixed partial derivatives of smooth functions commute, we obtain

$$\begin{aligned} \frac{\partial}{\partial t}\frac{\partial\theta^i}{\partial x^j}(t,x) &= \frac{\partial V^i}{\partial x^k}(\theta(t,x))\frac{\partial\theta^k}{\partial x^j}(t,x),\\ \frac{\partial\theta^i}{\partial x^j}(t_0,x) &= \delta^i_j = \begin{cases} 1 & \text{if } i=j,\\ 0 & \text{if } i\neq j. \end{cases} \end{aligned}$$

The idea of the proof for $k=1$ is to show that the partial derivatives $\partial\theta^i/\partial x^j$ exist and solve this system of equations, and then to use the continuity lemma to conclude that these partial derivatives are continuous.

To that end, we let $G\colon U\to \mathrm{M}(n,\mathbb{R})$ denote the matrix-valued function $G(x)=DV(x)$. The assumption that V is C^2 implies that G is C^1, so (shrinking U_0 if necessary) the map $U_0\times\mathrm{M}(n,\mathbb{R})\to\mathrm{M}(n,\mathbb{R})$ given by $(x,y)\mapsto G(x)y$ is Lipschitz.

Consider the following initial-value problem for the $n+n^2$ unknown functions (θ^i,ψ^i_j):

$$\begin{aligned} \frac{\partial}{\partial t}\theta(t,x) &= V(\theta(t,x)),\\ \frac{\partial}{\partial t}\psi(t,x) &= G(\theta(t,x))\psi(t,x);\\ \theta(t_0,x) &= x,\\ \psi(t_0,x) &= I_n. \end{aligned} \tag{12.11}$$

This system is called the *variational equation* for the system (12.4). By the existence and continuity theorems, for any $x_0\in U_0$ there exist an interval $J_1\subset J_0$ containing t_0, a neighborhood of (x_0,I_n) in $U_0\times\mathrm{M}(n,\mathbb{R})$ (which we may assume to be a product set $U_1\times W_1$), and a continuous map $(\theta,\psi)\colon J_1\times U_1\times W_1\to U_0\times\mathrm{M}(n,\mathbb{R})$ satisfying (12.11). If we can show that $\partial\theta^i/\partial x^j$ exists for each i and j and equals ψ^i_j on $J_1\times U_1$, then $\partial\theta^i/\partial x^j$ is continuous there. Moreover, (12.4) implies $\partial\theta^i/\partial t$ is continuous, so it follows that θ is C^1, at least on the set $J_1\times U_1$.

We will show that $\partial\theta^i/\partial x^j$ exists by working directly with the definition of the derivative. For any sufficiently small $h\in\mathbb{R}\smallsetminus\{0\}$, let $\Delta_h\colon J_1\times U_1\to$

$\mathrm{M}(n,\mathbb{R})$ be the difference quotient

$$(\Delta_h)^i_j(t,x) = \frac{\theta^i(t,x+he_j) - \theta^i(t,x)}{h}.$$

Then $\partial\theta^i/\partial x^j(t,x) = \lim_{h\to 0}(\Delta_h)^i_j(t,x)$ if the limit exists.

Because a C^2 map is differentiable, for sufficiently small $\delta > 0$ there is a map $R\colon B_\delta(0) \to \mathbb{R}^n$ such that

$$V(\widetilde{x}) - V(x) = DV(x)(\widetilde{x} - x) + R(\widetilde{x} - x) \tag{12.12}$$

and

$$\lim_{v\to 0} \frac{R(v)}{|v|} = 0. \tag{12.13}$$

Let us compute the t-derivative of Δ_h using (12.12):

$$\begin{aligned}
\frac{\partial}{\partial t}(\Delta_h)^i_j(t,x) &= \frac{1}{h}\left(\frac{\partial}{\partial t}\theta^i(t,x+he_j) - \frac{\partial}{\partial t}\theta^i(t,x)\right)\\
&= \frac{1}{h}\big(V^i(\theta(t,x+he_j)) - V^i(\theta(t,x))\big)\\
&= \frac{1}{h}\left(\frac{\partial V^i}{\partial x^k}(\theta(t,x))(\theta^k(t,x+he_j) - \theta^k(t,x))\right.\\
&\qquad \left. + R^i(\theta(t,x+he_j) - \theta(t,x))\right)\\
&= G^i_k(\theta(t,x))(\Delta_h)^k_j(t,x)) + \frac{R^i(\theta(t,x+he_j) - \theta(t,x))}{h}.
\end{aligned}$$

By 12.13, given any $\varepsilon > 0$ there exists $\delta > 0$ such that $|v| < \delta$ implies $|R(v)|/|v| < \varepsilon$. The Lipschitz estimate (12.8) for θ says that

$$|\theta(t,x+he_j) - \theta(t,x))| \le K|he_j| = K|h|$$

for some constant K, and therefore $|h| < \delta/K$ implies that

$$\left|\frac{R(\theta(t,x+he_j) - \theta(t,x))}{h}\right| \le \varepsilon.$$

Summarizing, we have shown that for h sufficiently small, Δ_h is an "ε-approximate solution" to the second equation of (12.11), in the sense that

$$\frac{\partial}{\partial t}\Delta_h(t,x) = G(\theta(t,x))\Delta_h(t,x) + E(t,x),$$

where $|E(t,x)| < \varepsilon$. It also satisfies the initial condition $\Delta_h(t_0,x) = I_n$ for each x.

Let (θ, ψ) be the exact solution to (12.11). We will show that Δ_h converges to ψ as $h \to 0$. To do so, we note that

$$\begin{aligned}
&\frac{\partial}{\partial t}|\psi(t,x) - \Delta_h(t,x)|^2 \\
&\quad = 2(\psi(t,x) - \Delta_h(t,x)) \cdot \\
&\qquad\quad \big(G(\theta(t,x))\psi(t,x) - G(\theta(t,x))\Delta_h(t,x) - E(t,x)\big) \\
&\quad \leq 2|\psi(t,x) - \Delta_h(t,x)| \left(|G(\theta(t,x)|\,|\psi(t,x) - \Delta_h(t,x)| + \varepsilon\right) \\
&\quad \leq 2B|\psi(t,x) - \Delta_h(t,x)|^2 + 2\varepsilon|\psi(t,x) - \Delta_h(t,x)| \\
&\quad \leq (2B+1)|\psi(t,x) - \Delta_h(t,x)|^2 + \varepsilon^2,
\end{aligned}$$

where $B = \sup|G|$, and the last line follows from the inequality $2ab \leq a^2+b^2$, which is proved just by expanding $(a-b)^2 \geq 0$. Thus if $J_1 \subset [-T,T]$,

$$\frac{\partial}{\partial t}\big(e^{-(2B+1)t}|\psi(t,x) - \Delta_h(t,x)|^2\big) \leq \varepsilon^2 e^{-(2B+1)t} \leq \varepsilon^2 e^{(2B+1)T}.$$

Since $\psi(t_0,x) = \Delta_h(t_0,x)$, it follows by elementary calculus that

$$e^{-(2B+1)t}|\psi(t,x) - \Delta_h(t,x)|^2 \leq \varepsilon^2 e^{(2B+1)T}|t - t_0|.$$

In particular, since ε can be made as small as desired by choosing h sufficiently small, this shows that $\lim_{h\to 0}\Delta_h(t,x) = \psi(t,x)$. Therefore, θ is C^1 on the set $J_1 \times U_1$.

To show that θ is actually C^1 on its entire domain $J_0 \times U_0$, we proceed just as in the proof of Theorem 12.9. For each $x_0 \in U_0$, the argument above shows that θ is C^1 in some neighborhood $J_1 \times U_1$ of (t_0, x_0). Let τ be the supremum of the set of $t > t_0$ such that θ is C^1 on a product neighborhood of $[t_0, t] \times \{x_0\}$. If $(\tau, x_0) \in U_0$, we will show that θ is C^1 on a product neighborhood of $[t_0, \tau] \times \{x_0\}$, which contradicts our choice of τ. (The argument for $t < t_0$ is similar.)

By the argument above, there exist $\varepsilon > 0$, a neighborhood U_2 of $\theta(\tau, x_0)$, and a C^1 map $\beta\colon (-\varepsilon, +\varepsilon) \times U_2 \to U_0$ such that

$$\begin{aligned}
\frac{\partial}{\partial t}\beta^i(t,x) &= V^i(\beta(t,x)), \\
\beta^i(0,x) &= x^i.
\end{aligned}$$

Choose $t_1 \in (\tau - \varepsilon, \tau)$ such that $\theta(t_1, x_0) \in U_2$. Then $\theta(t,x) = \beta(t - t_1, \theta(t_1, x))$ where both are defined, because they both solve the same ODE and are equal to $\theta(t_1, x)$ when $t = t_1$. Since the right-hand side is smooth for $x \in U_2$ and $t_1 - \varepsilon < t < t_1 + \varepsilon$, this shows that θ is C^1 on $[t_0, t_1 + \varepsilon) \times$

$(U_2 \cap U_1)$, which contradicts our choice of τ. Therefore, θ is C^1 on its whole domain, thus proving (12.9) for $k = 1$. Moreover, we have also shown that $(\theta^i, \partial\theta^i/\partial x^j)$ solves the variational equation (12.11).

Now assume by induction that (12.9) is true for $1 \le k < k_0$, and suppose V is of class C^{k_0+1}. By the argument above, $(\theta^i, \partial\theta^i/\partial x^j)$ solves the variational equation. Since the first partial derivatives of V are of class C^{k_0}, the inductive hypothesis applied to (12.11) shows that $(\theta^i, \partial\theta^i/\partial x^j)$ is of class C^{k_0-1}. Since θ is of class C^{k_0-1} by the inductive hypothesis and therefore $\partial\theta^i/\partial t$ is C^{k_0-1} by (12.10), it follows that all of the partial derivatives of θ are C^{k_0-1}, so θ itself is C^{k_0}, thus completing the induction. $\square$

Problems

12-1. Show that every smooth vector field with compact support is complete.

12-2. Let M be a compact Riemannian n-manifold, and $f \in C^\infty(M)$. Suppose f has only finitely many critical points $\{p_1, \dots, p_k\}$ with corresponding critical values $\{c_1, \dots, c_k\}$. (Assume without loss of generality that $c_1 \le \cdots \le c_k$.) For any $a, b \in \mathbb{R}$, define $M_a = f^{-1}(a)$ and $M_{[a,b]} = f^{-1}([a,b])$. If a is a regular value, note that M_a is an embedded hypersurface in M.

(a) Let X be the vector field $X = \operatorname{grad} f / |\operatorname{grad} f|^2$ on $M \smallsetminus \{p_1, \dots, p_k\}$, and let θ denote the flow of X. Show that $f(\theta_t(p)) = f(p) + t$ whenever $\theta_t(p)$ is defined.

(b) Let $[a,b] \subset \mathbb{R}$ be an interval containing no critical values of f. Show that
$$\theta : [0, b-a] \times M_a \to M_{[a,b]}$$
is a diffeomorphism, whose inverse is $p \mapsto \big(f(p) - a, \theta(a - f(p), p)\big)$.

[Remark: This result shows that M can be decomposed as a union of simpler "building blocks"—the product manifolds $M_{[c_i+\varepsilon, c_{i+1}-\varepsilon]} \approx I \times M_{c_i+\varepsilon}$, and the neighborhoods $f^{-1}((c_i - \varepsilon, c_i + \varepsilon))$ of the critical points. This is the starting point of *Morse theory*, which is one of the deepest applications of differential geometry to topology. It is enlightening to think about what this means when M is a torus of revolution in $\mathbb{R}^3$ obtained by revolving a circle around the z-axis, and $f(x,y,z) = x$.]

12-3. Let M be a connected smooth manifold. Show that the group of diffeomorphisms of M acts transitively on M. More precisely, for any two points $p, q \in M$, show that there is a diffeomorphism $F\colon M \to M$ such that $F(p) = q$. [Hint: First prove the following lemma: If $p, q \in \mathbb{B}^n$ (the open unit ball in $\mathbb{R}^n$), there is a compactly supported vector field on $\mathbb{B}^n$ whose flow θ satisfies $\theta_1(p) = q$.]

12-4. Let M be a smooth manifold. A smooth curve $\gamma\colon \mathbb{R} \to M$ is said to be *periodic* if there exists $T > 0$ such that $\gamma(t) = \gamma(t')$ if and only if $t - t' = kT$ for some $k \in \mathbb{Z}$. Suppose $X \in \mathcal{T}(M)$ and γ is a maximal integral curve of X.

(a) Show that exactly one of the following holds:

- γ is constant.
- γ is injective.

- γ is periodic.

(b) Show that the image of γ is an immersed submanifold of M.

12-5. Show that there is only one smooth structure on $\mathbb{R}$ up to diffeomorphism. More precisely, if M is any smooth manifold that is homeomorphic to $\mathbb{R}$, show that M is diffeomorphic to $\mathbb{R}$ (with its standard smooth structure). [Hint: First show that M admits a nowhere-vanishing smooth vector field. See Problem 10-1.]

12-6. Let θ be a flow on an oriented manifold. Show that for each $t \in \mathbb{R}$, θ_t is orientation-preserving wherever it is defined.

12-7. All of the systems of differential equations considered in this chapter have been of the form

$$\gamma^{i\prime}(t) = V^i(\gamma(t)),$$

in which the functions V^i do not depend explicitly on the independent variable t. (Such a system is said to be *autonomous*.) If instead V is a function of (t, x) in some subset of $\mathbb{R} \times \mathbb{R}^n$, the resulting system is called *nonautonomous*; it can be thought of as a "time-dependent vector field" on a subset of $\mathbb{R}^n$. This problem shows that local existence, uniqueness, and smoothness for a nonautonomous system follow from the corresponding results for autonomous ones. Suppose $U \subset \mathbb{R}^n$ is an open set, $J \subset \mathbb{R}$ is an open interval, and $V\colon J \times U \to \mathbb{R}^n$ is a smooth map. For any $(t_0, x_0) \in J \times U$ and any sufficiently small $\varepsilon > 0$, show that there exists a neighborhood U_0 of x_0 in U and a unique smooth map $\psi^{t_0}\colon (t_0 - \varepsilon, t_0 + \varepsilon) \times U_0 \to U$ such that for each $x \in U_0$, the curve $\gamma(t) = \psi^{t_0}(t, x)$ is the unique solution on $(t_0 - \varepsilon, t_0 + \varepsilon)$ to the nonautonomous initial-value problem

$$\begin{aligned} \gamma^{i\prime}(t) &= V^i(t, \gamma(t)), \\ \gamma^i(t_0) &= x^i. \end{aligned}$$

[Hint: Replace this system of ODEs by an autonomous system in $\mathbb{R}^{n+1}$.]

13
Lie Derivatives

This chapter is devoted primarily to the study of a particularly important construction involving vector fields, the Lie derivative. This is a method of computing the "directional derivative" of one vector field with respect to another. We will see in this and later chapters that this construction has applications to flows, symplectic manifolds, Lie groups, and partial differential equations, among other subjects.

The Lie Derivative

We already know how to compute "directional derivatives" of functions on a manifold: Indeed, a tangent vector V_p is by definition an operator that acts on a smooth function f to give a number $V_p f$, which we interpret as the directional derivative of f in the direction V_p.

What about the directional derivative of a *vector field*? In Euclidean space, we have a perfectly good way of making sense of this: If V is a tangent vector at $p \in \mathbb{R}^n$, and W is a smooth vector field on $\mathbb{R}^n$, we can define the "directional derivative" of W in the direction of V as the vector

$$D_V W = \left.\frac{d}{dt}\right|_{t=0} W_{p+tV} = \lim_{t\to 0} \frac{W_{p+tV} - W_p}{t}. \tag{13.1}$$

A standard calculation using the chain rule shows that $D_V W$ can be calculated by applying V to each component of W separately:

$$D_V W(p) = VW^i(p) \left.\frac{\partial}{\partial x^i}\right|_p .$$

Unfortunately, as we will see, this does not define a coordinate-independent operation. If we search for a way to make invariant sense of (13.1) on a manifold, we will see very quickly what the problem is. To begin with, we can replace $p + tV$ by any curve $\gamma(t)$ that starts at p and whose initial tangent vector is V. But even with this substitution, the difference quotient still makes no sense because $W_{\gamma(t)}$ and $W_{\gamma(0)}$ are elements of different vector spaces ($T_{\gamma(t)}M$ and $T_{\gamma(0)}M$). We got away with it in Euclidean space because there is a canonical identification of each tangent space with $\mathbb{R}^n$ itself; but on a manifold there is no such identification. Thus there is no coordinate-invariant way to make sense of the directional derivative of W in the direction of the vector V.

Now suppose that V itself is a vector field instead of a single vector. In this case, we can use the flow of V to push values of W back to p and then differentiate. Thus, for any smooth vector fields V and W on a manifold M, let θ be the flow of V, and define a vector $(\mathcal{L}_V W)_p$ at each $p \in M$, called the *Lie derivative* of W with respect to V, by

$$(\mathcal{L}_V W)_p = \left.\frac{d}{dt}\right|_{t=0} (\theta_{-t})_* W_{\theta_t(p)} = \lim_{t \to 0} \frac{(\theta_{-t})_* W_{\theta_t(p)} - W_p}{t}, \tag{13.2}$$

provided the derivative exists. For small $t \neq 0$, the difference quotient makes sense at least, because θ_t is defined in a neighborhood of p and both $(\theta_{-t})_* W_{\theta_t(p)}$ and W_p are elements of T_pM.

Exercise 13.1. If $V \in \mathbb{R}^n$ and W is a smooth vector field on an open subset of $\mathbb{R}^n$, show that the directional derivative $D_V W(p)$ defined by (13.1) is equal to $(\mathcal{L}_{\widetilde{V}} W)_p$, where $\widetilde{V}$ is the vector field $\widetilde{V} = V^i \partial/\partial x^i$ with constant coefficients in standard coordinates.

This definition raises a number of questions: Does $(\theta_{-t})_* W_{\theta_t(p)}$ always depend differentiably on t, so that the derivative in (13.2) always exists? If so, does the assignment $p \mapsto (\mathcal{L}_V W)_p$ define a smooth vector field on M? And most importantly, is there a reasonable way to compute $(\mathcal{L}_V W)_p$, given that the only way to find the integral curves of V is to solve a system of ODEs, and most such systems cannot be solved explicitly? Fortunately, there are good answers to all these questions, but we will need to develop a few more tools in order to describe them.

Lie Brackets

We begin by defining another way to combine two vector fields to obtain a new vector field, seemingly unrelated to the Lie derivative.

Let V and W be smooth vector fields on a smooth manifold M. Given a smooth function $f\colon M \to \mathbb{R}$, we can apply V to f and obtain another smooth function Vf, to which we can then apply the vector W_p to obtain a real number $W_p(Vf)$; of course we can also do the same thing the other way around. Applying both of these operators to f and subtracting, we obtain an operator $[V,W]_p\colon C^\infty(M) \to \mathbb{R}$, called the *Lie bracket* of V and W, defined by

$$[V,W]_p f = V_p(Wf) - W_p(Vf).$$

Lemma 13.1. *For any two vector fields* $V, W \in \mathcal{T}(M)$*, the Lie bracket* $[V,W]_p$ *satisfies the following properties.*

(*a*) $[V,W]_p \in T_pM$.

(*b*) *The assignment* $p \mapsto [V,W]_p$ *defines a smooth vector field* $[V,W]$ *on* M*, which satisfies*

$$[V,W]f = VWf - WVf. \tag{13.3}$$

(*c*) *If* (x^i) *are any local coordinates on* M*, then* $[V,W]$ *has the coordinate expression*

$$[V,W] = \left(V^i \frac{\partial W^j}{\partial x^i} - W^i \frac{\partial V^j}{\partial x^i}\right)\frac{\partial}{\partial x^j}, \tag{13.4}$$

or more concisely,

$$[V,W] = (VW^j - WV^j)\frac{\partial}{\partial x^j}.$$

Proof. First we prove that $[V,W]_p$ is a tangent vector, i.e., a linear derivation on the space of germs of functions at p. Clearly $[V,W]_p f$ depends only on the values of f in a neighborhood of p, so it is well-defined on germs. As a map from $C^\infty(p)$ to $\mathbb{R}$, it is obviously linear over $\mathbb{R}$, so only the product rule needs to be checked. If f and g are smooth functions defined in a neighborhood of p, then

$$\begin{aligned}
[V,W]_p(fg) &= V_p(W(fg)) - W_p(V(fg)) \\
&= V_p(fWg + gWf) - W_p(fVg + gVf) \\
&= (V_pf)(W_pg) + f(p)V_p(Wg) + (V_pg)(W_pf) + g(p)V_p(Wf) \\
&\quad - (W_pf)(V_pg) - f(p)W_p(Vg) - (W_pg)(V_pf) - g(p)W_p(Vf) \\
&= f(p)(V_p(Wg) - W_p(Vg)) + g(p)(V_p(Wf) - W_p(Vf)) \\
&= f(p)[V,W]_pg + g(p)[V,W]_pf.
\end{aligned}$$

This shows that $[V,W]_p$ satisfies the product rule and therefore defines a tangent vector at p.

Formula (13.3) is immediate from the definition of the Lie bracket, and it follows from this that $[V,W]$ is a smooth vector field, because (13.3) defines a smooth function whenever f is smooth in an open subset of M.

To prove (c), we just write $V = V^i \partial/\partial x^i$ and $W = W^j \partial/\partial x^j$ in coordinates, and compute:

$$\begin{aligned}
[V,W]f &= V^i \frac{\partial}{\partial x^i}\left(W^j \frac{\partial f}{\partial x^j}\right) - W^j \frac{\partial}{\partial x^j}\left(V^i \frac{\partial f}{\partial x^i}\right) \\
&= V^i \frac{\partial W^j}{\partial x^i}\frac{\partial f}{\partial x^j} + V^i W^j \frac{\partial^2 f}{\partial x^i \partial x^j} - W^j \frac{\partial V^i}{\partial x^j}\frac{\partial f}{\partial x^i} - W^j V^i \frac{\partial^2 f}{\partial x^j \partial x^i} \\
&= V^i \frac{\partial W^j}{\partial x^i}\frac{\partial f}{\partial x^j} - W^j \frac{\partial V^i}{\partial x^j}\frac{\partial f}{\partial x^i},
\end{aligned}$$

where in the last step we have used the fact that mixed partial derivatives of a smooth function commute. Reversing the roles of the summation indices i and j in the second term, we obtain (13.4). $\square$

Exercise 13.2. For each of the following pairs of vector fields V, W defined on $\mathbb{R}^3$, compute the Lie bracket $[V,W]$.

(a) $V = y\dfrac{\partial}{\partial z} - 2xy^2\dfrac{\partial}{\partial y}$; $\quad W = \dfrac{\partial}{\partial y}$.

(b) $V = x\dfrac{\partial}{\partial y} - y\dfrac{\partial}{\partial x}$; $\quad W = y\dfrac{\partial}{\partial z} - z\dfrac{\partial}{\partial y}$.

(c) $V = x\dfrac{\partial}{\partial y} - y\dfrac{\partial}{\partial x}$; $\quad W = x\dfrac{\partial}{\partial y} + y\dfrac{\partial}{\partial x}$.

Lemma 13.2 (Properties of the Lie Bracket). *The Lie bracket satisfies the following identities:*

(*a*) Bilinearity: *For* $a_1, a_2 \in \mathbb{R}$, $[a_1V_1 + a_2V_2, W] = a_1[V_1,W] + a_2[V_2,W]$ *and* $[V, a_1W_1 + a_2W_2] = a_1[V,W_1] + a_2[V,W_2]$.

(*b*) Antisymmetry: $[V,W] = -[W,V]$.

(*c*) Jacobi Identity: $[V,[W,X]] + [W,[X,V]] + [X,[V,W]] = 0$.

Proof. Bilinearity and antisymmetry are obvious consequences of the definition. The proof of the Jacobi identity is just a computation:

$$\begin{aligned}
&[V,[W,X]]f + [W,[X,V]]f + [X,[V,W]]f \\
&\quad = V[W,X]f - [W,X]Vf + W[X,V]f \\
&\qquad\quad - [X,V]Wf + X[V,W]f - [V,W]Xf \\
&\quad = VWXf - VXWf - WXVf + XWVf + WXVf - WVXf \\
&\qquad - XVWf + VXWf + XVWf - XWVf - VWXf + WVXf.
\end{aligned}$$

In this last expression, all the terms cancel in pairs. $\square$

Proposition 13.3 (Naturality of the Lie Bracket). *Let $F\colon M \to N$ be a smooth map, and let $V_1, V_2 \in \mathfrak{T}(M)$ and $W_1, W_2 \in \mathfrak{T}(N)$ be vector fields such that V_i is F-related to W_i, $i = 1, 2$. Then $[V_1, V_2]$ is F-related to $[W_1, W_2]$. If F is a diffeomorphism, then $F_*[V_1, V_2] = [F_*V_1, F_*V_2]$.*

Proof. Using Lemma 3.17 and fact that V_i and W_i are F-related,

$$\begin{aligned} V_1V_2(f \circ F) &= V_1((W_2f) \circ F) \\ &= (W_1W_2f) \circ F. \end{aligned}$$

Similarly,

$$V_2V_1(f \circ F) = (W_2W_1f) \circ F.$$

Therefore,

$$\begin{aligned} [V_1, V_2](f \circ F) &= V_1V_2(f \circ F) - V_2V_1(f \circ F) \\ &= (W_1W_2f) \circ F - (W_2W_1f) \circ F \\ &= ([W_1, W_2]f) \circ F. \end{aligned}$$

The result then follows from the lemma. The statement when F is a diffeomorphism is an obvious consquence of the general case, because $W_i = F_*V_i$ in that case. □

Proposition 13.4. *Let N be an immersed submanifold of M, and suppose $V, W \in \mathfrak{T}(M)$. If V and W are tangent to N, then so is $[V, W]$.*

Proof. This is a local question, so we may replace N by an open subset of N that is embedded. Then Proposition 5.8 shows that a vector $X \in T_pM$ is in T_pN if and only if $Xf = 0$ whenever $f \in C^\infty(M)$ vanishes on N. Suppose f is such a function. Then the fact that V and W are tangent to N implies that $Vf|_N = Wf|_N = 0$, and so

$$[V, W]_pf = V_p(Wf) - W_p(Vf) = 0.$$

This shows that $[V, W]_p \in T_pN$, which was to be proved. □

We will see shortly that the Lie bracket $[V, W]$ is equal to the Lie derivative $\mathcal{L}_VW$, even though the two quantities are defined in ways that seem totally unrelated. Before doing so, we need to prove one more result, which is of great importance in its own right. If V is a smooth vector field on M, a point $p \in M$ is said to be a *singular point* for V if $V_p = 0$, and a *regular point* otherwise.

Theorem 13.5 (Canonical Form for a Regular Vector Field). *Let V be a smooth vector field on a smooth manifold M, and let $p \in M$ be a regular point for V. There exist coordinates (x^i) on some neighborhood of p in which V has the coordinate expression $\partial/\partial x^1$.*

Proof. By the way we have defined coordinate vector fields on a manifold, a coordinate chart (U, φ) will satisfy the conclusion of the theorem provided that $(\varphi^{-1})_*(\partial/\partial x^1) = V$, which will be true if and only if φ^{-1} takes lines parallel to the x^1 axis to the integral curves of V. The flow of V is ideally suited to this purpose.

Begin by choosing any coordinates (y^i) on a neighborhood U of p, with p corresponding to 0. By composing with a linear transformation, we may assume that $V_p = \partial/\partial y^1|_p$. Let $\theta\colon \mathcal{D}(V) \to M$ be the flow of V. There exists $\varepsilon > 0$ and a neighborhood $U_0 \subset U$ of p such that the product open set $(-\varepsilon, \varepsilon) \times U_0$ is contained in $\mathcal{D}(V)$ and is mapped by θ into U.

Let $S \subset \mathbb{R}^{n-1}$ be the set

$$S = \{(x^2, \ldots, x^n) : (0, x^2, \ldots, x^n) \in U_0\},$$

and define a smooth map $\psi\colon (-\varepsilon, \varepsilon) \times S \to U$ by

$$\psi(t, x^2, \ldots, x^n) = \theta(t, (0, x^2, \ldots, x^n)).$$

Geometrically, for each fixed $(x^2, \ldots, x^n)$, ψ maps the interval $(-\varepsilon, \varepsilon) \times \{(x^2, \ldots, x^n)\}$ to the integral curve through $(0, x^2, \ldots, x^n)$.

First we will show that ψ pushes $\partial/\partial t$ forward to V. We have

$$\begin{aligned} \left(\psi_* \left.\frac{\partial}{\partial t}\right|_{(t_0,x_0)}\right) f &= \left.\frac{\partial}{\partial t}\right|_{(t_0,x_0)} (f \circ \psi) \\ &= \left.\frac{\partial}{\partial t}\right|_{t=t_0} (f(\theta(t, (0, x_0)))) \\ &= V_{\psi(t_0,x_0)} f, \end{aligned} \tag{13.5}$$

where we have used the fact that $t \mapsto \theta(t, (0, x_0))$ is an integral curve of V. On the other hand, when restricted to $\{0\} \times S$, $\psi(0, x^2, \ldots, x^n) = \theta(0, (0, x^2, \ldots, x^n)) = (0, x^2, \ldots, x^n)$, so

$$\psi_* \left.\frac{\partial}{\partial x^i}\right|_{(0,0)} = \left.\frac{\partial}{\partial y^i}\right|_{(0,0)}.$$

Since $\psi_*\colon T_{(0,0)}((-\varepsilon, \varepsilon) \times S) \to T_pM$ takes a basis to a basis, it is an isomorphism. Therefore, by the inverse function theorem, there are neighborhoods W of $(0, 0)$ and Y of p such that $\psi\colon W \to Y$ is a diffeomorphism.

Let $\varphi = \psi^{-1}\colon Y \to W$. Equation (13.5) says precisely that V is equal to the coordinate vector field $\partial/\partial t$ in these coordinates. Renaming t to x^1, this is what we wanted to prove. □

This theorem implies that the integral curves of V near a regular point behave, up to diffeomorphism, just like the x^1-lines in $\mathbb{R}^n$, so that all of the interesting local behavior is concentrated near the singular points. Of course, the flow near singular points can exhibit a wide variety of behaviors, such as closed orbits surrounding the singular point, orbits converging exponentially or spiraling into the singular point as $t \to \infty$ or $-\infty$, and many more complicated phenomena, as one can see in any good differential equations text that treats systems of ODEs in the plane. This is the starting point for the subject of *smooth dynamical systems*, which is the study of the global and long-time behavior of flows of vector fields.

We are now in a position to prove the promised formula for computing Lie derivatives.

Theorem 13.6. *For any smooth vector fields V and W on a smooth manifold M, $\mathcal{L}_V W = [V, W]$.*

Proof. We will show that $(\mathcal{L}_V W)_p = [V, W]_p$ for each $p \in M$. We consider two cases.

Case I: *p is a regular point for V.* In this case, we can choose coordinates (x^i) near p such that $V = \partial/\partial x^1$ in coordinates. In this case, the flow of V is easy to compute explicitly:

$$\theta_t(x) = (x^1 + t, x^2, \dots, x^n).$$

Therefore, for each fixed t, the matrix of $(\theta_t)_*$ in these coordinates (the Jacobian matrix of θ_t) is the identity at every point. Consequently,

$$\begin{aligned}
(\theta_{-t})_* W_{\theta_t(p)} &= (\theta_{-t})_* \left(W^j(x^1 + t, x^2, \dots, x^n) \left.\frac{\partial}{\partial x^j}\right|_{\theta_t(p)} \right) \\
&= W^j(x^1 + t, x^2, \dots, x^n) \left.\frac{\partial}{\partial x^j}\right|_p .
\end{aligned}$$

Using the definition of the Lie derivative,

$$\begin{aligned}
(\mathcal{L}_V W)_p &= \left.\frac{d}{dt}\right|_{t=0} W^j(x^1 + t, x^2, \dots, x^n) \left.\frac{\partial}{\partial x^j}\right|_p \\
&= \frac{\partial W^j}{\partial x^1}(x^1, \dots, x^n) \left.\frac{\partial}{\partial x^j}\right|_p .
\end{aligned}$$

On the other hand, using the formula (13.4) for the Lie bracket in coordinates, $[V, W]_p$ is easily seen to be equal to the same expression.

Case II: *p is a singular point for V.* In this case, we cannot write down the flow explicitly. However, it does have the property that $\theta_t(p) = p$ for all $t \in \mathbb{R}$ (Lemma 12.8), and therefore $(\theta_{-t})_*$ maps T_pM to itself. Since the matrix entries of $(\theta_{-t})_* \colon T_pM \to T_pM$ in any coordinate system are partial derivatives of θ and therefore are smooth functions of t, it follows that the components of the T_pM-valued function $t \mapsto (\theta_{-t})_* W_{\theta_t(p)} = (\theta_{-t})_* W_p$ with respect to the coordinate basis are also smooth functions of t. If $t \mapsto X(t)$

is any smooth curve in T_pM, then for any smooth function f defined on a neighborhood of p,

$$\begin{aligned}\left(\frac{d}{dt}X(t)\right)f &= \left(\frac{d}{dt}X^j(t)\right)\frac{\partial f}{\partial x^j}(p)\\ &= \left(\frac{d}{dt}X^j(t)\frac{\partial f}{\partial x^j}(p)\right)\\ &= \frac{d}{dt}(X(t)f).\end{aligned}$$

Applying this to $(\theta_{-t})_*W_p$ and using the definition of the Lie derivative,

$$\begin{aligned}(\mathcal{L}_V W)_p f &= \left(\left.\frac{d}{dt}\right|_{t=0}(\theta_{-t})_*W_p\right)f\\ &= \left.\frac{d}{dt}\right|_{t=0}((\theta_{-t})_*W_p f)\\ &= \left.\frac{d}{dt}\right|_{t=0}(W_p(f\circ\theta_{-t}))\\ &= \left.\frac{d}{dt}\right|_{t=0}\left(W^j(p)\left.\frac{\partial}{\partial x^j}\right|_p(f\circ\theta_{-t})\right)\\ &= W^j(p)\left.\frac{\partial}{\partial t}\right|_{t=0}\left.\frac{\partial}{\partial x^j}\right|_{x=p} f(\theta_{-t}(x))\\ &= W^j(p)\left.\frac{\partial}{\partial x^j}\right|_{x=p}\left.\frac{\partial}{\partial t}\right|_{t=0} f(\theta_{-t}(x))\\ &= W^j(p)\left.\frac{\partial}{\partial x^j}\right|_p(-Vf)\\ &= -W_p(Vf).\end{aligned}$$

On the other hand, since $V_p = 0$, we have

$$[V,W]_p f = V_p(Wf) - W_p(Vf) = -W_p(Vf),$$

which is equal to $(\mathcal{L}_V W)_p f$. □

Corollary 13.7. *If V and W are as in the statement of the theorem, then*

(*a*) *The assignment $p \mapsto (\mathcal{L}_V W)_p$ is a smooth vector field on M.*

(*b*) $\mathcal{L}_V W = -\mathcal{L}_W V$.

(*c*) *If $F\colon M \to N$ is a diffeomorphism, then $F_*(\mathcal{L}_V W) = \mathcal{L}_{F_*V} F_*W$.*

Exercise 13.3. Prove this corollary.

Commuting Vector Fields

Two smooth vector fields are said to *commute* if $[V, W] \equiv 0$, or equivalently if $VWf = WVf$ for every smooth function f. One simple example of a pair of commuting vector fields is $\partial/\partial x^i$ and $\partial/\partial x^j$ in any coordinate system: because their component functions are constants, their Lie bracket is identically zero.

We will see that commuting vector fields are closely related to another important concept. Suppose W is a smooth vector field on M, and θ is a flow on M. We say W is *invariant* under θ if $(\theta_t)_* W = W$ on the image of θ_t. More explicitly, this means that

$$(\theta_t)_* W_p = W_{\theta_t(p)} \quad \text{for all } (t,p) \text{ in the domain of } \theta.$$

We will show that commuting vector fields are invariant under each other's flows. The key is the following somewhat more general result about F-related vector fields.

Lemma 13.8. *Suppose $F\colon M \to N$ is a smooth map, $V \in \mathcal{T}(M)$, and $W \in \mathcal{T}(N)$, and let θ be the flow of V and ψ the flow of W. Then V and W are F-related if and only if for each $t \in \mathbb{R}$, $\psi_t \circ F = F \circ \theta_t$ on the domain of θ_t:*

$$\begin{tikzcd} M \arrow[r, "F"] \arrow[d, "\theta_t"'] & N \arrow[d, "\psi_t"] \\ M \arrow[r, "F"'] & N \end{tikzcd}$$

Proof. The commutativity of the diagram means that the following holds for all (t,p) in the domain of θ:

$$\psi_t \circ F(p) = F \circ \theta_t(p).$$

If we let $\mathcal{D}_p \subset \mathbb{R}$ denote the domain of $\theta^{(p)}$, this is equivalent to

$$\psi^{(F(p))}(t) = F \circ \theta^{(p)}(t), \quad t \in \mathcal{D}_p. \tag{13.6}$$

Suppose first that V and W are F-related. If we define $\gamma\colon \mathcal{D}_p \to N$ by $\gamma = F \circ \theta^{(p)}$, then

$$\begin{aligned} \gamma'(t) &= (F \circ \theta^{(p)})'(t) \\ &= F_*(\theta^{(p)\prime}(t)) \\ &= F_* V_{\theta^{(p)}(t)} \\ &= W_{F\circ\theta^{(p)}(t)} \\ &= W_{\gamma(t)}, \end{aligned}$$

so γ is an integral curve of W starting at $F \circ \theta^{(p)}(0) = F(p)$. By uniqueness of integral curves, therefore, the maximal integral curve $\psi^{(F(p))}$ must be defined at least on the interval $\mathcal{D}_p$, and $\gamma(t) = \psi^{(F(p))}(t)$ on that interval. This proves (13.6).

Conversely, if (13.6) holds, then for each $p \in M$ we have

$$\begin{aligned} F_* V_p &= F_*(\theta^{(p)\prime}(0)) \\ &= (F \circ \theta^{(p)})'(0) \\ &= \psi^{(F(p))\prime}(0) \\ &= W_{F(p)}, \end{aligned}$$

which shows that V and W are F-related. □

Proposition 13.9. *Let V and W be smooth vector fields on M, with flows θ and ψ, respectively. The following are equivalent:*

(*a*) *V and W commute.*

(*b*) $\mathcal{L}_V W = \mathcal{L}_W V = 0$.

(*c*) *W is invariant under the flow of V.*

(*d*) *V is invariant under the flow of W.*

(*e*) *$\theta_s \circ \psi_t = \psi_t \circ \theta_s$ wherever either side is defined.*

Proof. Clearly (a) and (b) are equivalent because $\mathcal{L}_V W = [V, W] = -\mathcal{L}_W V$. Part (c) means that $(\theta_{-t})_* W_{\theta_t(p)} = W_p$ whenever $(-t, p)$ is in the domain of θ, which obviously implies (b) directly from the definition of $\mathcal{L}_V W$. The same argument shows that (d) implies (b).

To prove that (b) implies (c), let $p \in M$ be arbitrary, let $\mathcal{D}_p \subset \mathbb{R}$ denote the domain of the integral curve $\theta^{(p)}$, and consider the map $X\colon \mathcal{D}_p \to T_pM$ given by the time-dependent vector

$$X(t) = (\theta_{-t})_* \left(W_{\theta_t(p)}\right) \in T_pM. \tag{13.7}$$

This can be considered as a smooth curve in the vector space T_pM. We will show that $X(t)$ is independent of t. Since $X(0) = W_p$, this implies that $X(t) = W_p$ for all $t \in \mathcal{D}_p$, which says that W is invariant under θ_t.

The assumption that $\mathcal{L}_V W = 0$ means precisely that the t-derivative of (13.7) is zero when $t = 0$; we need to show that this derivative is zero for all values of t. Making the change of variables $t = t_0 + s$, we obtain

$$\begin{aligned} X'(t_0) &= \left.\frac{d}{dt}\right|_{t=t_0} (\theta_{-t})_* W_{\theta_t(p)} \\ &= \left.\frac{d}{ds}\right|_{s=0} (\theta_{-t_0-s})_* W_{\theta_{s+t_0}(p)} \\ &= \left.\frac{d}{ds}\right|_{s=0} (\theta_{-t_0})_* (\theta_{-s})_* W_{\theta_s(\theta_{t_0}(p))} \\ &= (\theta_{-t_0})_* \left.\frac{d}{ds}\right|_{s=0} (\theta_{-s})_* W_{\theta_s(\theta_{t_0}(p))} \\ &= (\theta_{-t_0})_* (\mathcal{L}_V W)_{\theta_{t_0}(p)} = 0. \end{aligned} \tag{13.8}$$

(The equality on the next-to-last line follows because $(\theta_{-t_0})_* \colon T_{\theta_{t_0}(p)}M \to T_pM$ is a linear map that is independent of s.) The same proof also shows that (b) implies (d).

To prove that (c) and (e) are equivalent, we let M_t denote the domain of θ_t and use Lemma 13.8 applied to the map $F = \theta_t \colon M_t \to \theta_t(M_t)$. According to that lemma, $\theta_t \circ \psi_s = \psi_s \circ \theta_t$ on the set where $\theta_t \circ \psi_s$ is defined if and only if $(\theta_t)_* W = W$ on $\theta_t(M_t)$, which is to say if and only if W is invariant under θ. By reversing the roles of V and W, we see that this is also true on the set where $\psi_s \circ \theta_t$ is defined. □

As we mentioned above, one example of a family of commuting vector fields is given by the coordinate vector fields $\partial/\partial x^i$, $i = 1, \ldots, n$. The next theorem shows that up to diffeomorphism, any collection of independent commuting vector fields is of this form locally.

Theorem 13.10 (Canonical Form for Commuting Vector Fields). *Let M be a smooth n-manifold, and let $V_1, \ldots, V_k$ be smooth vector fields on an open subset of M whose values are linearly independent at each point. Then the following are equivalent:*

(*a*) *There exist coordinates (x^i) in a neighborhood of each point such that $V_i = \partial/\partial x^i$, $i = 1, \ldots, k$.*

(*b*) $[V_i, V_j] \equiv 0$ *for all i and j.*

Proof. The fact that (a) implies (b) is obvious because the coordinate vector fields commute and the Lie bracket is coordinate-independent.

To prove the converse, suppose $V_1, \ldots, V_k$ are vector fields satisfying (b). The basic outline of the proof is entirely analogous to that of the canonical form theorem for one nonvanishing vector field (Theorem 13.5), except that we have to do a bit of work to make use of the hypothesis that the vector fields commute.

Choose coordinates (y^i) on a neighborhood U of p such that p corresponds to 0 and $V_i = \partial/\partial y^i$ at 0 for $i = 1, \ldots, k$. Let θ_i be the flow of

V_i for $i = 1, \dots, k$. There exists $\varepsilon > 0$ and a neighborhood W of p such that the composition $(\theta_k)_{t_k} \circ (\theta_{k-1})_{t_{k-1}} \circ \dots \circ (\theta_1)_{t_1}$ is defined on W and maps W into U whenever $|t_1|, \dots, |t_k|$ are all less than ε. (Just choose $\varepsilon_1 > 0$ and $U_1 \subset U$ such that θ_1 maps $(-\varepsilon_1, \varepsilon_1) \times U_1$ into U, and then inductively choose ε_i and U_i such that θ_i maps $(-\varepsilon_i, \varepsilon_i) \times U_i$ into U_{i-1}. Taking $\varepsilon = \min\{\varepsilon_i\}$ and $W = U_k$ does the trick.)

As in the proof of the canonical form theorem for one nonvanishing vector field, let

$$S = \{(x^{k+1}, \dots, x^n) : (0, \dots, 0, x^{k+1}, \dots, x^n) \in W\}.$$

Define $\psi\colon (-\varepsilon, \varepsilon)^k \times W \to U$ by

$$\begin{aligned}\psi(x^1, &\dots, x^k, x^{k+1}, \dots, x^n) \\ &= (\theta_k)_{x^k} \circ \dots \circ (\theta_1)_{x^1}(0, (0, \dots, 0, x^{k+1}, \dots, x^n)).\end{aligned}$$

We will show first that

$$\psi_* \frac{\partial}{\partial x^i} = V_i, \quad i = 1, \dots, k.$$

Because all the flows θ_i commute with each other, we have

$$\begin{aligned}\left(\psi_* \frac{\partial}{\partial x^i}\right) f &= \frac{\partial}{\partial x^i} \left(f(\psi(x^1, \dots, x^n))\right) \\ &= \frac{\partial}{\partial x^i} \left(f((\theta_k)_{x^k} \circ \dots \circ (\theta_1)_{x^1}(0, (0, \dots, 0, x^{k+1}, \dots, x^n)))\right) \\ &= \frac{\partial}{\partial x^i} \Big(f((\theta_i)_{x^i} \circ (\theta_k)_{x^k} \circ \cdots \\ &\qquad \circ \widehat{(\theta_i)_{x^i}} \circ \dots \circ (\theta_1)_{x^1}(0, (0, \dots, 0, x^{k+1}, \dots, x^n))\Big),\end{aligned}$$

where the hat means that $(\theta_i)_{x^i}$ is omitted. Now, for any $p \in M$, $t \mapsto (\theta_i)_t(p)$ is an integral curve of V_i, so this last expression is equal to $(V_i)_{\psi(x)} f$, which proves the claim.

Next we will show that ψ_* is invertible at p. The computation above shows that for $i = 1, \dots, k$,

$$\psi_* \left. \frac{\partial}{\partial x^i} \right|_0 = V_p = \left. \frac{\partial}{\partial y^i} \right|_p.$$

On the other hand, since $\psi(0, \dots, 0, x^{k+1}, \dots, x^n) = (0, \dots, 0, x^{k+1}, \dots, x^n)$, it follows immediately that

$$\psi_* \left. \frac{\partial}{\partial x^i} \right|_0 = \left. \frac{\partial}{\partial y^i} \right|_p$$

for $i = k+1, \dots, n$ as well. Thus ψ_* takes a basis to a basis, and is therefore a diffeomorphism in a neighborhood of 0. It follows that $\varphi = \psi^{-1}$ is the desired coordinate map. □

Lie Derivatives of Tensor Fields

The Lie derivative operation can be extended to tensors of arbitrary rank. As usual, we focus on covariant tensors; the analogous results for contravariant or mixed tensors require only minor modifications.

Let X be a smooth vector field on a smooth manifold M, and let θ be its flow. Near any $p \in M$, if t is sufficiently close to zero, θ_t is a diffeomorphism from a neighborhood of p to a neighborhood of $\theta_t(p)$. Thus θ_t^* pulls back smooth tensor fields near $\theta_t(p)$ to ones near p.

Given a covariant tensor field τ on M, we define the *Lie derivative* of τ with respect to X, denoted by $\mathcal{L}_X\tau$, as

$$(\mathcal{L}_X\tau)_p = \left.\frac{\partial}{\partial t}\right|_{t=0} (\theta_t^*\tau)_p = \lim_{t\to 0} \frac{\theta_t^*\tau_{\theta_t(p)} - \tau_p}{t}.$$

Because the expression being differentiated lies in $T^k(T_pM)$ for all t, $(\mathcal{L}_X\tau)_p$ makes sense as an element of $T^k(T_pM)$. We will show below that $\mathcal{L}_X\tau$ is actually a smooth tensor field. First we prove the following important properties of Lie derivatives of tensors.

Proposition 13.11. *Let M be a smooth manifold. Suppose X, Y are smooth vector fields on M, σ, τ are smooth covariant tensor fields, ω, η are differential forms, and f is a smooth function* (*thought of as a* 0*-tensor field*)*.*

(*a*) $\mathcal{L}_X f = Xf$.

(*b*) $\mathcal{L}_X(f\sigma) = (\mathcal{L}_X f)\sigma + f\mathcal{L}_X\sigma$.

(*c*) $\mathcal{L}_X(\sigma\otimes\tau) = (\mathcal{L}_X\sigma)\otimes\tau + \sigma\otimes\mathcal{L}_X\tau$.

(*d*) $\mathcal{L}_X(\omega\wedge\eta) = \mathcal{L}_X\omega\wedge\eta + \omega\wedge\mathcal{L}_X\eta$.

(*e*) $\mathcal{L}_X(Y \lrcorner\, \omega) = (\mathcal{L}_X Y)\lrcorner\, \eta + Y \lrcorner\, \mathcal{L}_X\omega$.

(*f*) *For any smooth vector fields* $Y_1,\dots,Y_k$,

$$\begin{aligned}\mathcal{L}_X(\sigma(Y_1,\dots,Y_k)) &= (\mathcal{L}_X\sigma)(Y_1,\dots,Y_k)\\ &\quad + \sigma(\mathcal{L}_X Y_1,\dots,Y_k) + \cdots + \sigma(Y_1,\dots,\mathcal{L}_X Y_k).\end{aligned} \tag{13.9}$$

Proof. The first assertion is just a reinterpretation of the definition in the case of a 0-tensor. Because $\theta_t^* f = f\circ\theta_t$, the definition implies

$$\mathcal{L}_X f(p) = \left.\frac{\partial}{\partial t}\right|_{t=0} f(\theta_t(p)) = Xf(p).$$

The proofs of (b), (c), (d), (e), and (f) are essentially the same, so we will prove (c) and leave the others to you.

$$\begin{aligned}
(\mathcal{L}_X(\sigma\otimes\tau))_p &= \lim_{t\to 0}\frac{\theta_t^*(\sigma\otimes\tau)_{\theta_t(p)}-(\sigma\otimes\tau)_p}{t}\\
&= \lim_{t\to 0}\frac{\theta_t^*\sigma_{\theta_t(p)}\otimes\theta_t^*\tau_{\theta_t(p)}-\sigma_p\otimes\tau_p}{t}\\
&= \lim_{t\to 0}\frac{\theta_t^*\sigma_{\theta_t(p)}\otimes\theta_t^*\tau_{\theta_t(p)}-\theta_t^*\sigma_{\theta_t(p)}\otimes\tau_p}{t}\\
&\qquad +\lim_{t\to 0}\frac{\theta_t^*\sigma_{\theta_t(p)}\otimes\tau_p-\sigma_p\otimes\tau_p}{t}\\
&= \lim_{t\to 0}\theta_t^*\sigma_{\theta_t(p)}\otimes\frac{\theta_t^*\tau_{\theta_t(p)}-\tau_p}{t}+\lim_{t\to 0}\frac{\theta_t^*\sigma_{\theta_t(p)}-\sigma_p}{t}\otimes\tau_p\\
&= \sigma_p\otimes(\mathcal{L}_X\tau)_p+(\mathcal{L}_X\sigma)_p\otimes\tau_p.
\end{aligned}$$

The other parts are similar, and are left as an exercise. □

Exercise 13.4. Complete the proof of the preceding proposition.

Corollary 13.12. *If X is a smooth vector field and σ is a smooth covariant tensor field, then $\mathcal{L}_X\sigma$ can be computed by the following expression:*

$$\begin{aligned}
(\mathcal{L}_X\sigma)(Y_1,\dots,Y_k)) = X(\sigma(Y_1,\dots,Y_k) - \sigma([X,Y_1],Y_2,\dots,Y_k) - \dots\\
- \sigma(Y_1,\dots,Y_{k-1},[X,Y_k]). \quad (13.10)
\end{aligned}$$

It follows that $\mathcal{L}_X\sigma$ is smooth.

Proof. Formula (13.10) is obtained simply by solving (13.9) for $\mathcal{L}_X\sigma$, and replacing $\mathcal{L}_Xf$ by Xf and $\mathcal{L}_XY_i$ by $[X,Y_i]$. It then follows immediately that $\mathcal{L}_X\sigma$ is smooth, because its action on smooth vector fields yields a smooth function. □

Corollary 13.13. *If $f\in C^\infty(M)$, then $\mathcal{L}_X(df)=d(\mathcal{L}_Xf)$.*

Proof. Using (13.10), we compute

$$\begin{aligned}
(\mathcal{L}_X df)(Y) &= X(df(y)) - df[X,Y]\\
&= XYf-[X,Y]f\\
&= XYf-(XYf-YXf)\\
&= YXf\\
&= d(Xf)(Y)\\
&= d(\mathcal{L}_Xf)(Y).
\end{aligned}$$

□

Differential Forms

In the case of differential forms, the exterior derivative yields a much more powerful formula for computing Lie derivatives. Although Corollary 13.12 gives a general formula for computing the Lie derivative of any tensor field, this formula has a serious drawback: In order to calculate what $\mathcal{L}_X\sigma$ does to vectors $Y_1, \ldots, Y_k$ at a point $p \in M$, it is necessary first to extend the vectors to vector *fields* in a neighborhood of p. The formula in the next proposition overcomes this disadvantage.

Proposition 13.14. *For any vector field X and any differential k-form ω on a smooth manifold M,*

$$\mathcal{L}_X\omega = X \lrcorner (d\omega) + d(X \lrcorner \omega). \tag{13.11}$$

Proof. The proof is by induction on k. We begin with a 0-form f, in which case

$$X \lrcorner (df) + d(X \lrcorner f) = X \lrcorner df = df(X) = Xf = \mathcal{L}_X f,$$

which is (13.11).

Any 1-form can be written locally as a sum of terms of the form $u\,dv$ for smooth functions u and v, so to prove (13.11) for 1-forms, it suffices to consider the case $\omega = u\,dv$. In this case, using Proposition 13.11(d) and Corollary 13.13, the left-hand side of (13.11) reduces to

$$\begin{aligned}\mathcal{L}_X(u\,dv) &= (\mathcal{L}_X u)dv + u(\mathcal{L}_X dv)\\ &= (Xu)dv + u\,d(Xv).\end{aligned}$$

On the other hand, using the fact that interior multiplication is an antiderivation, the right-hand side is

$$\begin{aligned}X \lrcorner d(u\,dv) + d(X \lrcorner (u\,dv)) &= X \lrcorner (du \wedge dv) + d(uXv)\\ &= (X \lrcorner du) \wedge dv - du \wedge (X \lrcorner dv)\\ &\qquad + u\,d(Xv) + (Xv)du\\ &= (Xu)dv - (Xv)du + u\,d(Xv) + (Xv)du.\end{aligned}$$

(Remember that $X \lrcorner du = du(X) = Xu$, and a wedge product with a 0-form is just ordinary multiplication.) After cancelling the two $(Xv)du$ terms, this is equal to $\mathcal{L}_X(u\,dv)$.

Now let $k > 1$, and suppose (13.11) has been proved for forms of degree less than k. Let ω be an arbitrary k-form, written in local coordinates as

$$\omega = {\sum_I}' \omega_I\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

Writing $\alpha = \omega_I\, dx^{i_1}$ and $\beta = dx^{i_2} \wedge \cdots \wedge dx^{i_k}$, we see that ω can be written as a sum of terms of the form $\alpha \wedge \beta$, where α is a 1-form and β is a $(k-1)$-form. For such a term, Proposition 13.11(d) and the induction hypothesis imply

$$\begin{aligned}\mathcal{L}_X(\alpha\wedge\beta) &= (\mathcal{L}_X\alpha)\wedge\beta+\alpha\wedge(\mathcal{L}_X\beta)\\ &= (X \lrcorner\, d\alpha + d(X \lrcorner\, \alpha))\wedge\beta+\alpha\wedge(X \lrcorner\, d\beta + d(X \lrcorner\, \beta)).\end{aligned} \tag{13.12}$$

On the other hand, using the fact that both d and i_X are antiderivations, we compute

$$\begin{aligned}X \lrcorner\, d(\alpha\wedge\beta) &+ d(X \lrcorner\,(\alpha\wedge\beta))\\ &= X \lrcorner\,(d\alpha\wedge\beta-\alpha\wedge d\beta)+d((X \lrcorner\,\alpha)\wedge\beta-\alpha\wedge(X \lrcorner\,\beta))\\ &= (X \lrcorner\, d\alpha)\wedge\beta+d\alpha\wedge(X \lrcorner\,\beta)-(X \lrcorner\,\alpha)\wedge d\beta\\ &\quad +\alpha\wedge(X \lrcorner\, d\beta)+d(X \lrcorner\,\alpha)\wedge\beta+(X \lrcorner\,\alpha)\wedge d\beta\\ &\quad -d\alpha\wedge(X \lrcorner\,\beta)+\alpha\wedge d(X \lrcorner\,\beta).\end{aligned}$$

After the obvious cancellations are made, this is equal to (13.12). □

Corollary 13.15. *If X is a vector field and ω is a differential form, then*

$$\mathcal{L}_X(d\omega) = d(\mathcal{L}_X\omega).$$

Proof. This follows from the preceding proposition and the fact that $d^2 = 0$:

$$\begin{aligned}\mathcal{L}_X d\omega &= X \lrcorner\, d(d\omega) + d(X \lrcorner\, d\omega)\\ &= d(X \lrcorner\, d\omega);\\ d\mathcal{L}_X\omega &= d(X \lrcorner\, d\omega + d(X \lrcorner\,\omega))\\ &= d(X \lrcorner\, d\omega).\end{aligned}$$

□

As promised, Proposition 13.14 gives a formula for the Lie derivative of a differential form that can be computed easily in local coordinates, without having to go to the trouble of letting the form act on vector fields. In fact, this leads to an easy algorithm for computing Lie derivatives of *arbitrary* tensor fields, since any tensor field can be written locally as a linear combination of tensor products of 1-forms. This is easiest to illustrate with an example.

Example 13.16. Suppose T is an arbitrary smooth symmetric 2-tensor field on a smooth manifold M, and let Y be a smooth vector field. We will compute the Lie derivative $\mathcal{L}_Y T$ in coordinates (x^i). First, we observe that $\mathcal{L}_Y dx^i = d(Y \lrcorner\, dx^i) + Y \lrcorner\, d(dx^i) = dY^i$. Therefore,

$$\begin{aligned}\mathcal{L}_Y T &= \mathcal{L}_Y(T_{ij})dx^i\otimes dx^j + T_{ij}(\mathcal{L}_Y dx^i)\otimes dx^j + T_{ij}\,dx^i\otimes(\mathcal{L}_Y dx^j)\\ &= YT_{ij}\,dx^i + T_{ij}\,dY^i\otimes dx^j + T_{ij}\,dx^i\otimes dY^j\\ &= \left(YT_{ij}+T_{jk}\frac{\partial Y^k}{\partial x^i}+T_{ik}\frac{\partial Y^k}{\partial x^j}\right)dx^i\otimes dx^j.\end{aligned}$$

Applications

What is the meaning of the Lie derivative of a tensor field with respect to a vector field X? We have already seen that the Lie derivative of a vector field Y with respect to X is zero if and only if Y is invariant along the flow of X. It turns out that the Lie derivative of a covariant tensor field has exactly the same interpretation. We say that a tensor field σ is *invariant* under a flow θ if $\theta_t^*\sigma = \sigma$ on the domain of θ_t.

The next lemma shows how the Lie derivative can be used to compute t-derivatives at times other than $t = 0$; it is a generalization of formula (13.8) to tensor fields.

Lemma 13.17. *Let M be a smooth manifold, $X \in \mathcal{T}(M)$, and let θ be the flow of X. For any smooth covariant tensor field τ and any (t_0, p) in the domain of θ,*

$$\left.\frac{d}{dt}\right|_{t=t_0} \theta_t^*(\tau_{\theta_t(p)}) = \theta_{t_0}^*(\mathcal{L}_X\tau)_{\theta_{t_0}(p)}.$$

Proof. Just as in the proof of Proposition 13.9, the change of variables $t = t_0 + s$ yields

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=t_0} \theta_t^*(\tau_{\theta_t(p)}) &= \left.\frac{d}{ds}\right|_{s=0} (\theta_{t_0+s})^*\tau_{\theta_{s+t_0}(p)} \\
&= \left.\frac{d}{ds}\right|_{s=0} (\theta_{t_0})^*(\theta_s)^*\tau_{\theta_s(\theta_{t_0}(p))} \\
&= (\theta_{t_0})^* \left.\frac{d}{ds}\right|_{s=0} (\theta_s)^*\tau_{\theta_s(\theta_{t_0}(p))} \\
&= (\theta_{t_0})^*(\mathcal{L}_V W)_{\theta_{t_0}(p)}.
\end{aligned}$$

□

Proposition 13.18. *Let M be a smooth manifold and let $X \in \mathcal{T}(M)$. A smooth covariant tensor field τ is invariant under the flow of X if and only if $\mathcal{L}_X\tau = 0$.*

Proof. Let θ denote the flow of X. If τ is invariant under θ, then $\theta_t^*\tau = \tau$ for all t. Inserting this into the definition of the Lie derivative, we see immediately that $\mathcal{L}_X\tau = 0$.

Conversely, suppose $\mathcal{L}_X\tau = 0$. For any $p \in M$, let $\mathcal{D}_p$ denote the domain of $\theta^{(p)}$, and consider the smooth curve $T\colon \mathcal{D}_p \to T^k(T_pM)$ defined by

$$T(t) = \theta_t^*(\tau_{\theta_t(p)}).$$

Lemma 13.17 shows that $T'(t) = 0$ for all $t \in \mathcal{D}_p$. Because $\mathcal{D}_p$ is a connected interval containing zero, this implies that $T(t) = T(0) = \tau_p$ for all $t \in \mathcal{D}_p$.

This is the same as

$$\theta_t^*(\tau_{\theta_t(p)}) = \tau_p,$$

which says precisely that τ is invariant under θ. $\square$

Killing Fields

Let (M, g) be a Riemannian manifold. A vector field Y on M is called a *Killing field* for g if g is invariant under the flow of Y. By Proposition 13.18, this is the case if and only if $\mathcal{L}_Y g = 0$.

Example 13.16 applied to the case $T = g$ gives the following coordinate expression for $\mathcal{L}_Y g$:

$$(\mathcal{L}_Y g)\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = Y g_{ij} + g_{jk}\frac{\partial Y^k}{\partial x^i} + g_{ik}\frac{\partial Y^k}{\partial x^j}. \tag{13.13}$$

Example 13.19 (Euclidean Killing Fields). Let $\overline{g}$ be the Euclidean metric on $\mathbb{R}^n$. In standard coordinates, the condition for a vector field to be a Killing field with respect to $\overline{g}$ reduces to

$$\frac{\partial Y^j}{\partial x^i} + \frac{\partial Y^i}{\partial x^j} = 0.$$

It is easy to check that all constant vector fields satisfy this equation, as do the vector fields

$$x^i \frac{\partial}{\partial x^j} - x^j \frac{\partial}{\partial x^i},$$

which generate rotations in the (x^i, x^j)-plane.

The Divergence

For our next application, we let (M, g) be an oriented Riemannian n-manifold. Recall that the *divergence* of a smooth vector field $X \in \mathcal{T}(M)$ is the smooth function $\operatorname{div} X$ characterized by

$$(\operatorname{div} X) dV_g = d(X \lrcorner\, dV_g).$$

Now we can give a geometric interpretation to the divergence, which explains the choice of the term "divergence." Observe that formula (13.11) for the Lie derivative of a differential form implies

$$\mathcal{L}_X dV_g = X \lrcorner\, d(dV_g) + d(X \lrcorner\, dV_g) = (\operatorname{div} X) dV_g,$$

because the exterior derivative of any n-form on an n-manifold is zero.

A flow θ on M is said to be *volume preserving* if for every compact domain of integration $D \subset M$ and every $t \in \mathbb{R}$ such that D is contained in the domain of θ_t, $\mathrm{Vol}(\theta_t(D)) = \mathrm{Vol}(D)$. It is *volume increasing* if for any such D with positive volume, $\mathrm{Vol}(\theta_t(D))$ is a strictly increasing function of t, and *volume decreasing* if it is strictly decreasing. Note that the properties of flow domains ensure that, if D is contained in the domain of θ_t for some t, then the same is true for all times between 0 and t. The next proposition shows that the divergence of a vector field is a quantitative measure of the tendency of its flow to "spread out" or diverge.

Proposition 13.20. *Let M be an oriented Riemannian manifold and let $X \in \mathcal{T}(M)$.*

(*a*) *The flow of X is volume preserving if and only if* $\operatorname{div} X \equiv 0$.

(*b*) *If* $\operatorname{div} X > 0$, *then the flow of X is volume increasing, and if* $\operatorname{div} X < 0$, *then it is volume decreasing.*

Proof. Let θ be the flow of X, and for each t let M_t be the domain of θ_t. If D is a compact domain of integration contained in M_t, then

$$\mathrm{Vol}(\theta_t(D)) = \int_{\theta_t(D)} dV_g = \int_D \theta_t^* dV_g.$$

Because the integrand is a smooth function of t, we can differentiate this expression with respect to t by differentiating under the integral sign. (Strictly speaking, we should use a partition of unity to express the integral as a sum of integrals over domains in $\mathbb{R}^n$, and then differentiate under the integral signs there. The details are left to you.) Using Lemma 13.17, we obtain

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=t_0} \mathrm{Vol}(\theta_t(D)) &= \int_D \left.\frac{\partial}{\partial t}\right|_{t=t_0} (\theta_t^* dV_g) \\
&= \int_D \theta_{t_0}^* (\mathcal{L}_X dV_g) \\
&= \int_D \theta_{t_0}^* ((\operatorname{div} X) dV_g) \\
&= \int_{\theta_{t_0}(D)} (\operatorname{div} X) dV_g.
\end{aligned}$$

It follows that $\operatorname{div} X \equiv 0$ implies that $\mathrm{Vol}(\theta_t(D))$ is a constant function of t, while $\operatorname{div} X > 0$ or $\operatorname{div} X < 0$ implies that it is strictly increasing or strictly decreasing, respectively.

Now assume that θ is volume preserving. If $\operatorname{div} X \neq 0$ at some point $p \in M$, then there is some open set U containing p on which $\operatorname{div} X$ does not change sign. If $\operatorname{div} X > 0$ on U, then X generates a volume increasing flow on U by the argument above. In particular, for any coordinate ball

B such that $\overline{B} \subset U$ and any $t > 0$ sufficiently small that $\theta_t(B) \subset U$, we have $\mathrm{Vol}(\theta_t(B)) > \mathrm{Vol}(B)$, which contradicts the assumption that θ is volume preserving. The argument in the case $\operatorname{div} X < 0$ is exactly analogous. Therefore $\operatorname{div} X \equiv 0$. □

Symplectic Manifolds

Let (M, ω) be a symplectic manifold. (Recall that this means a smooth manifold M endowed with a symplectic form ω, which is a closed nondegenerate 2-form.) One of the most important constructions on symplectic manifolds is a symplectic analogue of the gradient, defined as follows. Because of the nondegeneracy of ω, the bundle map $\widetilde{\omega}\colon TM \to T^*M$ given by $\widetilde{\omega}(X)(Y) = \omega(X, Y)$ is an isomorphism. For any smooth function $f \in C^\infty(M)$, we define the *Hamiltonian vector field* of f to be the vector field X_f defined by

$$X_f = \widetilde{\omega}^{-1}(df),$$

so X_f is characterized by

$$\omega(X_f, Y) = df(Y) = Yf$$

for any vector field Y. Another way to write this is

$$X_f \lrcorner\, \omega = df.$$

Example 13.21. On $\mathbb{R}^{2n}$ with the standard symplectic form $\omega = \sum_{i=1}^n dx^i \wedge dy^i$, X_f can be computed explicitly as follows. Writing

$$X_f = \sum_{i=1}^n A^i \frac{\partial}{\partial x^i} + B^i \frac{\partial}{\partial y^i}$$

for some coefficient functions (A^i, B^i) to be determined, we compute

$$\begin{aligned} X_f \lrcorner\, \omega &= \sum_{j=1}^n \left(A^j \frac{\partial}{\partial x^j} + B^j \frac{\partial}{\partial y^j} \right) \lrcorner \sum_{i=1}^n dx^i \wedge dy^i \\ &= \sum_{i=1}^n A^i\, dy^i - B^i\, dx^i. \end{aligned}$$

(When working with the standard symplectic form, like the Euclidean metric, it is usually necessary to insert explicit summation signs.) On the other hand,

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x^i} dx^i + \frac{\partial f}{\partial y^i} dy^i.$$

Setting these two expressions equal to each other, we find that $A^i = \partial f/\partial y^i$ and $B^i = -\partial f/\partial x^i$, which yields the following formula for the Hamiltonian vector field of f:

$$X_f = \sum_{i=1}^{n} \frac{\partial f}{\partial y^i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial y^i}. \tag{13.14}$$

Although the definition of the Hamiltonian vector field is formally analogous to that of the gradient on a Riemannian manifold, Hamiltonian vector fields differ from gradients in some very significant ways, as the next lemma shows.

Proposition 13.22 (Properties of Hamiltonian Vector Fields). *Let (M, ω) be a symplectic manifold and let $f \in C^\infty(M)$.*

(*a*) *f is constant along the flow of X_f, i.e., if θ is the flow, then $f \circ \theta_t(p) = f(p)$ for all (t, p) in the domain of θ.*

(*b*) *At each regular point of f, the Hamiltonian vector field X_f is tangent to the level set of f.*

Proof. Both assertions follow from the fact that

$$X_f f = df(X_f) = \omega(X_f, X_f) = 0$$

because ω is alternating. □

A vector field X on M is said to be *symplectic* if ω is invariant under the flow of X. It is said to be *Hamiltonian* (or *globally Hamiltonian*) if there exists a smooth function f such that $X = X_f$, and *locally Hamiltonian* if every point p has a neighborhood on which X is Hamiltonian. Clearly every globally Hamiltonian vector field is locally Hamiltonian.

Proposition 13.23 (Hamiltonian and Symplectic Vector Fields). *Let (M, ω) be a symplectic manifold. A smooth vector field on M is symplectic if and only if it is locally Hamiltonian. Every locally Hamiltonian vector field on M is globally Hamiltonian if and only if $H^1_{dR}(M) = 0$.*

Proof. By Proposition 13.18, a vector field X is symplectic if and only if $\mathcal{L}_X\omega = 0$. Using formula (13.11) for the Lie derivative of a differential form, we compute

$$\mathcal{L}_X\omega = d(X \lrcorner\, \omega) + X \lrcorner\, (d\omega) = d(X \lrcorner\, \omega). \tag{13.15}$$

Therefore X is symplectic if and only if the 1-form $X \lrcorner\, \omega$ is closed. On the one hand, if X is locally Hamiltonian, then in a neighborhood of each point there is a function f such that $X = X_f$, so $X \lrcorner\, \omega = X_f \lrcorner\, \omega = df$, which is certainly closed. Conversely, if X is symplectic, then by the Poincaré lemma each point $p \in M$ has a neighborhood U on which the closed 1-form $X \lrcorner \omega$ is exact. This means there is a smooth function f defined on U such that $X \lrcorner \omega = df$, which means $X = X_f$ on U.

Now suppose $H^1_{dR}(M) = 0$. Then every closed form is exact, so for any locally Hamiltonian (hence symplectic) vector field X there is a smooth function f such that $X \lrcorner \omega = df$. This means that $X = X_f$, so X is globally Hamiltonian. Conversely, suppose every locally Hamiltonian vector field is globally Hamiltonian. Let η be a closed 1-form, and let X be the vector field $X = \widetilde{\omega}^{-1}\eta$. Then (13.15) shows that $\mathcal{L}_X \omega = 0$, so X is symplectic and therefore locally Hamiltonian. By hypothesis, there is a global smooth function f such that $X = X_f$, and then unwinding the definitions, we find that $\eta = df$. □

Using Hamiltonian vector fields, we define an operation on functions similar to the Lie bracket of vector fields. Given $f, g \in C^\infty(M)$, we define their *Poisson bracket* $\{f, g\} \in C^\infty(M)$ by

$$\{f, g\} = X_f\, g = \omega(X_g, X_f).$$

Two functions are said to *Poisson commute* if their Poisson bracket is zero.

Example 13.24. Using the result of Example 13.21, we can easily compute the Poisson bracket of two functions f, g on $\mathbb{R}^{2n}$:

$$\{f, g\} = \sum_{i=1}^{n} \frac{\partial f}{\partial y^i}\frac{\partial g}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y^i}. \tag{13.16}$$

Proposition 13.25 (Properties of the Poisson Bracket). *Let (M, ω) be a symplectic manifold, and $f, g \in C^\infty(M)$.*

(*a*) *$\{f, g\} = -\{g, f\}$.*

(*b*) *g is constant along the flow of X_f if and only if $\{f, g\} = 0$.*

(*c*) *$X_{\{f,g\}} = [X_f, X_g]$.*

Proof. Part (a) is evident from the characterization $\{f, g\} = \omega(X_g, X_f)$, and (b) from $\{f, g\} = X_f g$. The proof of (c) is a computation using Proposition 13.11(e):

$$\begin{aligned}
d\{f, g\} &= d(X_f g) \\
&= d(\mathcal{L}_{X_f} g) \\
&= \mathcal{L}_{X_f} dg \\
&= \mathcal{L}_{X_f}(X_g \lrcorner \omega) \\
&= (\mathcal{L}_{X_f} X_g) \lrcorner \omega + X_g \lrcorner \mathcal{L}_{X_f} \omega \\
&= [X_f, X_g] \lrcorner \omega,
\end{aligned}$$

which is equivalent to (c). □

Our next theorem, called the Darboux theorem, is central in the theory of symplectic structures. It is a nonlinear analogue of the canonical form for a symplectic tensor given in Proposition 9.17.

Theorem 13.26 (Darboux). *Let (M,ω) be a $2n$-dimensional symplectic manifold. Near every point $p \in M$, there are coordinates $(x^1, y^1, \dots, x^n, y^n)$ in which ω is given by*

$$\omega = \sum_{i=1}^{n} dx^i \wedge dy^i. \tag{13.17}$$

Any coordinates satisfying the conclusion of the Darboux theorem are called *Darboux coordinates*, *symplectic coordinates*, or *canonical coordinates*. Before we begin the proof, let us prove the following lemma, which shows that Darboux coordinates are characterized by the Poisson brackets of the coordinate functions.

Lemma 13.27. *Let (M,ω) be a symplectic manifold. Coordinates (x^i, y^i) on an open set $U \subset M$ are Darboux coordinates if and only if their Poisson brackets satisfy*

$$\{x^i, x^j\} = \{y^i, y^j\} = 0; \quad \{x^i, y^j\} = -\delta^{ij}. \tag{13.18}$$

Proof. One direction is easy: If ω is given by (13.17), then formula (13.16) shows that the Poisson brackets of the coordinate functions satisfy (13.18). Conversely, If (x^i, y^i) are coordinates whose Poisson brackets satisfy (13.18), then the Hamiltonian fields of the coordinates satisfy

$$\begin{aligned}
dx^j(X_{x^i}) &= X_{x^i}(x^j) = \{x^i, x^j\} = 0,\\
dy^j(X_{x^i}) &= X_{x^i}(y^j) = \{x^i, y^j\} = -\delta^{ij},\\
dx^j(X_{y^i}) &= X_{y^i}(x^j) = \{y^i, x^j\} = \delta^{ij},\\
dy^j(X_{y^i}) &= X_{y^i}(y^j) = \{y^i, y^j\} = 0.
\end{aligned}$$

This implies that

$$X_{x^i} = -\frac{\partial}{\partial y^i}, \qquad X_{y^i} = \frac{\partial}{\partial x^i}.$$

If we write

$$\omega = A_{ij} dx^i \wedge dx^j + B_{ij} dx^i \wedge dy^j + C_{ij} dy^i \wedge dy^j,$$

with $A_{ij} = -A_{ji}$ and $C_{ij} = -C_{ji}$, then the coefficients are determined by

$$\begin{aligned}
A_{ij} &= \tfrac{1}{2}\omega\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \tfrac{1}{2}\omega(X_{y^i}, X_{y^j}) = \tfrac{1}{2}\{y^j, y^i\} = 0.\\
B_{ij} &= \omega\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = -\omega(X_{y^i}, X_{x^j}) = -\{x^j, y^i\} = \delta^{ij}.\\
C_{ij} &= \tfrac{1}{2}\omega\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right) = \tfrac{1}{2}\omega(X_{x^i}, X_{x^j}) = \tfrac{1}{2}\{x^j, x^i\} = 0.
\end{aligned}$$

This shows that ω has the form (13.17). □

Proof of the Darboux Theorem. We will show by induction on k that for each $k = 0, \dots, n$ there are functions $(x^1, y^1, \dots, x^k, y^k)$ satisfying (13.18) near p such that $\{dx^1, dy^1, \dots, dx^k, dy^k\}$ are independent at p. When $k = n$, this proves the theorem.

For $n = 0$, there is nothing to prove. So suppose we have such $(x^1, y^1, \dots, x^k, y^k)$. Observe that the Hamiltonian vector field of any constant function is zero. Therefore, because all of the Poisson brackets of the coordinates are constants, the Hamiltonian vector fields of the coordinates satisfy

$$\begin{aligned}
[X_{x^i}, X_{x^j}] &= X_{\{x^i, x^j\}} = 0, \\
[X_{x^i}, X_{y^j}] &= X_{\{x^i, y^j\}} = 0, \\
[X_{y^i}, X_{y^j}] &= X_{\{y^i, y^j\}} = 0.
\end{aligned}$$

Because $\widetilde{\omega}\colon T_pM \to T_p^*M$ is an isomorphism and the differentials $\{dx^i, dy^i\}$ are independent, these Hamiltonian vector fields are all independent in a neighborhood of p. Thus by the normal form theorem for commuting vector fields, there are coordinates $(u^1, \dots, u^{2n})$ near p such that

$$\frac{\partial}{\partial u^i} = X_{x^i}, \quad \frac{\partial}{\partial u^{i+k}} = X_{y^i}, \quad i = 1, \dots, k. \tag{13.19}$$

Let $x^{k+1} = u^{2k+1}$ in this coordinate system. Then x^{k+1} Poisson commutes with x^i and y^i for $i = 1, \dots, k$, because

$$\begin{aligned}
\{x^i, x^{k+1}\} &= X_{x^i}(x^{k+1}) = \frac{\partial}{\partial u^i} u^{2k+1} = 0, \\
\{y^i, x^{k+1}\} &= X_{y^i}(x^{k+1}) = \frac{\partial}{\partial u^{i+k}} u^{2k+1} = 0.
\end{aligned} \tag{13.20}$$

The restriction of ω to $\operatorname{span}\{X_{x^i}, X_{y^i}, i = 1, \dots, k\}$ is nondegenerate, as can easily be verified by computing the action of ω on these vectors using (13.18). Therefore, $X_{x^{k+1}}$ is independent of $\{X_{x^i}, X_{y^i}, i = 1, \dots, k\}$ because (13.20) implies $\omega(X_{x^{k+1}}, X_{x^i}) = \omega(X_{x^{k+1}}, X_{y^i}) = 0$. Just as before, the $2k+1$ Hamiltonian vector fields of the functions $(x^1, y^1, \dots, x^k, y^k, x^{k+1})$ all commute, so we can find new coordinates (v^i) such that

$$\begin{aligned}
\frac{\partial}{\partial v^i} &= X_{x^i}, \quad i = 1, \dots, k+1, \\
\frac{\partial}{\partial v^{i+k+1}} &= X_{y^i}, \quad i = 1, \dots, k.
\end{aligned}$$

Finally, let $y^{k+1} = -v^{k+1}$ in these coordinates. The independence condition is satisfied as before, and we compute

$$
\begin{aligned}
\{x^{k+1}, y^{k+1}\} &= H_{x^{k+1}} y^{k+1} = \frac{\partial}{\partial v^{k+1}}(-v^{k+1}) = -1. \\
\{x^j, y^{k+1}\} &= -H_{x^j} y^{k+1} = \frac{\partial}{\partial v^j} v^{k+1} = 0, \quad j = 1, \dots, k, \\
\{y^j, y^{k+1}\} &= -H_{y^j} y^{k+1} = \frac{\partial}{\partial v^{j+k+1}} v^{k+1} = 0, \quad j = 1, \dots, k.
\end{aligned}
$$

This completes the inductive step. □

This theorem was first proved in 1882 by Darboux. A much more elegant proof was discovered in the 1960s by Jürgen Moser [Mos65] and Alan Weinstein [Wei69]. It requires a bit more machinery than we have developed, but you can look it up, for example, in [Wei77] or [AM78].

The theory of symplectic manifolds is central to the study of classical mechanics. Many classical dynamical systems moving under the influence of Newton's laws of motion can be modeled naturally as the flow of a Hamiltonian vector field on a symplectic manifold. If one can find one or more functions that Poisson commute with the Hamiltonian, then they must be constant along the flow, so by restricting attention to a common level set of these functions one can often reduce the problem to one with fewer degrees of freedom. For much more on these ideas, see [AM78].

Problems

13-1. Let $V, W, X \in \mathfrak{T}(M)$ and $f, g \in C^\infty(M)$. Show that

(a) $[fV, gW] = fg[V,W] + f(Vg)W - g(Wf)V$.

(b) $\mathcal{L}_V(fW) = (Vf)W + f\mathcal{L}_V W$.

(c) $\mathcal{L}_{[V,W]}X = \mathcal{L}_V\mathcal{L}_W X - \mathcal{L}_W\mathcal{L}_V X$.

13-2. Let V and W be the vector fields of Exercise 13.2(b). Compute the flows θ, ψ of V and W, and verify that they do not commute by finding explicit times s and t such that $\theta_s \circ \psi_t \neq \psi_t \circ \theta_s$.

13-3. Give an example of vector fields V, $\widetilde{V}$, and W on $\mathbb{R}^2$ such that $V = \widetilde{V} = \partial/\partial x$ along the x-axis but $\mathcal{L}_V W \neq \mathcal{L}_{\widetilde{V}} W$ at the origin. [This shows that it is really necessary to know the vector field V to compute $(\mathcal{L}_V W)_p$; it is not sufficient just to know the vector V_p, or even to know the values of V along an integral curve of V.]

13-4. Determine all Killing fields on $(\mathbb{R}^n, \overline{g})$.

13-5. Let M be a smooth manifold and ω a 1-form on M. Show that for any smooth vector fields X, Y,

$$d\omega(X,Y) = X(\omega(Y)) - Y(\omega(X)) - \omega[X,Y].$$

13-6. Generalize the result of Problem 13-5 to a k-form ω by showing that

$$\begin{aligned} d\omega(X_1, \dots, X_{k+1}) &= \sum_{1 \le i \le k+1} (-1)^{i-1} X_i(\omega(X_1, \dots, \widehat{X}_i, \dots, X_{k+1})) \\ &+ \sum_{1 \le i < j \le k+1} (-1)^{i+j} \omega([X_i, X_j], X_1, \dots, \widehat{X}_i, \dots, \widehat{X}_j, \dots, X_{k+1}), \end{aligned}$$

where the hats indicate omitted arguments. [This formula can be used to give a coordinate-free definition of the exterior derivative. However, this definition has the serious flaw that, in order to compute the action of $d\omega$ on vectors $(X_1, \dots, X_k)$ at a point $p \in M$, one must first extend them to vector *fields* in a neighborhood of p. It is not evident from this formula that the resulting value is independent of the extensions chosen.]

13-7. For each k-tuple of vector fields on $\mathbb{R}^3$ shown below, either find coordinates (u^1, u^2, u^3) in a neighborhood of $(1,0,0)$ such that $V_i = \partial/\partial u^i$ for $i = 1, \dots, k$, or explain whey there are none.

(a) $k = 1$; $V_1 = x\dfrac{\partial}{\partial y} - y\dfrac{\partial}{\partial x}$.

(b) $k = 2;\ V_1 = \dfrac{\partial}{\partial x},\ V_2 = \dfrac{\partial}{\partial x} + \dfrac{\partial}{\partial y}.$

(c) $k = 2;\ V_1 = x\dfrac{\partial}{\partial y} - y\dfrac{\partial}{\partial x},\ V_2 = x\dfrac{\partial}{\partial x} + y\dfrac{\partial}{\partial y}.$

(d) $k = 3;\ V_1 = x\dfrac{\partial}{\partial y} - y\dfrac{\partial}{\partial x},\ V_2 = y\dfrac{\partial}{\partial z} - z\dfrac{\partial}{\partial y},\ V_3 = z\dfrac{\partial}{\partial x} - x\dfrac{\partial}{\partial z}.$

13-8. Let (M, ω) be a symplectic manifold. Show that the Poisson bracket satisfies the following identity for all $f, g, h \in C^\infty(M)$:

$$\{\{f, g\}, h\} + \{\{g, h\}, f\} + \{\{h, f\}, g\} = 0.$$

14
Integral Manifolds and Foliations

Suppose V is a nonvanishing vector field on a manifold M. The results of Chapter 12 tell us that each integral curve of V is an immersion, and that locally the images of the integral curves fit together nicely like parallel lines in Euclidean space. In particular, the fundamental theorem on flows tells us that these curves are determined by the knowledge of their tangent vectors.

In this chapter we explore an important generalization of this idea to higher-dimensional submanifolds. The general setup is this: Suppose we are given a k-dimensional subspace of T_pM at each point $p \in M$, varying smoothly from point to point. (Such a collection of subspaces is called a "tangent distribution.") Is there a k-dimensional submanifold (called an "integral manifold" of the tangent distribution) whose tangent space at each point is the given subspace? The answer in this case is more complicated than in the case of vector fields: There is a necessary condition, called involutivity, that must be satisfied by the tangent distribution. The main theorem of this chapter, the Frobenius theorem, tells us that this condition is also sufficient.

We will prove the Frobenius theorem in two forms: a local form, which says that a neighborhood of every point is filled up with integral manifolds, fitting together nicely like parallel subspaces of $\mathbb{R}^n$, and a global form, which says that the entire manifold is the disjoint union of immersed integral manifolds.

Tangent Distributions

Let M be a smooth manifold. Given a smooth vector bundle $\pi\colon E \to M$, a (*smooth*) *subbundle* of E is a subset $D \subset E$ with the following properties:

(i) D is an embedded submanifold of E.

(ii) For each $p \in M$, the fiber $D_p = D \cap \pi^{-1}(p)$ is a linear subspace of $E_p = \pi^{-1}(p)$.

(iii) With the vector space structure on each D_p inherited from E_p and the projection $\pi|_D\colon D \to M$, D is a smooth vector bundle over M.

Note that the condition that D be a vector bundle implies that the projection $\pi|_D\colon D \to M$ must be surjective, and that all the fibers D_p must have the same dimension.

Exercise 14.1. If $D \subset E$ is a smooth subbundle, show that the inclusion map $\iota\colon D \to E$ is a bundle map.

A subbundle of TM is called a *tangent distribution* on M, or just a *distribution* if there is no opportunity for confusion with the use of the term "distribution" for generalized functions in analysis. Other common names for distributions are *tangent subbundles* or *plane fields*. The dimension of each fiber of D is called *dimension* of the distribution.

The following lemma gives a convenient condition for checking that a collection of subspaces $\{D_p \subset T_pM : p \in M\}$ is a distribution.

Lemma 14.1. *Let M be a smooth manifold, and suppose for each $p \in M$ we are given a k-dimensional linear subspace $D_p \subset T_pM$. Then $D = \coprod_{p\in M} D_p \subset TM$ is a distribution if and only if the following condition is satisfied:*

$$\text{Each point } p \in M \text{ has a neighborhood } U \text{ on which there are smooth vector fields } Y_1, \dots, Y_k\colon U \to TM \text{ such that } Y_1|_p, \dots, Y_k|_p \text{ form a basis for } D_p \text{ at each } p \in U. \tag{14.1}$$

Proof. If D is a distribution, then by definition any $p \in M$ has a neighborhood U over which there exists a local trivialization of D, and by Problem 3-5 there exists a smooth local frame for D over any such set U. Such a local frame is by definition a collection of smooth sections $\sigma_1, \dots, \sigma_k\colon U \to D$ whose images form a basis for D_p at each point $p \in U$. The smooth vector fields we seek are given by $Y_j = \iota \circ \sigma_j$, where $\iota\colon D \hookrightarrow TM$ is inclusion.

Conversely, suppose that D satisfies (14.1). Condition (ii) in the definition of a subbundle is true by hypothesis, so we need to show that D satisfies conditions (i) and (iii).

To prove that D is an embedded submanifold, it suffices to show that each point $p \in M$ has a neighborhood U such that $D \cap \pi^{-1}(U)$ is an embedded submanifold of $\pi^{-1}(U) \subset TM$. Given $p \in M$, let $Y_1, \dots, Y_k$ be vector fields defined on a neighborhood of p and satisfying (14.1). The independent vectors $Y_1|_p, \dots, Y_k|_p$ can be extended to a basis $Y_1|_p, \dots, Y_n|_p$ for T_pM, and then $Y_{k+1}|_p, \dots, Y_n|_p$ can be extended to vector fields in a neighborhood of p. By continuity, they will still be independent in some neighborhood U of p, so they form a local frame for TM over U. By Problem 3-5 again, this yields a local trivialization $\pi^{-1}(U) \to U \times \mathbb{R}^n$ defined by

$$y^i Y_i|_q \mapsto (q, (y^1, \dots, y^n)).$$

In terms of this trivialization, $D \cap \pi^{-1}(U)$ corresponds to $U \times \mathbb{R}^k = \{(q, (y^1, \dots, y^k, 0, \dots, 0))\} \subset U \times \mathbb{R}^n$, which is obviously a regular submanifold. Moreover, the map $\Phi|_{D \cap \pi^{-1}(U)} \colon D \cap \pi^{-1}(U) \to U \times \mathbb{R}^k$ is obviously a local trivialization of D, showing that D is itself a vector bundle. $\square$

In the situation of the preceding lemma, we say D is the distribution (locally) spanned by the vector fields $Y_1, \dots, Y_k$.

Integral Manifolds and Involutivity

Suppose $D \subset TM$ is a distribution. An immersed submanifold $N \subset M$ is called an *integral manifold* of D if $T_pN = D_p$ at each point $p \in N$. The main question we want to address in this chapter is the existence of integral manifolds.

Before we proceed with the general theory, let us describe some examples of distributions and integral manifolds that you should keep in mind.

Example 14.2 (Tangent distributions).

(a) If V is any nowhere-vanishing vector field on a manifold M, then V spans a 1-dimensional distribution on M (i.e., $D_p = \operatorname{span}(V_p)$ for each $p \in M$). The smoothness criterion (14.1) is obviously satisfied. The image of any integral curve of V is an integral manifold of D.

(b) In $\mathbb{R}^n$, the vector fields $\partial/\partial x^1, \dots, \partial/\partial x^k$ span a k-dimensional distribution. The k-dimensional affine subspaces parallel to $\mathbb{R}^k$ are integral manifolds.

(c) Define a distribution on $\mathbb{R}^3 \smallsetminus \{0\}$ by letting D_p be the tangent space to the sphere through p and centered at 0. Away from the north and south poles, D is locally spanned by the coordinate vector fields $\partial/\partial\theta$ and $\partial/\partial\varphi$ in spherical coordinates; near the poles, we can use spherical coordinates composed with a suitable rotation. Thus D is a 2-dimensional distribution on $\mathbb{R}^3 \smallsetminus \{0\}$. Through each point $p \in \mathbb{R}^3 \smallsetminus \{0\}$, the sphere of radius $|p|$ around 0 is an integral manifold.

(d) Let X and Y be the following vector fields on $\mathbb{R}^3$:

$$X = \frac{\partial}{\partial x} + y\frac{\partial}{\partial z} \qquad Y = \frac{\partial}{\partial y},$$

and let D be the distribution spanned by X and Y. It turns out that D has no integral manifolds. To get an idea why, suppose N is an integral manifold through the origin. Because X and Y are tangent to N, any integral curve of X or Y that starts in N will have to stay in N, at least for a short time. Thus N contains an open subset of the x-axis (which is an integral curve of X). It also contains, for each sufficiently small x, an open subset of the line parallel to the y-axis and passing through $(x, 0, 0)$ (which is an integral curve of Y). Therefore N contains an open subset of the (x, y)-plane. However, the tangent planes to points of the (x, y)-plane at points off of the x-axis are not contained in D. Therefore, no such integral manifold exists.

The last example shows that, in general, integral manifolds may fail to exist. The reason for this failure is expressed in the following proposition. We say that a tangent distribution D is *involutive* if given any pair of local sections of D (i.e., vector fields X, Y defined on an open subset of M such that $X_p, Y_p \in D_p$ for each p), their Lie bracket is also a section of D. We say D is *integrable* if through each point of M there exists an integral manifold of D.

Proposition 14.3. *Every integrable distribution is involutive.*

Proof. Suppose X and Y are local sections of D defined on some open subset $U \subset M$. Let p be any point in U, and let N be an integral manifold of D passing through p. The fact that X and Y are sections of D means that X and Y are tangent to N. By Proposition 13.4, $[X, Y]$ is also tangent to N, and therefore $[X, Y]_p \in D_p$. □

The next lemma shows that the involutivity condition does not have to be checked for every pair of vector fields, just those of a local frame near each point.

Lemma 14.4. *Let $D \subset TM$ be a distribution. If in a neighborhood of every point of M there exists a local frame $(V_1, \dots, V_k)$ for D such that $[V_i, V_j]$ is a section of D for each $i, j = 1, \dots, k$, then D is involutive.*

Proof. Suppose the hypothesis holds, and suppose X and Y are sections of D over some open subset $U \subset M$. Given $p \in M$, choose a local frame $(V_1, \dots, V_k)$ satisfying the hypothesis in a neighborhood of p, and write $X = X^i V_i$ and $Y = Y^i V_i$. Then (using the result of Problem 13-1),

$$\begin{aligned}[X, Y] &= \left[X^i V_i, Y^j V_j\right] \\ &= X^i Y^j [V_i, V_j] + X^i (V_i Y^j) V_j - Y^j (V_j X^i) V_i.\end{aligned}$$

It follows from the hypothesis that this last expression is a section of D. $\square$

The Frobenius Theorem

In Example 14.2, all of the tangent distributions we defined except the last one had the property that there was an integral manifold through each point. Moreover, these submanifolds all "fit together" nicely like parallel affine subspaces of $\mathbb{R}^n$. Given a k-dimensional distribution $D \subset TM$, let us say that a coordinate chart (U, φ) on M is *flat* for D if at points of U, D is spanned by the first k coordinate vector fields $\partial/\partial x^1, \dots, \partial/\partial x^k$. It is obvious that each slice of the form $x^{k+1} = c^{k+1}, \dots, x^n = c^n$ for constants $c^{k+1}, \dots, c^n$ is an integral manifold of D. This is the nicest possible local situation for integral manifolds We say that a distribution $D \subset TM$ is *completely integrable* if there exists a flat chart for D in a neighborhood of every point of M. Obviously every completely integrable distribution is integrable and therefore involutive.

The next theorem is the main result of this chapter, and indeed one of the central theorems in smooth manifold theory. (The others are the inverse function theorem, the fundamental theorem on flows, and Stokes's theorem.)

Theorem 14.5 (Frobenius). *Every involutive distribution is completely integrable.*

Proof. The canonical form theorem for commuting vector fields (Theorem 13.10) implies that any distribution locally spanned by *commuting* vector fields is completely integrable. Thus it suffices to show that any involutive distribution is locally spanned by commuting vector fields.

Let D be a k-dimensional involutive distribution on an n-dimensional manifold M. Given $p \in M$, choose a neighborhood U of p on which there exist coordinates $(x^1, \dots, x^n)$ centered at p and a local frame $Y_1, \dots, Y_k$ for D. By a linear change of coordinates, we may assume that $\partial/\partial x^i|_p = Y_i|_p$.

Let $\Pi \colon U \to \mathbb{R}^k$ be the smooth map whose coordinate representation is the projection onto the first k coordinates: $\Pi(x^1, \dots, x^n) = (x^1, \dots, x^k)$. This induces a smooth map $\Pi_* \colon TU \to T\mathbb{R}^k$, which can be written

$$\Pi_* \left(\sum_{i=1}^{n} v^i \left. \frac{\partial}{\partial x^i} \right|_q \right) = \sum_{i=1}^{k} v^i \left. \frac{\partial}{\partial x^i} \right|_{\Pi(q)} .$$

(Notice that the summation is only over $i = 1, \dots, k$ on the right-hand side.) Clearly, the restriction of Π_* to $D_p \subset T_pM$ is an isomorphism, and thus by continuity the same is true of $\Pi_* \colon D_q \to T_{\Pi(q)}\mathbb{R}^k$ for q in a neighborhood of p. The matrix entries of $\Pi_*|_D$ with respect to the frames $\{Y_i|_q\}$ and $\{\partial/\partial x^j|_{\Pi(q)}\}$ are smooth functions of q, and thus so are the matrix entries of $\left(\Pi_*|_{D_q}\right)^{-1}: T_{\Pi(q)}\mathbb{R}^k \to D_q$. Define a new local frame $X_1, \dots, X_k$ for D near p by

$$X_i|_q = \left(\Pi_*|_{D_q}\right)^{-1} \left.\frac{\partial}{\partial x^i}\right|_{\Pi(q)}. \tag{14.2}$$

The theorem will be proved if we can show that $[X_i, X_j] = 0$ for all i, j.

First observe that X_i and $\partial/\partial x^i$ are Π-related, because (14.2) implies that

$$\left.\frac{\partial}{\partial x^i}\right|_{\Pi(q)} = \left(\Pi_*|_{D_q}\right) X_i|_q = \Pi_* X_i|_q.$$

Therefore, by the naturality of Lie brackets,

$$\Pi_*[X_i, X_j]_q = \left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right]_{\Pi(q)} = 0.$$

Since $[X_i, X_j]$ takes its values in D by involutivity, and Π_* is injective on D, this implies that $[X_i, X_j]_q = 0$ for each q, thus completing the proof. □

One immediate consequences of the theorem is the following lemma, which we will use repeatedly.

Proposition 14.6 (Local Structure of Integral Manifolds). *Let D be an involutive k-dimensional distribution on a smooth manifold M, and let (U, φ) be a flat chart for D. If N is any integral manifold of D, then $N \cap U$ is a countable disjoint union of open subsets of k-dimensional slices of U, each of which is open in N and embedded in M.*

Proof. Because the inclusion map $\iota\colon N \hookrightarrow M$ is continuous, $N \cap U = \iota^{-1}(U)$ is open in N, and thus consists of a countable disjoint union of connected components, each of which is open in N.

Let V be any component of $N \cap U$. We will show first that V is contained in a single slice. Choosing some $p \in V$, it suffices to show that $x^i(q) = x^i(p)$ for $i = k+1, \dots, n$ for any $q \in V$. Since V is connected, there exists a path γ in V from p to q, which we may take to be smooth by Problem 6-1. Because γ lies in V and V is an integral manifold of D, we have $\gamma'(t) \in T_{\gamma(t)}V = D_{\gamma(t)}$. Because D is spanned by $\partial/\partial x^1, \dots, \partial/\partial x^k$ in U, this implies that the last $n-k$ components of $\gamma'(t)$ are zero: $(\gamma^i)'(t) \equiv 0$ for $i \geq k+1$. Since the domain of γ is connected, this means γ^i is constant for these values of i, so p and q lie in a single slice S.

Because S is embedded in M, the inclusion map $V \hookrightarrow M$ is also smooth as a map into S by Corollary 5.38. The inclusion $V \hookrightarrow S$ is thus an injective immersion between manifolds of the same dimension, and therefore a local diffeomorphism, an open map, and a homeomorphism onto an open subset of S. □

This local characterization of integral manifolds implies the following strengthening of our theorem about restricting the range of a smooth map. (This result will be used in Chapter 15.)

Proposition 14.7. *Suppose $H \subset N$ is an integral manifold of an involutive distribution D on N. If $F\colon M \to N$ is a smooth map such that $F(M) \subset H$, then F is smooth as a map from M to H.*

Proof. Let $p \in M$ be arbitrary, and set $q = F(p) \in H$. Let $(y^1, \dots, y^n)$ be flat coordinates for D on a neighborhood U of q. Choose coordinates (x^i) for M on a connected neighborhood B of p. Writing the coordinate representation of F as

$$(y^1, \dots, y^n) = (F^1(x), \dots, F^n(x)),$$

the fact that $F(B) \subset H$ means that the coordinate functions $F^{k+1}, \dots, F^n$ take on only countably many values. Because B is connected, the intermediate value theorem implies that these coordinate functions are constant, and thus $F(B)$ lies in a single slice. On this slice, $(y^1, \dots, y^k)$ are coordinates for H, so $F\colon N \to H$ has the local coordinate representation

$$(y^1, \dots, y^k) = (F^1(x), \dots, F^k(x)),$$

which is smooth. □

Applications

In the next chapter, we will see some geometric applications of the Frobenius theorem. In this chapter, we concentrate on some applications to the study of partial differential equations (PDEs).

Our first application is not so much an application as a simple rephrasing of the theorem. Because explicitly finding integral manifolds boils down to solving a system of PDEs, we can interpret the Frobenius theorem as an existence and uniqueness result for such equations.

Suppose we are given $n-1$ smooth vector fields $Y_1, \dots, Y_{n-1}$ on an open subset of $\mathbb{R}^n$, and suppose we seek a smooth function f that satisfies the system of equations $Y_1 f = \cdots Y_{n-1} f = 0$. Away from the critical points of f, the level sets of f will be integral manifolds of the distribution spanned by the Y_i's. Thus the involutivity of this distribution is a necessary and sufficient condition for the existence of a solution f on any open set.

Example 14.8. Consider the system

$$-2z^2\frac{\partial f}{\partial x} + 2x\frac{\partial f}{\partial z} = 0, \qquad -3z^2\frac{\partial f}{\partial y} + 2y\frac{\partial f}{\partial z} = 0 \tag{14.3}$$

of PDEs for a real-valued function f of three variables. It is an example of a *linear*, *first-order* system of PDEs for f (linear because the left-hand sides of (14.3) depend linearly on f, and first-order because it involves only first derivatives of f). It is also *overdetermined*, because there are more equations than unknown functions. In general, overdetermined systems have solutions only if they satisfy certain compatibility conditions; in this case, involutivity is the key.

To see whether (14.3) has nonconstant solutions, we let X and Y denote the vector fields

$$X = -2z^2\frac{\partial}{\partial x} + 2x\frac{\partial}{\partial z}, \qquad Y = -3z^2\frac{\partial}{\partial y} + 2y\frac{\partial}{\partial z}.$$

These span a 2-dimensional distribution D on the subset U of $\mathbb{R}^3$ where $z \neq 0$, as you can easily check. By direct computation, if $z \neq 0$,

$$[X, Y] = -12xz\frac{\partial}{\partial y} + 8yz\frac{\partial}{\partial x} = \frac{4x}{z}Y - \frac{4y}{z}X,$$

which shows that D is involutive on U. Therefore, on a neighborhood of any point (x_0, y_0, z_0) with $z_0 \neq 0$, there are flat coordinates (u, v, w) for D; because X and Y are tangent to the level sets of w it follows that $f = w$ is a solution to (14.3) defined in a neighborhood of (x_0, y_0, z_0). In this flat chart, (14.3) is equivalent to $\partial f/\partial u = \partial f/\partial v = 0$, because $\partial/\partial u$ and $\partial/\partial v$ also span D. Therefore, (assuming that the w-slices are connected, which will be the case if U is small enough), the solutions in U are precisely those smooth functions that depend on w alone in these coordinates, i.e., those of the form $f(x, y, z) = g(w(x, y, z))$ for some smooth function g of one variable.

Exercise 14.2. What can you say about solutions in a neighborhood of a point $(x_0, y_0, 0)$?

The next application is more substantial. We introduce it first in the case of a single function $f(x, y)$ of two independent variables, in which case the notation is considerably simpler.

Suppose we seek a solution f to the system

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \alpha(x, y, f(x, y)), \\ \frac{\partial f}{\partial y}(x, y) &= \beta(x, y, f(x, y)), \end{aligned} \tag{14.4}$$

where α and β are smooth functions defined on some open subset $U \subset \mathbb{R}^3$. This is an overdetermined system of (possibly nonlinear) first-order partial differential equations.

To determine the necessary conditions that α and β must satisfy for solvability of (14.4), assume f is a solution on some open set in $\mathbb{R}^2$. Because $\partial^2 f/\partial x\partial y = \partial^2 f/\partial y\partial x$, (14.4) implies

$$\frac{\partial}{\partial y}\big(\alpha(x,y,f(x,y))\big) = \frac{\partial}{\partial x}\big(\beta(x,y,f(x,y))\big)$$

and therefore by the chain rule

$$\frac{\partial \alpha}{\partial y} + \beta\frac{\partial \alpha}{\partial z} = \frac{\partial \beta}{\partial x} + \alpha\frac{\partial \beta}{\partial z}. \tag{14.5}$$

This is true at any point $(x,y,z) \in U$ provided there is a smooth solution f with $f(x,y) = z$. In particular, (14.5) is a necessary condition for (14.4) to have a solution in a neighborhood of any point (x_0,y_0) with arbitrary initial value $f(x_0,y_0) = z_0$. Using the Frobenius theorem, we can show that this condition is sufficient.

Proposition 14.9. *Suppose α and β are smooth functions defined on some open set $U \subset \mathbb{R}^3$ and satisfying* (14.5) *there. For any $(x_0,y_0,z_0) \in U$, there is a neighborhood V of (x_0,y_0) in $\mathbb{R}^2$ and a unique smooth function $f\colon V \to \mathbb{R}$ satisfying* (14.4) *and $f(x_0,y_0) = z_0$.*

Proof. The idea of the proof is that the system (14.4) determines the partial derivatives of f in terms of its values, and therefore determines the tangent plane to the graph of f at each point in terms of the coordinates of the point on the graph. This collection of tangent planes defines a 2-dimensional distribution on U, and (14.5) is equivalent to the involutivity condition for this distribution.

If there were a solution f on an open set $V \subset \mathbb{R}^2$, the map $\Phi\colon V \to \mathbb{R}^3$ given by

$$\Phi(x,y) = (x,y,f(x,y))$$

would be a diffeomorphism onto the graph $\Gamma(f) \subset \mathbb{R}^2$. At any point $p = \Phi(x,y)$, the tangent space $T_p\Gamma(f)$ is spanned by the vector fields

$$\begin{aligned}
\Phi_* \left.\frac{\partial}{\partial x}\right|_{(x,y)} &= \left.\frac{\partial}{\partial x}\right|_p + \frac{\partial f}{\partial x}(x,y) \left.\frac{\partial}{\partial z}\right|_p, \\
\Phi_* \left.\frac{\partial}{\partial y}\right|_{(x,y)} &= \left.\frac{\partial}{\partial y}\right|_p + \frac{\partial f}{\partial y}(x,y) \left.\frac{\partial}{\partial z}\right|_p.
\end{aligned} \tag{14.6}$$

The system (14.4) is satisfied if and only if

$$\begin{aligned}
\Phi_* \left.\frac{\partial}{\partial x}\right|_{(x,y)} &= \left.\frac{\partial}{\partial x}\right|_p + \alpha(x,y,f(x,y)) \left.\frac{\partial}{\partial z}\right|_p, \\
\Phi_* \left.\frac{\partial}{\partial y}\right|_{(x,y)} &= \left.\frac{\partial}{\partial y}\right|_p + \beta(x,y,f(x,y)) \left.\frac{\partial}{\partial z}\right|_p.
\end{aligned} \tag{14.7}$$

Let X and Y be the vector fields

$$X = \frac{\partial}{\partial x} + \alpha(x,y,z)\frac{\partial}{\partial z},$$
$$Y = \frac{\partial}{\partial y} + \beta(x,y,z)\frac{\partial}{\partial z}$$

on U, and let D be the distribution on U spanned by X and Y. A little linear algebra will convince you that (14.7) holds if and only if $\Gamma(f)$ is an integral manifold of D. A straightforward computation using (14.5) shows that $[X,Y] \equiv 0$, so through each point $p = (x_0, y_0, z_0) \in U$ there is an integral manifold N of D. Let $F\colon W \to \mathbb{R}$ be a defining function for N on some neighborhood W of p; for example, we could take F to be the third coordinate function in a flat chart. The tangent space to N at each point $p \in N$ (namely D_p) is equal to the kernel of $F_*\colon T_p\mathbb{R}^3 \to T_p\mathbb{R}$. Since $\partial/\partial z|_p \notin D_p$ at any point p, this implies that $\partial F/\partial z \neq 0$ at p, so by the implicit function theorem N is the graph of a smooth function $z = f(x,y)$ in some neighborhood of p. You can verify easily that f is a solution to the problem. Uniqueness follows immediately from Proposition 14.6. □

There is a straightforward generalization of this result to higher dimensions. The general statement of the theorem may seem hopelessly complicated, but verifying the necessary conditions in specific examples usually just amounts to computing mixed partial derivatives and applying the chain rule.

Proposition 14.10. *Suppose α^i_j, $i = 1,\dots,k$, $j = 1,\dots,n$ are smooth functions defined on some open set $U \subset \mathbb{R}^{n+k}$ and satisfying*

$$\frac{\partial \alpha^i_j}{\partial x^k} + \alpha^l_k \frac{\partial \alpha^i_j}{\partial z^l} = \frac{\partial \alpha^i_k}{\partial x^j} + \alpha^l_j \frac{\partial \alpha^i_k}{\partial z^l},$$

where we use $(x,z) = (x^1,\dots,x^n,z^1,\dots,z^k)$ to denote a point in $\mathbb{R}^{n+k}$. For any $(x_0,z_0) \in U$, there is a neighborhood V of x_0 in $\mathbb{R}^n$ and unique smooth map $f\colon V \to \mathbb{R}^k$ satisfying

$$\frac{\partial f^i}{\partial x^j}(x^1,\dots,x^n) = \alpha^i_j(x^1,\dots,x^n,f^1(x),\dots,f^k(x))$$

and $f(x_0) = z_0$.

Exercise 14.3. Prove Proposition 14.10.

Foliations

When we put together all the maximal integral manifolds of a k-dimensional involutive distribution D, we obtain a decomposition of M into k-dimensional submanifolds that “fit together” locally like the slices in a flat chart.

We define a *foliation* of dimension k on an n-manifold M to be a collection $\mathcal{F}$ of disjoint, connected, immersed k-dimensional submanifolds of M (called the *leaves* of the foliation), whose union is M, and such that in a neighborhood of each point $p \in M$ there is a chart $(U, (x^i))$ such that each leaf of the foliation intersects U in either the empty set or a countable union of k-dimensional slices of the form $x^{k+1} = c^{k+1}, \ldots, x^n = c^n$. (Such a chart is called a *flat chart* for the foliation.)

Example 14.11 (Foliations).

(a) The collection of all k-dimensional affine subspaces of $\mathbb{R}^n$ parallel to $\mathbb{R}^k$ is a k-dimensional foliation of $\mathbb{R}^n$.

(b) The collection of all open rays of the form $\{\lambda x_0 : \lambda > 0\}$ is a 1-dimensional foliation of $\mathbb{R}^n \smallsetminus \{0\}$.

(c) The collection of all spheres centered at 0 is an $(n-1)$-dimensional foliation of $\mathbb{R}^n \smallsetminus \{0\}$.

(d) The collection of all circles of the form $\mathbb{S}^1 \times \{q\} \subset \mathbb{T}^2$ as q ranges over $\mathbb{S}^1$ yields a foliation of the torus $\mathbb{T}^2$. A different foliation of $\mathbb{T}^2$ is given by the collection of circles of the form $\{p\} \times \mathbb{S}^1$ as p ranges over $\mathbb{S}^1$.

(e) If α is a fixed irrational number, the images of all curves of the form

$$\gamma_\theta(t) = (\cos t, \sin t, \cos(\alpha t + \theta), \sin(\alpha t + \theta))$$

as θ ranges over $[0, 2\pi)$ form a 1-dimensional foliation of the torus in which each leaf is dense (cf. Example 5.3 and Problem 5-4).

(f) The collection of curves in $\mathbb{R}^2$ satisfying one of the following equations is a foliation of the plane:

$$\begin{aligned} y &= \sec(x + k\pi) + c, \qquad && k \in \mathbb{Z}, c \in \mathbb{R}; \\ x &= \frac{\pi}{2} + l, && l \in \mathbb{Z}. \end{aligned}$$

(g) If we rotate the curves of the previous example around the z-axis, we obtain a 2-dimensional foliation of $\mathbb{R}^3$ in which some of the leaves are diffeomorphic to disks and some are diffeomorphic to annuli.

The main fact about foliations is that they are in one-to-one correspondence with involutive distributions. One direction, expressed in the next lemma, is an easy consequence of the definition.

Lemma 14.12. *Let $\mathcal{F}$ be a foliation of a smooth manifold M. The collection of tangent spaces to the leaves of $\mathcal{F}$ forms an involutive distribution on M.*

Exercise 14.4. Prove Lemma 14.12.

The Frobenius theorem allows us to conclude the following converse, which is much more profound. By the way, it is worth noting that this result is one of the two main reasons why the notion of immersed submanifold has been defined.

Theorem 14.13 (Global Frobenius Theorem). *Let D be an involutive k-dimensional tangent distribution on a smooth manifold M. The collection of all maximal connected integral manifolds of D forms a foliation of M.*

The theorem will be a consequence of the following lemma.

Lemma 14.14. *Let $\{N_\alpha\}_{\alpha\in A}$ be any collection of connected integral manifolds of D with a point p in common. Then $N = \bigcup_\alpha N_\alpha$ has a unique smooth manifold structure making it into a connected integral manifold of D in which each N_α is an open submanifold.*

Proof. Define a topology on N by declaring a subset U of N to be open if and only if $U \cap N_\alpha$ is open in N_α for each α. It is easy to check that this is a topology. To prove that each N_α is open in N, we need to show that $N_\alpha \cap N_\beta$ is open in N_β for each β. Let q be an arbitrary point of $N_\alpha \cap N_\beta$, and choose a flat chart for D on a neighborhood U of q. Let V_α, V_β denote the components of $N_\alpha \cap U$ and $N_\beta \cap U$, respectively, containing q. By the preceding lemma, V_α and V_β are open subsets of single slices with the subspace topology, and since both contain q, they both must lie in the same slice S. Thus $V_\alpha \cap V_\beta$ is open in S and also in both N_α and N_β. Since each $q \in N_\alpha \cap N_\beta$ has a neighborhood in N_β contained in the intersection, it follows that $N_\alpha \cap N_\beta$ is open in N_β as claimed. Clearly this is the unique topology on N with the property that each N_α is a subspace of N.

With this topology, N is locally Euclidean of dimension k, because each point $q \in N$ has a Euclidean neighborhood V in some N_α, and V is an open subset of N because its intersection with each N_β is open in N_β by the argument in the preceding paragraph. Moreover, the inclusion map $N \hookrightarrow M$ is continuous: For any open subset $U \subset M$, $U \cap N$ is open in N because $U \cap N_\alpha$ is open in N_α for each α.

To see that N is Hausdorff, let $q, q' \in N$ be given. There are disjoint open sets $U, U' \subset M$ containing q and q', respectively, and then (because inclusion $N \hookrightarrow M$ is continuous) $N \cap U$ and $N \cap U'$ are disjoint open subsets of N containing q.

Next we show that N is second countable. We can cover M with countably many flat charts for D, say $\{W_i\}_{i\in\mathbb{N}}$. It suffices to show that $N \cap W_i$ is contained in a countable union of slices for each i, for then we can choose a countable basis for the portion of N in each such slice, and the union of all such bases forms a countable basis for the topology of N.

Suppose W_k is one of these flat charts and $S \subset W_k$ is a slice containing a point $q \in N$. There is some connected integral manifold N_α containing p and q. Because connected manifolds are path connected (***appendix?), there is a continuous path $\gamma\colon [0,1] \to N_\alpha$ connecting p and q. Since $\gamma[0,1]$ is compact, there exist finitely many numbers $0 = t_0 < t_1 < \cdots < t_m = 1$ such that $\gamma[t_{j-1}, t_j]$ is contained in one of the flat charts W_{i_j} for each j. Since $\gamma[t_{j-1}, t_j]$ is connected, it is contained in a single component of $W_{i_j} \cap N_\alpha$ and therefore in a single slice $S_{i_j} \subset W_{i_j}$.

Let us say that a slice S of some W_k is *accessible from* p if there is a finite sequence of indices $i_0, \ldots, i_m$ and for each i_j a slice $S_{i_j} \subset W_{i_j}$, with the properties that $p \in S_{i_0}$, $S_{i_m} = S$, and $S_{i_j} \cap S_{i_{j+1}} \neq \varnothing$ for each j. The discussion in the preceding paragraph showed that every slice that contains a point of N is accessible from p. To complete the proof of second countability, we just note that S_{i_j} is itself an integral manifold, and therefore it meets at most countably many slices of $W_{i_{j+1}}$ by Proposition 14.6; thus there are only countably many slices accessible from p. Therefore, N is a topological manifold of dimension k. It is connected because it is a union of connected subspaces with a point in common.

To construct a smooth structure on N, we define an atlas consisting of all charts of the form $(S \cap N, \psi)$, where S is a single slice of some flat chart, and $\psi\colon S \to \mathbb{R}^k$ is the map whose coordinate representation is projection onto the first k coordinates: $\psi(x^1, \ldots, x^n) = (x^1, \ldots, x^k)$. Because any slice is an embedded submanifold, its smooth structure is uniquely determined, and thus whenever two such slices S, S' overlap the transition map $\psi' \circ \psi$ is smooth.

With respect to this smooth structure, the inclusion map $N \hookrightarrow M$ is an immersion (because it is an embedding on each slice), and the tangent space to N at each point $q \in N$ is equal to D_q (because this is true for slices). The smooth structure is uniquely determined by the requirement that the inclusion $N \hookrightarrow M$ be an immersion, because this implies N is locally embedded and thus its smooth structure must match those of the slices on small enough open sets. □

Proof of the global frobenius theorem. For each $p \in M$, let L_p be the union of all connected integral manifolds of D containing p. By the preceding lemma, L_p is a connected integral manifold of D containing p, and it is clearly maximal. If any two such maximal integral manifolds L_p and $L_{p'}$ intersect, their union $L_p \cup L_{p'}$ is an integral manifold containing both p and p', so by maximality $L_p = L_{p'}$. Thus the various maximal integral manifolds are either disjoint or identical.

If (U, φ) is any flat chart for D, then $L_p \cap U$ is a countable union of open subsets of slices by Proposition 14.6. For any such slice S, if $L_p \cap S$ is neither empty nor all of S, then $L_p \cup S$ is a connected integral manifold properly containing L_p, which contradicts the maximality of L_p. Therefore, $L_p \cap U$

is precisely a countable union of slices, so the collection $\{L_p : p \in M\}$ is the desired foliation. $\square$

Exercise 14.5. If M and N are smooth manifolds, show that TM and TN define integrable distributions on $M \times N$, whose leaves are diffeomorphic to M and N, respectively.

Problems

14-1. If G is a connected Lie group acting smoothly, freely, and properly on a smooth manifold M, show that the orbits of G form a foliation of M.

14-2. Let D be the distribution on $\mathbb{R}^3$ spanned by

$$X = \frac{\partial}{\partial x} + yz\frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y}$$

(a) Find an integral submanifold of D passing through the origin.

(b) Is D involutive? Explain your answer in light of part (a).

14-3. Of the systems of partial differential equations below, determine which ones have solutions $z(x, y)$ (or, for part (c), $z(x, y)$ and $w(x, y)$) in a neighborhood of the origin for arbitrary positive values of $z(0, 0)$ (respectively, $z(0, 0)$ and $w(0, 0)$).

(a) $\dfrac{\partial z}{\partial x} = z \cos y; \qquad \dfrac{\partial z}{\partial y} = -z \log z \tan y.$

(b) $\dfrac{\partial z}{\partial x} = e^{xz}; \qquad \dfrac{\partial z}{\partial y} = xe^{yz}.$

(c) $\dfrac{\partial z}{\partial x} = z; \qquad \dfrac{\partial z}{\partial y} = w; \qquad \dfrac{\partial w}{\partial x} = w; \qquad \dfrac{\partial w}{\partial y} = z.$

14-4. This problem outlines an alternative characterization of involutivity in terms of differential forms. Suppose D is a k-dimensional tangent distribution on an n-manifold M.

(a) Show that near any point of M there are $n - k$ independent 1-forms $\omega^1, \dots, \omega^{n-k}$ such that $X \in D$ if and only if $\omega^i(X) = 0$ for $i = 1, \dots, k$.

(b) Show that D is involutive if and only if whenever $(\omega^1, \dots, \omega^{n-k})$ are forms as above, there are 1-forms $\{\alpha^i_j : i, j = 1, \dots, n - k\}$ such that

$$d\omega^i = \sum_{j=1}^{k} \alpha^i_j \wedge \omega^j.$$

[Hint: Use Problem 13-5.]

(c) Let $\mathcal{A}^*(M)$ denote the vector space $\mathcal{A}^0(M) \oplus \cdots \oplus \mathcal{A}^n(M)$. With the wedge product, $\mathcal{A}^*(M)$ is an associative ring. Show that the set

$$\mathcal{I}(D) = \{\omega \in \mathcal{A}^*(M) : \omega|_D = 0\}$$

is an ideal in $\mathcal{A}^*(M)$.

(d) An ideal $\mathcal{I}$ is said to be a *differential ideal* if $d(\mathcal{I}) \subset \mathcal{I}$, that is, if whenever $\omega \in \mathcal{I}$, $d\omega \in \mathcal{I}$ as well. Show that D is involutive if and only if $\mathcal{I}(D)$ is a differential ideal.

15

Lie Algebras and Lie Groups

The set of vector fields on a Lie group that are invariant under all left translations forms a finite-dimensional vector space, which carries a natural bilinear product making it into an algebraic structure known as a Lie algebra. Many of the properties of a Lie group are reflected in the algebraic structure of its Lie algebra.

In this chapter, we introduce the definition of an abstract Lie algebra, define the Lie algebra of a Lie group, and explore some of its important properties, including the relationships among Lie algebras, homomorphisms of Lie groups, and one-parameter subgroups of Lie groups. Later we introduce the exponential map, a smooth map from the Lie algebra into the group that shows in a very explicit way how the group structure near the identity is reflected in the algebraic structure of the Lie algebra. The culmination of the chapter is a complete description of the fundamental correspondence between Lie groups and Lie algebras: There is a one-to-one correspondence between finite-dimensional Lie algebras and simply-connected Lie groups, and all of the connected Lie groups with a given Lie algebra are quotients of the simply connected one by discrete normal subgroups.

Lie Algebras

A *Lie algebra* is a real vector space $\mathfrak{b}$ endowed with a bilinear map $\mathfrak{b}\times\mathfrak{b} \to \mathfrak{b}$, denoted by $(X, Y) \mapsto [X, Y]$ and called the *bracket* of X and Y, satisfying the following two properties for all $X, Y, Z \in \mathfrak{b}$:

(i) ANTISYMMETRY: $[X,Y] = -[Y,X]$.

(ii) JACOBI IDENTITY: $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$.

Notice that the Jacobi identity is a substitute for associativity, which does not hold in general for brackets in a Lie algebra.

Example 15.1 (Lie algebras).

(a) The vector space $\mathrm{M}(n,\mathbb{R})$ of $n \times n$ real matrices is an n^2-dimensional Lie algebra under the *commutator bracket*

$$[A,B] = AB - BA.$$

Antisymmetry is obvious from the definition, and the Jacobi identity follows from a straightforward calculation. When we are thinking of $\mathrm{M}(n,\mathbb{R})$ as a Lie algebra with this bracket, we will denote it by $\mathfrak{gl}(n,\mathbb{R})$.

(b) Similarly, $\mathfrak{gl}(n,\mathbb{C})$ is the $2n^2$-dimensional (real) Lie algebra obtained by endowing $\mathrm{M}(n,\mathbb{C})$ with the commutator bracket.

(c) The space $\mathfrak{T}(M)$ of all smooth vector fields on a smooth manifold M is an infinite-dimensional Lie algebra under the Lie bracket by Lemma 13.1.

(d) If (M,ω) is a symplectic manifold, the space $C^\infty(M)$ becomes a Lie algebra under the Poisson bracket. Problem 13-8 shows that this bracket satisfies the Jacobi identity.

(e) Any vector space V becomes a Lie algebra if we define all brackets to be zero. Such a Lie algebra is said to be *abelian*. (The name refers to the fact that the bracket in most Lie algebras, as in the two preceding examples, is defined as a commutator operation in terms of an underlying associative product; so "abelian" refers to the fact that all brackets are zero precisely when the underlying product is commutative.)

(f) If $\mathfrak{g}$ and $\mathfrak{h}$ are Lie algebras, then $\mathfrak{g} \times \mathfrak{h}$ is a Lie algebra with the bracket operation defined by

$$[(X,Y),(X',Y')] = ([X,X'],[Y,Y']).$$

With this bracket, $\mathfrak{g} \times \mathfrak{h}$ is called a *product Lie algebra*.

If $\mathfrak{b}$ is a Lie algebra, a linear subspace $\mathfrak{a} \subset \mathfrak{b}$ is called a *Lie subalgebra* of $\mathfrak{b}$ if it is closed under Lie brackets. In this case $\mathfrak{a}$ is itself a Lie algebra with the same bracket operation.

If $\mathfrak{a}$ and $\mathfrak{b}$ are Lie algebras, a linear map $A\colon \mathfrak{a} \to \mathfrak{b}$ is called a *Lie algebra homomorphism* if it preserves brackets: $A[X,Y] = [AX,AY]$. An invertible

Lie algebra homomorphism is called a *Lie algebra isomorphism*. If there exists a Lie algebra isomorphism from $\mathfrak{a}$ to $\mathfrak{b}$, we say they are *isomorphic* as Lie algebras.

Exercise 15.1. Verify that the kernel and image of a Lie algebra homomorphism are Lie subalgebras.

Exercise 15.2. If $\mathfrak{g}$ and $\mathfrak{h}$ are Lie algebras and $A\colon \mathfrak{g} \to \mathfrak{h}$ is a linear map, show that A is a Lie algebra homomorphism if and only if $A[E_i, E_j] = [AE_i, AE_j]$ for some basis $(E_1, \dots, E_n)$ of $\mathfrak{g}$.

Now suppose G is a Lie group. Recall that each element $g \in G$ defines a diffeomorphism $L_g\colon G \to G$ called *left translation*, given by $L_g(h) = gh$. A smooth vector field $X \in \mathcal{T}(G)$ is said to be *left-invariant* if it is invariant under all left translations: $L_{g*}X = X$ for every $g \in G$. (Because L_g is a diffeomorphism, $L_{g*}X$ is a well-defined vector field on G, and $L_{g*}X = X$ means $L_{g*}(X_{g'}) = X_{gg'}$ for every pair g, g' of elements of G.)

Lemma 15.2. *Let G be a Lie group, and let $\mathfrak{g}$ denote the set of all left-invariant vector fields on G. Then $\mathfrak{g}$ is a Lie subalgebra of $\mathcal{T}(M)$.*

Proof. Because $L_{g*}(aX + bY) = aL_{g*}X + bL_{g*}Y$, it is clear that $\mathfrak{g}$ is a linear subspace of $\mathcal{T}(M)$. By the naturality of Lie brackets, if $X, Y \in \mathfrak{g}$,

$$L_{g*}[X, Y] = [L_{g*}X, L_{g*}Y] = [X, Y].$$

Thus $\mathfrak{g}$ is closed under Lie brackets. □

The Lie algebra $\mathfrak{g}$ is called the *Lie algebra of the Lie group G*, and is denoted by $\mathrm{Lie}(G)$. The fundamental fact is that $\mathrm{Lie}(G)$ is finite-dimensional, and in fact has the same dimension as G itself, as the following theorem shows.

Theorem 15.3. *Let G be a Lie group and let $\mathfrak{g} = \mathrm{Lie}(G)$. The evaluation map $\mathfrak{g} \to T_eG$, given by $X \mapsto X_e$, is a vector space isomorphism. Thus $\mathfrak{g}$ is finite-dimensional, with dimension equal to $\dim G$.*

Proof. We will prove the theorem by constructing an inverse to the evaluation map. For each $V \in T_eG$, define a section $\widetilde{V}$ of TG by

$$\widetilde{V}_g = L_{g*}V.$$

If there is a left-invariant vector field on G whose value at the identity is V, clearly it has to be given by this formula.

First we need to check that $\widetilde{V}$ is in fact a smooth vector field. By Lemma 3.14(c) it suffices to show that $\widetilde{V}f$ is smooth whenever f is a smooth function on an open set $U \subset G$. Choose a smooth curve $\gamma\colon (-\varepsilon, \varepsilon) \to G$ such that $\gamma(0) = e$ and $\gamma'(0) = V$. Then for $g \in U$,

$$\begin{aligned}
(\widetilde{V} f)(g) &= \widetilde{V}_g f \\
&= (L_{g*} V) f \\
&= V(f \circ L_g) \\
&= \gamma'(0)(f \circ L_g) \\
&= \left.\frac{d}{dt}\right|_{t=0} (f \circ L_g \circ \gamma) \\
&= \left.\frac{d}{dt} f(g\gamma(t))\right|_{t=0}.
\end{aligned}$$

The expression $\varphi(g, t) = f(g\gamma(t))$ depends smoothly on (g, t), because it is a composition of group multiplication, f, and γ. Thus its t-derivative depends smoothly on g, and so $\widetilde{V} f$ is smooth.

Next we need to verify that $\widetilde{V}$ is left-invariant, which is to say that $L_{h*}\widetilde{V}_g = \widetilde{V}_{hg}$ for all $g, h \in G$. This follows from the definition of $\widetilde{V}$ and the fact that $L_h \circ L_g = L_{hg}$:

$$L_{h*}\widetilde{V}_g = L_{h*}(L_{g*}V) = L_{hg*}V = V_{hg}.$$

Thus $\widetilde{V} \in \mathfrak{g}$.

Finally, we check that the map $\tau\colon V \mapsto \widetilde{V}$ is an inverse for the evaluation map $\varepsilon\colon X \mapsto X_e$. On the one hand, given a vector $V \in T_eG$,

$$\varepsilon(\tau(V)) = (\widetilde{V})_e = L_{e*}V = V,$$

which shows that $\varepsilon \circ \tau$ is the identity on T_eG. On the other hand, given a vector field $X \in \mathfrak{g}$,

$$\tau(\varepsilon(X))_g = \widetilde{X_e}|_g = L_{g*}X_e = X_g,$$

which shows that $\tau \circ \varepsilon = \mathrm{Id}_{\mathfrak{g}}$. □

Example 15.4. Let us compute the Lie algebras of some familiar Lie groups.

(a) *Euclidean space* $\mathbb{R}^n$: Left translation by an element $b \in \mathbb{R}^n$ is given by the affine map $L_b(x) = x + b$, whose push-forward L_{b*} is represented by the identity matrix in standard coordinates. This implies that a vector field $V^i \partial/\partial x^i$ is left-invariant if and only if its coefficients V^i are constants. Because any two constant-coefficient vector fields commute (by formula (13.4)), the Lie algebra of $\mathbb{R}^n$ is abelian, and is isomorphic to $\mathbb{R}^n$ itself with the trivial bracket operation. In brief, $\mathrm{Lie}(\mathbb{R}^n) \cong \mathbb{R}^n$.

(b) *The circle group* $\mathbb{S}^1$: There is a unique unit tangent vector field T on $\mathbb{S}^1$ that is positively oriented with respect to the orientation induced by the outward normal. (If θ is any local angle coordinate on an open set $U \subset \mathbb{S}^1$, then $T = \partial/\partial\theta$ on U.) Because left translations are rotations, which preserve T, it follows that T is left-invariant, and therefore T spans the Lie algebra of $\mathbb{S}^1$. This Lie algebra is 1-dimensional and abelian, and therefore $\mathrm{Lie}(\mathbb{S}^1) \cong \mathbb{R}$.

(c) *Product groups*: If G and H are Lie groups, it is easy to check that the Lie algebra of the product group $G \times H$ is the product Lie algebra $\mathrm{Lie}(G) \times \mathrm{Lie}(H)$.

(d) *The n-torus* $\mathbb{T}^n$: Since $\mathbb{T}^n$ is the n-fold product of $\mathbb{S}^1$ with itself, its Lie algebra is isomorphic to $\mathbb{R} \times \cdots \times \mathbb{R} = \mathbb{R}^n$. In particular, it is abelian. If T_i is the oriented unit vector field on the ith $\mathbb{S}^1$ factor, then $(T_1, \ldots, T_n)$ is a basis for $\mathrm{Lie}(\mathbb{T}^n)$.

Theorem 15.3 has several useful corollaries. Recall that a smooth manifold is said to be *parallelizable* if its tangent bundle admits a global frame, or equivalently if its tangent bundle is trivial.

Corollary 15.5. *Every Lie group is parallelizable.*

Proof. Let G be a Lie group and let $\mathfrak{g}$ be its Lie algebra. Choosing any basis $X_1, \ldots, X_n$ for $\mathfrak{g}$, Theorem 15.3 shows that $X_1|_e, \ldots, X_n|_e$ form a basis for T_eG. For each $g \in G$, $L_{g*}\colon T_eG \to T_gG$ is an isomorphism taking $X_i|_e$ to $X_i|_g$, so $X_1|_g, \ldots, X_n|_g$ form a basis for T_gG at each $g \in G$. In other words (X_i) is a global frame for G. □

The proof of the preceding corollary actually shows that any basis for the Lie algebra of G is a global frame consisting of left-invariant vector fields. We will call any such frame a *left-invariant frame*.

Corollary 15.6. *Every Lie group is orientable.*

Proof. Proposition 10.5 shows that every parallelizable manifold is orientable. □

A tensor or differential form σ on a Lie group is said to be left-invariant if $L_g^*\sigma = \sigma$ for all $g \in G$.

Corollary 15.7. *Every compact oriented Lie group G has a unique left-invariant orientation form Ω with the property that $\int_G \Omega = 1$.*

Proof. Let $E_1, \ldots, E_n$ be a left-invariant global frame on G. By replacing E_1 with $-E_1$ if necessary, we may assume this frame is positively oriented. Let $\varepsilon^1, \ldots, \varepsilon^n$ be the dual coframe. Left invariance of E_j implies that

$$(L_g^*\varepsilon^i)(E_j) = \varepsilon^i(L_{g*}E_j) = \varepsilon^i(E_j),$$

which shows that $L_g^*\varepsilon^i = \varepsilon^i$, so ε^i is left-invariant.

Let $\Omega = \varepsilon^1 \wedge \cdots \wedge \varepsilon^n$. Then

$$L_g^*\Omega = L_g^*\varepsilon^1 \wedge \cdots \wedge L_g^*\varepsilon^n = \varepsilon^1 \wedge \cdots \wedge \varepsilon^n = \Omega,$$

so Ω is left-invariant as well. Because $\Omega(E_1, \dots, E_n) = 1 > 0$, Ω is an orientation form for the given orientation. Clearly any positive constant multiple of Ω is also a left-invariant orientation form. Conversely, if $\widetilde{\Omega}$ is any other left-invariant orientation form, we can write $\widetilde{\Omega}_e = c\Omega_e$ for some positive number c. Using left-invariance, we find that

$$\widetilde{\Omega}_g = L_{g^{-1}}^*\widetilde{\Omega}_e = cL_{g^{-1}}^*\Omega_e = c\Omega_g,$$

which proves that $\widetilde{\Omega}$ is a positive constant multiple of Ω.

Since G is compact and oriented, $\int_G \Omega$ makes sense, so we can define $\widetilde{\Omega} = (\int_G \Omega)^{-1}\Omega$. Clearly $\widetilde{\Omega}$ is the unique left-invariant orientation form for which G has unit volume. □

Remark. The n-form whose existence is asserted in this proposition is called the *Haar volume form* on G, and is often denoted dV. Similarly, the map $f \mapsto \int_G f\, dV$ is called the *Haar integral*. Observe that the proof above did not use the fact that G was compact until the last paragraph; thus every Lie group has a left-invariant orientation form that is uniquely defined up to a positive constant. It is only in the compact case, however, that we can use the volume normalization to single out a unique one.

The General Linear Group

Before going on with the general theory, we will explore one more fundamental example: the general linear group. In this chapter, let us denote the $n \times n$ identity matrix by I instead of I_n for brevity. Theorem 15.3 gives a vector space isomorphism between $\mathrm{Lie}(\mathrm{GL}(n,\mathbb{R}))$ and $T_I\,\mathrm{GL}(n,\mathbb{R})$ as usual. Because $\mathrm{GL}(n,\mathbb{R})$ is an open subset of the vector space $\mathfrak{gl}(n,\mathbb{R})$, its tangent space is naturally isomorphic to $\mathfrak{gl}(n,\mathbb{R})$ itself. The composition of these two isomorphisms gives a vector space isomorphism $\mathrm{Lie}(\mathrm{GL}(n,\mathbb{R})) \cong \mathfrak{gl}(n,\mathbb{R})$.

Both $\mathrm{Lie}(\mathrm{GL}(n,\mathbb{R})$ and $\mathfrak{gl}(n,\mathbb{R})$ have independently defined Lie algebra structures—the first coming from Lie brackets of vector fields and the second from commutator brackets of matrices. The next proposition shows that the natural vector space isomorphism between these spaces is in fact a Lie algebra isomorphism. Problem 15-14 shows that the analogous result holds for the complex general linear group.

Proposition 15.8 (Lie Algebra of the General Linear Group). *The composite map*

$$\mathfrak{gl}(n,\mathbb{R}) \to T_I\,\mathrm{GL}(n,\mathbb{R}) \to \mathrm{Lie}(\mathrm{GL}(n,\mathbb{R})) \tag{15.1}$$

gives a Lie algebra isomorphism between $\operatorname{Lie}(\mathrm{GL}(n,\mathbb{R}))$ *and the matrix algebra* $\mathfrak{gl}(n,\mathbb{R})$.

Proof. Using the matrix entries A^i_j as global coordinates on $\mathrm{GL}(n,\mathbb{R}) \subset \mathfrak{gl}(n,\mathbb{R})$, the natural isomorphism $\mathfrak{gl}(n,\mathbb{R}) \longleftrightarrow T_I \mathrm{GL}(n,\mathbb{R})$ takes the form

$$(B^i_j) \longleftrightarrow \sum_{ij} B^i_j \left.\frac{\partial}{\partial A^i_j}\right|_I .$$

(Because of the dual role of the indices i, j as coordinate indices and matrix row and column indices, it is impossible to use our summation conventions consistently in this context, so we will write the summation signs explicitly.)

Let $\mathfrak{g}$ denote the Lie algebra of $\mathrm{GL}(n,\mathbb{R})$. For any matrix $B = (B^i_j) \in \mathfrak{gl}(n,\mathbb{R})$, the left-invariant vector field $\widetilde{B} \in \mathfrak{g}$ corresponding to B is given by

$$\widetilde{B}_A = L_{A*}B = L_{A*}\left(\sum_{ij} B^i_j \left.\frac{\partial}{\partial A^i_j}\right|_I\right).$$

Since L_A is the restriction of the linear map $B \mapsto AB$ on $\mathfrak{gl}(n,\mathbb{R})$, its push-forward is represented in coordinates by exactly the same linear map. In other words,

$$\widetilde{B}_A = \sum_{ijk} A^i_j B^j_k \left.\frac{\partial}{\partial A^i_k}\right|_A . \tag{15.2}$$

Given two matrices $B, C \in \mathfrak{gl}(n,\mathbb{R})$, the Lie bracket of the corresponding left-invariant vector fields is given by

$$\begin{aligned}
[\widetilde{B}, \widetilde{C}] &= \left[\sum_{ijk} A^i_j B^j_k \frac{\partial}{\partial A^i_k}, \sum_{pqr} A^p_q C^q_r \frac{\partial}{\partial A^p_r}\right] \\
&= \sum_{ijkpqr} A^i_j B^j_k \frac{\partial}{\partial A^i_k}(A^p_q C^q_r) \frac{\partial}{\partial A^p_r} \\
&\qquad - \sum_{ijkpqr} A^p_q C^q_r \frac{\partial}{\partial A^p_r}(A^i_j B^j_k) \frac{\partial}{\partial A^i_k} \\
&= \sum_{ijkr} A^i_j B^j_k C^k_r \frac{\partial}{\partial A^i_r} - \sum_{pqrk} A^p_q C^q_r B^r_k \frac{\partial}{\partial A^p_k} \\
&= \sum_{ijkr} (A^i_j B^j_k C^k_r - A^i_j C^j_k B^k_r) \frac{\partial}{\partial A^i_r},
\end{aligned}$$

where we have used the fact that $\partial(A^p_q)/\partial A^i_k$ is equal to one if $p = i$ and $q = k$ and zero otherwise, and B^i_j and C^i_j are constants. Evaluating this last expression when A is equal to the identity matrix, we get

$$[\widetilde{B}, \widetilde{C}]_I = \sum_{ikr} (B^i_k C^k_r - C^i_k B^k_r) \left.\frac{\partial}{\partial A^i_r}\right|_I .$$

This is the vector corresponding to the matrix commutator bracket $[B, C]$. Since the left-invariant vector field $[\widetilde{B}, \widetilde{C}]$ is determined by its value at the identity, this implies that

$$[\widetilde{B}, \widetilde{C}] = \widetilde{[B, C]},$$

which is precisely the statement that the composite map (15.1) is a Lie algebra isomorphism. □

Induced Lie Algebra Homomorphisms

In this section, we show that a Lie homomorphism between Lie groups induces a Lie algebra homomorphism between their Lie algebras, and explore some of the consequences of this fact.

Theorem 15.9. *Let G and H be Lie groups and $\mathfrak{g}$ and $\mathfrak{h}$ their Lie algebras, and suppose $F\colon G \to H$ is a Lie group homomorphism. For every $X \in \mathfrak{g}$, there is a unique vector field in $\mathfrak{h}$ that is F-related to X. The map $F_*\colon \mathfrak{g} \to \mathfrak{h}$ so defined is a Lie algebra homomorphism.*

Proof. If there is any vector field $Y \in \mathfrak{h}$ that is F-related to X, it must satisfy $Y_e = F_* X_e$, and thus it must be uniquely defined by

$$Y = \widetilde{F_* X_e}.$$

To show that this Y is F-related to X, we note that the fact that F is a homomorphism implies

$$\begin{aligned} F(g_1 g_2) &= F(g_1) F(g_2) \\ &\implies F(L_{g_1} g_2) = L_{F(g_1)} F(g_2) \\ &\implies F \circ L_g = L_{F(g)} \circ F \\ &\implies F_* \circ L_{g*} = L_{F(g)*} \circ F_*. \end{aligned}$$

Thus

$$\begin{aligned} F_* X_g &= F_*(L_{g*} X_e) \\ &= L_{F(g)*} F_* X_e \\ &= L_{F(g)*} Y_e \\ &= Y_{F(g)}. \end{aligned}$$

This says precisely that X and Y are F-related.

Now, for each $X \in \mathfrak{g}$, let F_*X denote the unique vector field in $\mathfrak{h}$ that is F-related to X. It then follows immediately from the naturality of Lie brackets that $F_*[X,Y] = [F_*X, F_*Y]$, so F_* is a Lie algebra homomorphism. □

The map $F_*\colon \mathfrak{g} \to \mathfrak{h}$ whose existence is asserted in this theorem will be called the *induced Lie algebra homomorphism*.

Proposition 15.10 (Properties of the Induced Homomorphism).

(*a*) *The homomorphism* $\mathrm{Id}_{G*}\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ *induced by the identity map of* G *is the identity of* $\mathrm{Lie}(G)$.

(*b*) *If* $F_1\colon G \to H$ *and* $F_2\colon H \to K$ *are Lie group homomorphisms, then* $(F_2 \circ F_1)_* = (F_2)_* \circ (F_1)_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(K)$.

Proof. Both of these relations hold for push-forwards, and the value of the induced homomorphism on a left-invariant vector field X is defined by the push-forward of X_e. □

In the language of category theory, this proposition says that the assignment $G \mapsto \mathrm{Lie}(G)$, $F \mapsto F_*$ is a covariant functor from the category of Lie groups to the category of Lie algebras. The following corollary is immediate.

Corollary 15.11. *Isomorphic Lie groups have isomorphic Lie algebras.*

The next corollary has a bit more substance to it.

Corollary 15.12 (The Lie Algebra of a Lie Subgroup). *Suppose* $H \subset G$ *is a Lie subgroup. The subset* $\widetilde{\mathfrak{h}} \subset \mathrm{Lie}(G)$ *defined by*

$$\widetilde{\mathfrak{h}} = \{X \in \mathrm{Lie}(G) : X_e \in T_eH\}$$

is a Lie subalgebra of $\mathrm{Lie}(G)$ *canonically isomorphic to* $\mathrm{Lie}(H)$.

Proof. Let $\mathfrak{g} = \mathrm{Lie}(G)$ and $\mathfrak{h} = \mathrm{Lie}(G)$. It is clear that the inclusion map $\iota_H\colon H \hookrightarrow G$ is a Lie group homomorphism, so $\iota_{H*}(\mathfrak{h})$ is a Lie subalgebra of $\mathfrak{g}$. By the way we defined the induced Lie algebra homomorphism, this subalgebra is precisely the set of left-invariant vector fields on G whose value at the identity is of the form $\iota_{H*}V$ for some $V \in T_eH$. Since the push-forward map $\iota_{H*}\colon T_eH \to T_eG$ is the inclusion of T_eH as a subspace in T_eG, it follows that $\iota_{H*}(\mathfrak{h}) = \widetilde{\mathfrak{h}}$. To complete the proof, therefore, we need only show that $\iota_{H*}\colon \mathfrak{h} \to \mathfrak{g}$ is injective, so that it is an isomorphism onto its image. If $\iota_{H*}X = 0$, then in particular $\iota_{H*}X_e = 0$. Since ι_H is an immersion, this implies that $X_e = 0$ and therefore by left-invariance $X = 0$. Thus ι_{H*} is injective as claimed. □

Using this corollary, whenever H is a Lie subgroup of G, we will generally identify $\operatorname{Lie}(H)$ as a subalgebra of $\operatorname{Lie}(G)$. It is important to remember that elements of $\operatorname{Lie}(H)$ are only vector fields on H, and so, strictly speaking, are not elements of $\operatorname{Lie}(G)$. However, by the preceding lemma, every element of $\operatorname{Lie}(H)$ corresponds to a unique element of $\operatorname{Lie}(G)$, determined by its value at the identity, and the injection of $\operatorname{Lie}(H)$ into $\operatorname{Lie}(G)$ thus determined respects Lie brackets; so by thinking of $\operatorname{Lie}(H)$ as a subalgebra of $\operatorname{Lie}(G)$ we are not committing a grave error.

This identification is especially illuminating in the case of Lie subgroups of $\mathrm{GL}(n,\mathbb{R})$.

Example 15.13. Consider $\mathrm{O}(n)$ as a Lie subgroup of $\mathrm{GL}(n,\mathbb{R})$. By Example 5.26, it is equal to the level set $F^{-1}(I)$, where $F\colon \mathrm{GL}(n,\mathbb{R}) \to \mathrm{S}(n,\mathbb{R})$ is the submersion $F(A) = A^TA$. By Lemma 5.29, $T_I\,\mathrm{O}(n) = \operatorname{Ker} F_*\colon T_I\,\mathrm{GL}(n,\mathbb{R}) \to T_I\,\mathrm{S}(n,\mathbb{R})$. By the computation in Example 5.26, this push-forward is $F_*B = B^T + B$, so

$$\begin{aligned} T_I\,\mathrm{O}(n) &= \{B \in \mathfrak{gl}(n,\mathbb{R}) : B^T + B = 0\} \\ &= \{\text{skew-symmetric } n\times n \text{ matrices}\}. \end{aligned}$$

We denote this subspace of $\mathfrak{gl}(n,\mathbb{R})$ by $\mathfrak{o}(n)$. The preceding corollary then implies that $\mathfrak{o}(n)$ is a Lie subalgebra of $\mathfrak{gl}(n,\mathbb{R})$ isomorphic to $\operatorname{Lie}(\mathrm{O}(n))$. Notice that we did not even have to verify directly that $\mathfrak{o}(n)$ is a subalgebra.

As another application, we consider Lie covering groups. A Lie group homomorphism $F\colon G \to H$ that is also a smooth covering map is called a *covering homomorphism*, and we say G is a *(Lie) covering group* of H. By Problem 7-5, if G is connected and $F\colon G \to H$ is a surjective Lie group homomorphism with discrete kernel, then F is a covering homomorphism. Every connected Lie group G has a universal covering space $\widetilde{G}$, which is naturally a smooth manifold by Proposition 2.8 and a Lie group by Problem 7-6. We call $\widetilde{G}$ the *universal covering group* of G.

The next proposition shows how Lie algebras behave under covering homomorphisms.

Proposition 15.14. *Suppose G and H are Lie groups and $F\colon G \to H$ is a covering homomorphism. Then $F_*\colon \operatorname{Lie}(G) \to \operatorname{Lie}(H)$ is an isomorphism.*

Proof. Because a smooth covering map is a local diffeomorphism, the push-forward from T_eG to T_eH is an isomorphism, and therefore the induced Lie algebra homomorphism is an isomorphism. □

One-Parameter Subgroups

In this section, we explore another set of relationships among Lie algebras, vector fields, and Lie groups. It will give us yet another characterization of the Lie algebra of a Lie group.

Let G be a Lie group. We define a *one-parameter subgroup* of G to be a Lie group homomorphism $F\colon \mathbb{R} \to G$. Notice that, by this definition, a one-parameter subgroup is *not* a Lie subgroup of G, but rather a homomorphism into G. (However, as Problem 15-5 shows, the *image* of a one-parameter subgroup is a Lie subgroup.)

We will see shortly that the one-parameter subgroups are precisely the integral curves of left-invariant vector fields starting at the identity. Before we do so, however, we need the following lemma.

Lemma 15.15. *Every left-invariant vector field on a Lie group is complete.*

Proof. Let $\mathfrak{g}$ be the Lie algebra of the Lie group G, let $X \in \mathfrak{g}$, and let θ denote the flow of X. Suppose some maximal integral curve $\theta^{(g)}$ is defined on an interval $(a, b) \subset \mathbb{R}$, and assume that $b < \infty$. (The case $a > -\infty$ is handled similarly.) We will use left-invariance to define an integral curve on a slightly larger interval.

By Lemma 13.8, left-invariance of X means that

$$L_g \circ \theta_t = \theta_t \circ L_g \tag{15.3}$$

whenever the left-hand side is defined. Observe that the integral curve $\theta^{(e)}$ starting at the identity is defined at least on some interval $(-\varepsilon, \varepsilon)$ for $\varepsilon > 0$. Choose some $s \in (b - \varepsilon, b)$, and define a new curve $\gamma\colon (a, s + \varepsilon) \to G$ by

$$\gamma(t) = \begin{cases} \theta^{(g)}(t) & t \in (a, b) \\ L_{\theta_s(g)}(\theta_{t-s}(e)) & t \in (s - \varepsilon, s + \varepsilon). \end{cases}$$

By (15.3), when $s \in (a, b)$ and $|t - s| < \varepsilon$, we have

$$\begin{aligned} L_{\theta_s(g)}(\theta_{t-s}(e)) &= \theta_{t-s}(L_{\theta_s(g)}(e)) \\ &= \theta_{t-s}(\theta_s(g)) \\ &= \theta_t(g) \\ &= \theta^{(g)}(t), \end{aligned}$$

so the two definitions of γ agree where they overlap.

Now, γ is clearly an integral curve of X on (a,b), and for $t_0 \in (s-\varepsilon, s+\varepsilon)$ we use left-invariance of X to compute

$$\begin{aligned}
\gamma'(t_0) &= \left.\frac{d}{dt}\right|_{t=t_0} L_{\theta_s(g)}(\theta_{t-s}(e)) \\
&= L_{\theta_s(g)*} \left.\frac{d}{dt}\right|_{t=t_0} \theta^{(e)}(t-s) \\
&= L_{\theta_s(g)*} X_{\theta^{(e)}(t_0-s)} \\
&= X_{\gamma(t_0)}.
\end{aligned}$$

Thus γ is an integral curve of X defined for $t \in (a, s+\varepsilon)$. Since $s+\varepsilon > b$, this contradicts the maximality of $\theta^{(g)}$. □

Proposition 15.16. *Let G be a Lie group, and let $X \in \mathrm{Lie}(G)$. The integral curve of X starting at e is a one-parameter subgroup of G.*

Proof. Let θ be the flow of X, so that $\theta^{(e)} \colon \mathbb{R} \to G$ is the integral curve in question. Clearly $\theta^{(e)}$ is smooth, so we need only show that it is a group homomorphism, i.e., that $\theta^{(e)}(s+t) = \theta^{(e)}(s)\theta^{(e)}(t)$ for all $s, t \in \mathbb{R}$. Using (15.3) once again, we compute

$$\begin{aligned}
\theta^{(e)}(s)\theta^{(e)}(t) &= L_{\theta^{(e)}(s)}\theta_t(e) \\
&= \theta_t(L_{\theta^{(e)}(s)}(e)) \\
&= \theta_t(\theta^{(e)}(s)) \\
&= \theta_t(\theta_s(e)) \\
&= \theta_{t+s}(e) \\
&= \theta^{(e)}(t+s).
\end{aligned}$$

□

The main result of this section is that all one-parameter subgroups are obtained in this way.

Theorem 15.17. *Every one-parameter subgroup of a Lie group is an integral curve of a left-invariant vector field. Thus there are one-to-one correspondences*

$$\{\textit{one-parameter subgroups of } G\} \longleftrightarrow \mathrm{Lie}(G) \longleftrightarrow T_eG.$$

In particular, a one-parameter subgroup is uniquely determined by its initial tangent vector in T_eG.

Proof. Let $F \colon \mathbb{R} \to G$ be a one-parameter subgroup, and let $X = F_*(d/dt) \in \mathrm{Lie}(G)$, where we think of d/dt as a left-invariant vector field on

$\mathbb{R}$. To prove the theorem, it suffices to show that F is an integral curve of X. Recall that $F_*(d/dt)$ is defined as the unique left-invariant vector field on G that is F-related to d/dt. Therefore, for any $t_0 \in \mathbb{R}$,

$$F'(t_0) = F_* \left.\frac{d}{dt}\right|_{t_0} = X_{F(t_0)},$$

so F is an integral curve of X. □

Given $X \in \mathrm{Lie}(G)$, we will call the one-parameter subgroup determined in this way the *one-parameter subgroup generated by* X.

The one-parameter subgroups of the general linear group are not hard to compute explicitly.

Proposition 15.18. *For any* $B \in \mathfrak{gl}(n,\mathbb{R})$*, let*

$$e^B = \sum_{k=0}^{\infty} \frac{1}{k!} B^k. \tag{15.4}$$

This series converges to an invertible matrix $e^B \in \mathrm{GL}(n,\mathbb{R})$*, and the one-parameter subgroup of* $\mathrm{GL}(n,\mathbb{R})$ *generated by* $B \in \mathfrak{gl}(n,\mathbb{R})$ *is* $F(t) = e^{tB}$*.*

Proof. First we verify convergence. From Exercise A.24 in the Appendix, matrix multiplication satisfies $|AB| \le |A|\,|B|$, where the norm is the Euclidean norm on $\mathfrak{gl}(n,\mathbb{R})$ under its obvious identification with $\mathbb{R}^{n^2}$. It follows by induction that $|B^k| \le |B|^k$. The Weierstrass M-test shows that (15.4) converges uniformly on any bounded subset of $\mathfrak{gl}(n,\mathbb{R})$ (by comparison with the series $\sum_k (1/k!)C^k = e^C$).

Fix $B \in \mathfrak{gl}(n,\mathbb{R})$. The one-parameter subgroup generated by B is an integral curve of the left-invariant vector field $\widetilde{B}$, and therefore satisfies the ODE initial value problem

$$F'(t) = \widetilde{B}_{F(t)},$$
$$F(0) = I.$$

Using formula (15.2) for $\widetilde{B}$, the condition for F to be an integral curve can be rewritten as

$$(F^i_k)'(t) = F^i_j(t) B^j_k,$$

or in matrix notation

$$F'(t) = F(t)B.$$

We will show that $F(t) = e^{tB}$ satisfies this equation. Since $F(0) = I$, this implies that F is the unique integral curve of $\widetilde{B}$ starting at the identity and is therefore the desired one-parameter subgroup.

To see that F is differentiable, we note that differentiating the series (15.4) formally term-by-term yields the result

$$\begin{aligned} F'(t) &= \sum_{k=1}^{\infty} \frac{k}{k!} t^{k-1} B^k \\ &= \left(\sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} B^{k-1} \right) B \\ &= F(t)B. \end{aligned}$$

Since the differentiated series converges uniformly on compact sets (because it is the same series!), the term-by-term differentiation is justified. A similar argument shows that $F'(t) = BF(t)$. By smoothness of solutions to ODEs, F is a smooth curve.

It remains only to show that $F(t)$ is invertible for all t, so that F actually takes its values in $\mathrm{GL}(n, \mathbb{R})$. If we let $\sigma(t) = F(t)F(-t) = e^{tB}e^{-tB}$, then σ is a smooth curve in $\mathfrak{gl}(n, \mathbb{R})$, and by the previous computation and the product rule it satisfies

$$\sigma'(t) = (F(t)B)F(-t) - F(t)(BF(-t)) = 0.$$

Therefore σ is the constant curve $\sigma(t) \equiv \sigma(0) = I$, which is to say that $F(t)F(-t) = I$. Substituting $-t$ for t, we obtain $F(-t)F(t) = I$, which shows that $F(t)$ is invertible and $F(t)^{-1} = F(-t)$. □

Next we would like to compute the one-parameter subgroups of subgroups of $\mathrm{GL}(n, \mathbb{R})$, such as $\mathrm{O}(n)$. To do so, we need the following result.

Proposition 15.19. *Suppose $H \subset G$ is a Lie subgroup. The one-parameter subgroups of H are precisely those one-parameter subgroups of G whose initial tangent vectors lie in T_eH.*

Proof. Let $F\colon \mathbb{R} \to H$ be a one-parameter subgroup. Then the composite map

$$\mathbb{R} \xrightarrow{F} H \hookrightarrow G$$

is a Lie group homomorphism and thus a one-parameter subgroup of G, which clearly satisfies $F'(0) \in T_eH$.

Conversely, suppose $F\colon \mathbb{R} \to G$ is a one-parameter subgroup whose initial tangent vector lies in T_eH. Let $\widetilde{F}\colon \mathbb{R} \to H$ be the one-parameter subgroup of H with the same initial tangent vector $\widetilde{F}'(0) = F'(0) \in T_eH \subset T_eG$. As in the preceding paragraph, by composing with the inclusion map, we can also consider $\widetilde{F}$ as a one-parameter subgroup of G. Since F and $\widetilde{F}$ are both one-parameter subgroups of G with the same initial tangent vector, they must be equal. □

Example 15.20. If H is a Lie subgroup of $\mathrm{GL}(n,\mathbb{R})$, the preceding proposition shows that the one-parameter subgroups of H are precisely the maps of the form $F(t) = e^{tB}$ for $B \in \mathfrak{h}$, where $\mathfrak{h} \subset \mathfrak{gl}(n,\mathbb{R})$ is the subalgebra corresponding to $\mathrm{Lie}(H)$ as in Corollary 15.12. For example, taking $H = \mathrm{O}(n)$, this shows that the exponential of any skew-symmetric matrix is orthogonal.

The Exponential Map

In the preceding section, we saw that the matrix exponential maps $\mathfrak{gl}(n,\mathbb{R})$ to $\mathrm{GL}(n,\mathbb{R})$ and takes each line through the origin to a one-parameter subgroup. This has a powerful generalization to arbitrary Lie groups.

Given a Lie group G with Lie algebra $\mathfrak{g}$, define a map $\exp\colon \mathfrak{g} \to G$, called the *exponential map* of G, by letting $\exp X = F(1)$, where F is the one-parameter subgroup generated by X, or equivalently the integral curve of X starting at the identity.

Example 15.21. The results of the preceding section show that the exponential map of $\mathrm{GL}(n,\mathbb{R})$ (or any Lie subgroup of it) is given by $\exp A = e^A$. This, obviously, is the reason for the term exponential map.

Proposition 15.22 (Properties of the Exponential Map). *Let G be a Lie group and $\mathfrak{g}$ its Lie algebra.*

(*a*) *The exponential map is smooth.*

(*b*) *For any $X \in \mathfrak{g}$, $F(t) = \exp tX$ is the one-parameter subgroup of G generated by X.*

(*c*) *For any $X \in \mathfrak{g}$, $\exp(s+t)X = \exp sX \exp tX$.*

(*d*) *The push-forward $\exp_*\colon T_0\mathfrak{g} \to T_eG$ is the identity map, under the canonical identifications of both $T_0\mathfrak{g}$ and T_eG with $\mathfrak{g}$ itself.*

(*e*) *The exponential map is a diffeomorphism from some neighborhood of 0 in $\mathfrak{g}$ to a neighborhood of e in G.*

(*f*) *For any Lie group homomorphism $F\colon G \to H$, the following diagram commutes:*

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "F_*"] \arrow[d, "\exp"'] & \mathfrak{h} \arrow[d, "\exp"] \\ G \arrow[r, "F"'] & H \end{tikzcd}$$

(*g*) *The flow θ of a left-invariant vector field X is given by $\theta_t = R_{\exp tX}$ (right multiplication by $\exp tX$).*

Proof. For this proof, for any $X \in \mathfrak{g}$ we let $\theta_{(X)}$ denote the flow of X. To prove (a), we need to show that the expression $\theta^{(e)}_{(X)}(1)$ depends smoothly on X, which amounts to showing that the flow varies smoothly as the vector field varies. This is a situation not covered by the fundamental theorem on flows, but we can reduce it to that theorem by the following trick. Define a smooth vector field Ξ on the product manifold $G \times \mathfrak{g}$ by

$$\Xi_{(g,X)} = (X_g, 0) \in T_g G \times T_X \mathfrak{g} \cong T_{(g,X)}(G \times \mathfrak{g}).$$

It is easy to verify that the flow Θ of Ξ is given by

$$\Theta_t(g, X) = (\theta_{(X)t}(g), X).$$

By the fundamental theorem on flows, Θ is a smooth map. Since $\exp X = \pi_1(\Theta_1(e, X))$, where $\pi_1\colon G \times \mathfrak{g} \to G$ is the projection, it follows that exp is smooth.

Since the one-parameter subgroup generated by X is equal to the integral curve of X starting at e, to prove (b) it suffices to show that $\exp tX = \theta^{(e)}_{(X)}(t)$, or in other words that

$$\theta^{(e)}_{(tX)}(1) = \theta^{(e)}_{(X)}(t). \tag{15.5}$$

In fact, we will prove that for all $s, t \in \mathbb{R}$,

$$\theta^{(e)}_{(tX)}(s) = \theta^{(e)}_{(X)}(st), \tag{15.6}$$

which clearly implies (15.5).

To prove (15.6), fix $t \in \mathbb{R}$ and define a smooth curve $\gamma\colon \mathbb{R} \to G$ by

$$\gamma(s) = \theta^{(e)}_{(X)}(st).$$

By the chain rule,

$$\gamma'(s) = t(\theta^{(e)}_{(X)})'(st) = tX_{\gamma(s)},$$

so γ is an integral curve of the vector field tX. Since $\gamma(0) = e$, by uniqueness of integral curves we must have $\gamma(s) = \theta^{(e)}_{(tX)}(s)$, which is (15.6). This proves (b).

Next, (c) follows immediately from (b), which shows that $t \mapsto \exp tX$ is a group homomorphism.

To prove (d), let $X \in \mathfrak{g}$ be arbitrary, and let $\sigma\colon \mathbb{R} \to \mathfrak{g}$ be the curve $\sigma(t) = tX$. Then $\sigma'(0) = X$, and (b) implies

$$\begin{aligned}
\exp_* X &= \exp_* \sigma'(0) \\
&= (\exp \circ \sigma)'(0) \\
&= \left.\frac{d}{dt}\right|_{t=0} \exp tX \\
&= X.
\end{aligned}$$

Part (e) then follows immediately from (d) and the inverse function theorem.

Next, to prove (f) we need to show that $\exp(F_*X) = F(\exp X)$ for any $X \in \mathfrak{g}$. In fact, we will show that for all $t \in \mathbb{R}$,

$$\exp(tF_*X) = F(\exp tX).$$

The left-hand side is, by (b), the one-parameter subgroup generated by F_*X. Thus, if we put $\sigma(t) = F(\exp tX)$, it suffices to show that σ is a group homomorphism satisfying $\sigma'(0) = F_*X$. We compute

$$\sigma'(0) = \left.\frac{d}{dt}\right|_{t=0} F(\exp tX) = F_* \left.\frac{d}{dt}\right|_{t=0} \exp tX = F_*X$$

and

$$\begin{aligned} \sigma(s+t) &= F(\exp(s+t)X) \\ &= F(\exp sX \exp tX) && \text{(by (c))} \\ &= F(\exp sX) F(\exp tX) && \text{(since } F \text{ is a homomorphism)} \\ &= \sigma(s)\sigma(t). \end{aligned}$$

Finally, to show that $\theta_{(X)t} = R_{\exp tX}$, we use part (b) and (15.3) to show that for any $g \in G$,

$$\begin{aligned} R_{\exp tX}(g) &= g \exp tX \\ &= L_g(\exp tX) \\ &= L_g(\theta_{(X)t}(e)) \\ &= \theta_{(X)t}(L_g(e)) \\ &= \theta_{(X)t}(g). \end{aligned}$$

□

In the remainder of this chapter, we present a variety of important applications of the exponential map. The first is a powerful technique for computing the Lie algebra of a subgroup.

Lemma 15.23. *Let G be a Lie group, and let $H \subset G$ be a Lie subgroup. Inclusion $H \hookrightarrow G$ induces an isomorphism between* $\mathrm{Lie}(H)$ *and the subalgebra* $\mathfrak{h} \subset \mathrm{Lie}(G)$ *given by*

$$\mathfrak{h} = \{X \in \mathrm{Lie}(G) : \exp tX \in H \text{ for all } t \in \mathbb{R}\}.$$

Proof. Let X be an arbitrary element of $\mathrm{Lie}(G)$. Suppose first that $X \in \mathfrak{h}$. Letting γ denote the curve $\gamma(t) = \exp tX$, the fact that $\gamma(t)$ lies in H for all t means that $X_e = \gamma'(0) \in T_eH$, which means that X is in the image of $\mathrm{Lie}(H)$ under inclusion (see Corollary 15.12). Conversely, if X is in the image of $\mathrm{Lie}(H)$, then $X_e \in T_eH$, which implies by Proposition 15.19 that $\exp tX \in H$ for all t. □

Corollary 15.24. *Let $G \subset \mathrm{GL}(n,\mathbb{R})$ be a Lie subgroup, and define a subset $\mathfrak{g} \subset \mathfrak{gl}(n,\mathbb{R})$ by*

$$\mathfrak{g} = \{B \in \mathfrak{gl}(n,\mathbb{R}) : e^{tB} \in G \text{ for all } t \in \mathbb{R}\}.$$

Then $\mathfrak{g}$ is a Lie subalgebra of $\mathfrak{gl}(n,\mathbb{R})$ canonically isomorphic to $\mathrm{Lie}(G)$.

As an application, we will determine the Lie algebra of $\mathrm{SL}(n,\mathbb{R})$. First we need the following lemma.

Lemma 15.25. *The matrix exponential satisfies the identity*

$$\det e^{A} = e^{\operatorname{tr} A}. \tag{15.7}$$

Proof. Let $A \in \mathfrak{gl}(n,\mathbb{R})$ be arbitrary, and consider the smooth function $\tau\colon \mathbb{R} \to \mathbb{R}$ given by

$$\tau(t) = e^{-\operatorname{tr} tA} \det e^{tA}.$$

We compute the derivatives of the two factors separately. First, using the result of Problem 4-10 and the chain rule,

$$\begin{aligned}
\frac{d}{dt}(\det e^{tA}) &= d(\det)_{e^{tA}}\left(\frac{d}{dt}e^{tA}\right)\\
&= d(\det)_{e^{tA}}(e^{tA}A)\\
&= (\det e^{tA}) \operatorname{tr}\left((e^{tA})^{-1}(e^{tA}A)\right)\\
&= (\det e^{tA}) \operatorname{tr} A.
\end{aligned}$$

Then, because the trace is linear,

$$\begin{aligned}
\frac{d}{dt}(e^{-\operatorname{tr} tA}) &= e^{-\operatorname{tr} tA}\frac{d}{dt}(-\operatorname{tr} tA)\\
&= e^{-\operatorname{tr} tA}(-\operatorname{tr} A).
\end{aligned}$$

Therefore, by the product rule,

$$\begin{aligned}
\tau'(t) &= e^{-\operatorname{tr} tA}\frac{d}{dt}(\det e^{tA}) + \det e^{tA}\frac{d}{dt}(e^{-\operatorname{tr} tA})\\
&= e^{-\operatorname{tr} tA}\det e^{tA}\operatorname{tr} A + \det e^{tA}e^{-\operatorname{tr} tA}(-\operatorname{tr} A)\\
&= 0.
\end{aligned}$$

Consequently $\tau(t) = \tau(0) = 1$ for all t. In particular, taking $t = 1$, this implies (15.7). □

Example 15.26. Let $\mathfrak{sl}(n,\mathbb{R}) \subset \mathfrak{gl}(n,\mathbb{R})$ be the set of trace-free matrices:

$$\mathfrak{sl}(n,\mathbb{R}) \subset \mathfrak{gl}(n,\mathbb{R}) = \{B \in \mathfrak{gl}(n,\mathbb{R}) : \operatorname{tr} B = 0\}.$$

If $B \in \mathfrak{sl}(n,\mathbb{R})$, then the preceding lemma shows that $\det e^{tB} = e^{trtB} = 1$, so $e^{tB} \in \mathrm{SL}(n,\mathbb{R})$ for all t. Conversely, if $e^{tB} \in \mathrm{SL}(n,\mathbb{R})$ for all t, then $1 = \det e^{tB} = e^{\operatorname{tr} tB} = e^{t \operatorname{tr} B}$, which immediately implies that $\operatorname{tr} B = 0$. Thus by Corollary 15.24, $\mathfrak{sl}(n,\mathbb{R})$ is a Lie subalgebra of $\mathfrak{gl}(n,\mathbb{R})$, isomorphic to $\operatorname{Lie}(\mathrm{SL}(n,\mathbb{R}))$.

The next lemma is a technical result that will be used below.

Lemma 15.27. *Let G be a Lie group and $\mathfrak{g}$ its Lie algebra.*

(*a*) *If $m\colon G \times G \to G$ denotes the multiplication map, then $m_*\colon T_eG \times T_eG \to T_eG$ is given by $m_*(X,Y) = X + Y$.*

(*b*) *If $A, B \subset \mathfrak{g}$ are complementary linear subspaces of $\mathfrak{g}$, then the map $A \times B \to G$ given by $(X,Y) \mapsto \exp X \exp Y$ is a diffeomorphism from some neighborhood of $(0,0)$ in $A \times B$ to a neighborhood of e in G.*

Exercise 15.3. Prove this lemma.

The following proposition shows how the group structure of a Lie group is reflected "infinitesimally" in the algebraic structure of its Lie algebra. The second formula, in particular, shows how the Lie bracket expresses the leading term in the Taylor series expansion of a group commutator.

In the statement of the proposition, we use the following standard notation from analysis: The expression $O(t^k)$ means any $\mathfrak{g}$-valued function of t that is bounded by a constant multiple of $|t|^k$ as $t \to 0$. (The bound can be expressed in terms of any norm on $\mathfrak{g}$; since all norms on a finite-dimensional vector space are equivalent, the definition is independent of which norm is chosen.)

Proposition 15.28. *Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. For any $X, Y \in \mathfrak{g}$, the exponential map satisfies*

$$(\exp tX)(\exp tY) = \exp(t(X+Y) + \tfrac{1}{2}t^2[X,Y] + O(t^3)), \tag{15.8}$$

$$(\exp tX)(\exp tY)(\exp -tX)(\exp -tY) = \exp(t^2[X,Y] + O(t^3)), \tag{15.9}$$

whenever $t \in \mathbb{R}$ is sufficiently close to 0.

Proof. Let $X \in \mathfrak{g}$, and let θ denote the flow of X. If f is any smooth function defined on an open subset of G,

$$\frac{d}{dt} f(g \exp tX) = \frac{d}{dt} f(\theta_t(g)) = X_{\theta_t(g)} f = Xf(g \exp tX), \tag{15.10}$$

because $t \mapsto g \exp tX = \theta_t(g)$ is an integral curve of X. Applying this same formula to the function Xf yields

$$\frac{d^2}{dt^2} f(g \exp tX) = \frac{d}{dt} Xf(g \exp tX) = X^2 f(g \exp tX).$$

In particular, if f is defined in a neighborhood of the identity, evaluating these equations at $g = e$ and $t = 0$ yields

$$\left.\frac{d}{dt} f(\exp tX)\right|_{t=0} = Xf(e), \tag{15.11}$$

$$\left.\frac{d^2}{dt^2} f(\exp tX)\right|_{t=0} = X^2 f(e). \tag{15.12}$$

By Taylor's theorem in one variable, therefore, we can write

$$f(\exp tX) = f(e) + tXf(e) + \tfrac{1}{2}t^2 X^2 f(e) + O(t^3). \tag{15.13}$$

Now, let $X, Y \in \mathfrak{g}$ be fixed, and define a function $u\colon \mathbb{R}^2 \to \mathbb{R}$ by $u(s,t) = f(\exp sX \exp tY)$. We will prove the proposition by analyzing the Taylor series of $u(t,t)$ about $t = 0$. First, by the chain rule and (15.11),

$$\left.\frac{d}{dt} u(t,t)\right|_{t=0} = \frac{\partial u}{\partial s}(0,0) + \frac{\partial u}{\partial t}(0,0) = Xf(e) + Yf(e). \tag{15.14}$$

By (15.12), the pure second derivatives of u are given by

$$\frac{\partial^2 u}{\partial t^2}(0,0) = X^2 f(e), \qquad \frac{\partial^2 u}{\partial s^2}(0,0) = Y^2 f(e).$$

To compute the mixed second derivative, we apply (15.10) twice to obtain

$$\begin{aligned}\frac{\partial^2 u}{\partial s \partial t}(0,0) &= \left.\frac{\partial}{\partial s}\right|_{s=0} \left(\left.\frac{\partial}{\partial t}\right|_{t=0} f(\exp sX \exp tY) \right) \\ &= \left.\frac{\partial}{\partial s}\right|_{s=0} Yf(\exp sX) \\ &= XYf(e).\end{aligned} \tag{15.15}$$

Therefore, by the chain rule,

$$\begin{aligned}\left.\frac{d^2}{dt^2} u(t,t)\right|_{t=0} &= \frac{\partial^2 u}{\partial s^2}(0,0) + 2\frac{\partial^2 u}{\partial s \partial t}(0,0) + \frac{\partial^2 u}{\partial t^2}(0,0) \\ &= X^2 f(e) + 2XYf(e) + Y^2 f(e).\end{aligned}$$

Taylor's theorem then yields

$$\begin{aligned}f(\exp tX \exp tY) &= u(t,t) \\ &= u(0,0) + t\left.\frac{d}{dt} u(t,t)\right|_{t=0} + \tfrac{1}{2}t^2 \left.\frac{d^2}{dt^2} u(t,t)\right|_{t=0} \\ &\qquad + O(t^3) \\ &= f(e) + t(Xf(e) + Yf(e)) \\ &\qquad + \tfrac{1}{2}t^2(X^2 f(e) + 2XYf(e) + Y^2 f(e)) + O(t^3).\end{aligned} \tag{15.16}$$

Because the exponential map is a diffeomorphism on some neighborhood U of the origin in $\mathfrak{g}$, there is a smooth curve in U defined by $\gamma(t) = \exp^{-1}(\exp tX \exp tY)$ for t sufficiently near zero. It obviously satisfies $\gamma(0) = 0$ and

$$\exp tX \exp tY = \exp \gamma(t).$$

The implied constant in the $O(t^3)$ error term in (15.13) can be taken to be independent of X as long as X stays in a compact subset of $\mathfrak{g}$, as can be seen by expressing the error explicitly in terms of the remainder term in the Taylor series of the smooth function $f \circ \exp$ (in terms of any basis for $\mathfrak{g}$). Therefore, writing the Taylor series of $\gamma(t)$ as

$$\gamma(t) = tA + t^2 B + O(t^3)$$

for some fixed $A, B \in \mathfrak{g}$, we can substitute $\gamma(t)$ for tX in (15.13) and expand to obtain

$$\begin{aligned} f(\exp tX \exp tY) &= f(\exp \gamma(t)) \\ &= f(e) + tAf(e) + t^2 Bf(e) + \tfrac{1}{2}t^2 A^2 f(e) + O(t^3). \end{aligned} \tag{15.17}$$

Comparing like powers of t in (15.17) and (15.16), we see that

$$\begin{aligned} Af(e) &= Xf(e) + Yf(e), \\ Bf(e) + \tfrac{1}{2}A^2 f(e) &= \tfrac{1}{2}(X^2 f(e) + 2XYf(e) + Y^2 f(e)). \end{aligned}$$

Since this is true for every f, the first equation implies $A = X+Y$. Inserting this into the second equation, we obtain

$$\begin{aligned} Bf(e) &= \tfrac{1}{2}(X^2 f(e) + 2XYf(e) + Y^2 f(e)) - \tfrac{1}{2}(X+Y)^2 f(e) \\ &= \tfrac{1}{2}X^2 f(e) + XYf(e) + \tfrac{1}{2}Y^2 f(e) \\ &\qquad - \tfrac{1}{2}X^2 f(e) - \tfrac{1}{2}XYf(e) - \tfrac{1}{2}YXf(e) - \tfrac{1}{2}Y^2 f(e) \\ &= \tfrac{1}{2}(XYf(e) - YXf(e)) \\ &= \tfrac{1}{2}[X,Y]f(e). \end{aligned}$$

This implies $B = \frac{1}{2}[X,Y]$, which completes the proof of (15.8).

Finally, (15.9) is proved by applying (15.8) twice:

$$\begin{aligned} &(\exp tX)(\exp tY)(\exp -tX)(\exp -tY) \\ &\qquad = \exp(t(X+Y) + \tfrac{1}{2}t^2[X,Y] + O(t^3)) \times \\ &\qquad\qquad \exp(t(-X-Y) + \tfrac{1}{2}t^2[X,Y] + O(t^3)) \\ &\qquad = \exp(t^2[X,Y] + O(t^3)). \end{aligned}$$

$\square$

The formulas above are special cases of a much more general formula, called the *Baker-Campbell-Hausdorff formula*, which gives recursive expressions for all the terms of the Taylor series of $\gamma(t)$ in terms of X, Y, $[X, Y]$, and iterated brackets such as $[X, [X, Y]]$ and $[Y, [X, [X, Y]]]$. The full formula can be found in [Var84].

The Closed Subgroup Theorem

The next theorem is one of the most powerful applications of the exponential map. For example, it allows us to strengthen Theorem 7.15 about quotients of Lie groups, because we need only assume that the subgroup H is topologically closed in G, not that it is a closed Lie subgroup. A similar remark applies to Proposition 7.21.

Theorem 15.29 (Closed Subgroup Theorem). *Suppose G is a Lie group and $H \subset G$ is a subgroup that is also a closed subset. Then H is an embedded Lie subgroup.*

Proof. By Proposition 5.41, it suffices to show that H is an embedded submanifold of G. We begin by identifying a subspace of the Lie algebra of G that will turn out to be the Lie algebra of H.

Let $\mathfrak{g}$ denote the Lie algebra of G, and define a subset $\mathfrak{h} \subset \mathfrak{g}$ by

$$\mathfrak{h} = \{X \in \mathfrak{g} : \exp tX \in H \text{ for all } t \in \mathbb{R}\}.$$

We need to show that $\mathfrak{h}$ is a vector subspace of $\mathfrak{g}$. It is obvious from the definition that if $X \in \mathfrak{h}$, then $tX \in \mathfrak{h}$ for all $t \in \mathbb{R}$. To see that $\mathfrak{h}$ is closed under vector addition, let $X, Y \in \mathfrak{h}$ be arbitrary. Observe that (15.8) implies that for any $t \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$\exp \frac{t}{n} X \exp \frac{t}{n} Y = \exp\left(\frac{t}{n}(X+Y) + O\left(\frac{t^2}{n^2}\right)\right),$$

and a simple induction using (15.8) again shows that

$$\begin{aligned}\left(\exp \frac{t}{n} X \exp \frac{t}{n} Y\right)^n &= \left(\exp\left(\frac{t}{n}(X+Y) + O\left(\frac{t^2}{n^2}\right)\right)\right)^n \\ &= \exp\left(t(X+Y) + O\left(\frac{t^2}{n}\right)\right).\end{aligned}$$

Taking the limit as $n \to \infty$, we obtain

$$\exp t(X+Y) = \lim_{n\to\infty}\left(\exp \frac{t}{n} X \exp \frac{t}{n} Y\right)^n,$$

which is in H because H is closed in G. Thus $X+Y \in \mathfrak{h}$, and so $\mathfrak{h}$ is a subspace. (In fact, (15.9) can be used in a similar way to prove that $\mathfrak{h}$ is a Lie subalgebra, but we will not need this.)

Next we will show that there is a neighborhood U of the origin in $\mathfrak{g}$ on which the exponential map of G is a diffeomorphism, and which has the property that

$$\exp(U \cap \mathfrak{h}) = (\exp U) \cap H. \tag{15.18}$$

This will enable us to construct a slice chart for H near the identity, and we will then use left translation to get a slice chart in a neighborhood of any point of H.

If U is any neighborhood of $0 \in \mathfrak{g}$ on which exp is a diffeomorphism, then $\exp(U \cap \mathfrak{h}) \subset (\exp U) \cap H$ by definition of $\mathfrak{h}$. So to find a neighborhood satisfying (15.18), all we need to do is to show that U can be chosen small enough that $(\exp U) \cap H \subset \exp(U \cap \mathfrak{h})$. Assume this is not possible. Let $\{U_i\}$ be any countable neighborhood basis at $0 \in \mathfrak{g}$ (for example, a countable sequence of coordinate balls whose radii approach zero). The assumption implies that for each i, there exists $h_i \in (\exp U_i) \cap H$ such that $h_i \notin \exp(U_i \cap \mathfrak{h})$.

Choose a basis $E_1, \dots, E_k$ for $\mathfrak{h}$ and extend it to a basis $E_1, \dots, E_m$ for $\mathfrak{g}$. Let $\mathfrak{b}$ be the subspace spanned by $E_{k+1}, \dots, E_m$, so that $\mathfrak{g} = \mathfrak{h} \times \mathfrak{b}$ as vector spaces. By Lemma 15.27, as soon as i is large enough, the map from $\mathfrak{h} \times \mathfrak{b}$ to G given by $(X, Y) \mapsto \exp X \exp Y$ is a diffeomorphism from U_i to a neighborhood of e in G. Therefore we can write

$$h_i = \exp X_i \exp Y_i$$

for some $X_i \in U_i \cap \mathfrak{h}$ and $Y_i \in U_i \cap \mathfrak{b}$, with $Y_i \neq 0$ because $h_i \notin \exp(U_i \cap \mathfrak{h})$. Since $\{U_i\}$ is a neighborhood basis, $Y_i \to 0$ as $i \to \infty$. Observe that $\exp X_i \in H$ by definition of $\mathfrak{h}$, so it follows that $\exp Y_i = (\exp X_i)^{-1} h_i \in H$ as well.

The basis $\{E_j\}$ induces a vector space isomorphism $E\colon \mathfrak{g} \cong \mathbb{R}^m$. Let $|\cdot|$ denote the Euclidean norm induced by this isomorphism and define $c_i = |Y_i|$, so that $c_i \to 0$ as $i \to \infty$. The sequence $\{c_i^{-1} Y_i\}$ lies in the unit sphere in $\mathfrak{b}$ with respect to this norm, so replacing it by a subsequence we may assume that $c_i^{-1} Y_i \to Y \in \mathfrak{b}$, with $|Y| = 1$ by continuity. In particular, $Y \neq 0$. We will show that $\exp tY \in H$ for all $t \in \mathbb{R}$, which implies that $Y \in \mathfrak{h}$. Since $\mathfrak{h} \cap \mathfrak{b} = \{0\}$, this is a contradiction.

Let $t \in \mathbb{R}$ be arbitrary, and for each i, let n_i be the greatest integer less than or equal to t/c_i. Then

$$\left| n_i - \frac{t}{c_i} \right| \leq 1,$$

which implies

$$|n_i c_i - t| \leq c_i \to 0,$$

so $n_i c_i \to t$. Thus

$$n_i Y_i = (n_i c_i)(c_i^{-1} Y_i) \to tY,$$

which implies $\exp n_i Y_i \to \exp tY$ by continuity. But $\exp n_i Y_i = (\exp Y_i)^n \in H$, so the fact that H is closed implies $\exp tY \in H$. This completes the proof of the existence of U satisfying (15.18).

The composite map $\varphi = E \circ \exp^{-1} \colon \exp U \to \mathbb{R}^m$ is easily seen to be a coordinate chart for G, and by our choice of basis, $\varphi((\exp U) \cap H) = E(U \cap \mathfrak{h})$ is the slice obtained by setting the last $m-k$ coordinates equal to zero. Moreover, if $h \in H$ is arbitrary, left multiplication L_h is a diffeomorphism from $\exp U$ to a neighborhood of h. Since H is a subgroup, $L_h(H) = H$, and so

$$L_h((\exp U) \cap H) = L_h(\exp U) \cap H,$$

and $\varphi \circ L_h^{-1}$ is easily seen to be a slice chart for H in a neighborhood of h. Thus H is a regular submanifold of G, hence a Lie subgroup. □

The following corollary summarizes the closed subgroup theorem and Proposition 5.41.

Corollary 15.30. *If G is a Lie group and H is any subgroup of G, the following are equivalent:*

(*a*) *H is closed in G.*

(*b*) *H is an embedded Lie subgroup.*

Lie Subalgebras and Lie Subgroups

Earlier in this chapter, we saw that a Lie subgroup of a Lie group gives rise to a Lie subalgebra of its Lie algebra. In this section, we show that the converse is true: Every Lie subalgebra corresponds to some Lie subgroup. This result has important consequences that we will explore in the remainder of the chapter.

Theorem 15.31. *Suppose G is a Lie group with Lie algebra $\mathfrak{g}$. If $\mathfrak{h}$ is any Lie subalgebra of $\mathfrak{g}$, then there is a unique connected Lie subgroup of G whose Lie algebra is $\mathfrak{h}$* (*under the canonical identification of the Lie algebra of a subgroup with a Lie subalgebra of $\mathfrak{g}$*).

Proof. Define a distribution $D \subset TG$ by

$$D_g = \{X_g \in T_gG : X \in \mathfrak{h}\}.$$

If $(X_1, \dots, X_k)$ is any basis for $\mathfrak{h}$, then clearly D_g is spanned by $X_1|_g, \dots, X_k|_g$ at any $g \in G$. Thus D is locally (in fact, globally) spanned by smooth vector fields, so it is a smooth subbundle of TG. Moreover, because $[X_i, X_j] \in \mathfrak{h}$ for each i, j, D is involutive. Let $\mathcal{H}$ denote the foliation determined by D, and for any $g \in G$, let $\mathcal{H}_g$ denote the leaf of $\mathcal{H}$ containing g.

If g, g' are arbitrary elements of G, then

$$\begin{aligned} L_{g*}(D_{g'}) &= \operatorname{span}(L_{g*}X_1|_{g'}, \dots, L_{g*}X_k|_{g'}) \\ &= \operatorname{span}(X_1|_{gg'}, \dots, X_k|_{gg'}) \\ &= D_{gg'}, \end{aligned}$$

so D is invariant under all left translations. If M is any connected integral manifold of D, then so is $L_g(M)$, since

$$T_{g'}L_g(M) = L_{g*}(T_{g^{-1}g'}M) = L_{g*}(D_{g^{-1}g'}) = D_{g'}.$$

If M is maximal, it is easy to see that $L_g(M)$ is as well. It follows that left multiplication takes leaves to leaves: $L_g(\mathcal{H}_{g'}) = \mathcal{H}_{gg'}$ for any $g, g' \in G$.

Define $H = \mathcal{H}_e$, the leaf containing the identity. We will show that H is the desired Lie subgroup.

First, to see that H is a subgroup, observe that for any $h, h' \in H$,

$$hh' = L_h(h') \in L_h(H) = L_h(\mathcal{H}_e) = \mathcal{H}_h = H.$$

Similarly,

$$h^{-1} = h^{-1}e \in L_{h^{-1}}(\mathcal{H}_e) = L_{h^{-1}}(\mathcal{H}_h) = \mathcal{H}_{h^{-1}h} = H.$$

To show that H is a Lie group, we need to show that the map $\mu\colon (h, h') \mapsto hh'^{-1}$ is smooth as a map from $H \times H$ to H. Because $H \times H$ is a submanifold of $G \times G$, it is immediate that $\mu\colon H \times H \to G$ is smooth. Since H is an integral manifold of an involutive distribution, Proposition 14.7 shows that μ is also smooth as a map into H.

The fact that H is a leaf of $\mathcal{H}$ implies that the Lie algebra of H is $\mathfrak{h}$, because the tangent space to H at the identity is $D_e = \{X_e : X \in \mathfrak{h}\}$. To see that H is the unique connected subgroup with Lie algebra $\mathfrak{h}$, suppose $\widetilde{H}$ is any other connected subgroup with the same Lie algebra. Any such Lie subgroup is easily seen to be an integral manifold of D, so by maximality of H, we must have $\widetilde{H} \subset H$. On the other hand, if U is the domain of a flat chart for D near the identity, then by Proposition 14.6. $\widetilde{H} \cap U$ is a union of open subsets of slices. Since the slice containing e is an open subset of H, this implies that $\widetilde{H}$ contains a neighborhood V of the identity in H.

Moreover, for any other point $h \in \widetilde{H}$, $L_h(V)$ is an open subset of H that is also contained in $\widetilde{H}$, so $\widetilde{H}$ is open in H. By Problem 7-9, this implies that $\widetilde{H} = H$. □

The most important application of Theorem 15.31 is in the proof of the next theorem.

Theorem 15.32. *Suppose G and H are Lie groups with G simply connected, and let $\mathfrak{g}$ and $\mathfrak{h}$ denote their Lie algebras. For any Lie algebra homomorphism $\varphi\colon \mathfrak{g} \to \mathfrak{h}$, there is a unique Lie group homomorphism $\Phi\colon G \to H$ such that $\Phi_* = \varphi$.*

Proof. The Lie algebra of $G \times H$ is the product Lie algebra $\mathfrak{g} \times \mathfrak{h}$. Let $\mathfrak{k} \subset \mathfrak{g} \times \mathfrak{h}$ be the graph of φ:

$$\mathfrak{k} = \{(X, \varphi X) : X \in \mathfrak{g}\}.$$

Then $\mathfrak{k}$ is a vector subspace of $\mathfrak{g} \times \mathfrak{h}$ because φ is linear, and in fact it is a Lie subalgebra because φ is a homomorphism:

$$[(X, \varphi X), (X', \varphi X')] = ([X, X'], [\varphi X, \varphi X']) = ([X, X'], \varphi[X, X']) \in \mathfrak{k}.$$

Therefore, by the preceding theorem, there is a unique connected Lie subgroup $K \subset G \times H$ whose Lie algebra is $\mathfrak{k}$.

The restrictions to K of the projections

$$\pi_1|_K\colon K \to G, \qquad \pi_2|_K\colon K \to H$$

are Lie group homomorphisms because π_1 and π_2 are. Let $\Pi = \pi_1|_K\colon K \to G$. We will show that Π is a smooth covering map. Since G is simply connected, this will imply that Π is a diffeomorphism and thus a Lie group isomorphism.

To show that Π is a smooth covering map, it suffices by Problem 7-5 to show that Π is surjective and has discrete kernel. Consider the sequence of maps

$$K \hookrightarrow G \times H \xrightarrow{\pi_1} G$$

whose composition is Π. The induced Lie algebra homomorphism Π_* is just inclusion followed by projection on the algebra level:

$$\mathfrak{k} \hookrightarrow \mathfrak{g} \times \mathfrak{h} \xrightarrow{\pi_1} \mathfrak{g}.$$

This last composition is nothing more than the restriction to $\mathfrak{k}$ of the projection $\mathfrak{g} \times \mathfrak{h} \to \mathfrak{g}$. Because $\mathfrak{k} \cap \mathfrak{h} = \{0\}$ (since $\mathfrak{k}$ is a graph), it follows that $\Pi_*\colon \mathfrak{k} \to \mathfrak{g}$ is an isomorphism. Because Π is a Lie group homomorphism, it has constant rank, and therefore it is a local diffeomorphism and an open map. Moreover, its kernel is an embedded Lie subgroup of dimension zero, which is to say a discrete group. Because $\Pi(K)$ is an open subgroup of the connected Lie group G, it is all of G by Problem 7-9. Thus Π is a surjective Lie group homomorphism with discrete kernel, so it is a smooth covering map and thus a Lie group isomorphism.

Define a Lie group homomorphism $\Phi\colon G \to H$ by $\Phi = \pi_2|_K \circ \Pi^{-1}$. Note that the definition of Φ implies that

$$\pi_2|_K = \Phi \circ \pi_1|_K.$$

Because the Lie algebra homomorphism induced by the projection $\pi_1\colon G \times H \to H$ is just the linear projection $\pi_1\colon \mathfrak{g} \times \mathfrak{h} \to \mathfrak{h}$, this implies

$$\pi_2|_{\mathfrak{k}} = \Phi_* \circ \pi_1|_{\mathfrak{k}}\colon \mathfrak{k} \to \mathfrak{h}.$$

Thus if $X \in \mathfrak{g}$ is arbitrary,

$$\begin{aligned}
\varphi X &= \pi_2|_{\mathfrak{k}}(X, \varphi X) \\
&= \Phi_* \circ \pi_1|_{\mathfrak{k}}(X, \varphi X) \\
&= \Phi_* X,
\end{aligned}$$

which shows that $\Phi_* = \varphi$.

The proof is completed by showing that Φ is the unique homomorphism with this property. This is left as an exercise. □

Exercise 15.4. Show that the homomorphism Φ constructed in the preceding proof is the unique Lie group homomorphism such that $\Phi_* = \varphi$. [Hint: Consider the graph.]

Corollary 15.33. *If G and H are simply connected Lie groups with isomorphic Lie algebras, then G and H are Lie isomorphic.*

Proof. Let $\mathfrak{g}$, $\mathfrak{h}$ be the Lie algebras, and let $\varphi\colon \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra isomorphism between them. By the preceding theorem, there are Lie group homomorphisms $\Phi\colon G \to H$ and $\Psi\colon H \to G$ satisfying $\Phi_* = \varphi$ and $\Psi_* = \varphi^{-1}$. Both the identity map of G and the composition $\Psi \circ \Phi$ are maps from G to itself whose induced homomorphisms are equal to the identity, so the uniqueness part of Theorem 15.32 implies that $\Psi \circ \Phi = \mathrm{Id}$. Similarly, $\Phi \circ \Psi = \mathrm{Id}$, so Φ is a Lie group isomorphism. □

A version of this theorem was proved in the nineteenth century by Sophus Lie. However, since global topological notions such as simple connectedness had not yet been formulated, what he was able to prove was essentially a local version of this corollary. Two Lie groups G and H are said to be *locally isomorphic* if there exist neighborhoods of the identity $U \subset G$ and $V \subset H$, and a diffeomorphism $F\colon U \to V$ such that $F(g_1 g_2) = F(g_1)F(g_2)$ whenever g_1, g_2, and $g_1 g_2$ are all in U.

Theorem 15.34 (Fundamental Theorem of Sophus Lie). *Two Lie groups are locally isomorphic if and only if they have isomorphic Lie algebras.*

The proof in one direction is essentially to follow the arguments in Theorem 15.32 and Corollary 15.33, except that one just uses the inverse function theorem to show that Φ is a local isomorphism instead of appealing to the theory of covering spaces. The details are left as an exercise.

Exercise 15.5. Carry out the details of the proof of Lie's fundamental theorem.

The Fundamental Correspondence Between Lie Algebras and Lie Groups

Many of the results of this chapter show how essential properties of a Lie group are reflected in its Lie algebra. This raises a natural question: To what extent is the correspondence between Lie groups and Lie algebras (or at least between their isomorphism classes) one-to-one? We have already seen in Corollary 15.11 that isomorphic Lie groups have isomorphic Lie algebras. The converse is easily seen to be false: Both $\mathbb{R}^n$ and $\mathbb{T}^n$ have n-dimensional abelian Lie algebras, which are obviously isomorphic, but $\mathbb{R}^n$ and $\mathbb{T}^n$ are certainly not isomorphic Lie groups. However, if we restrict our attention to simply connected Lie groups, then we do obtain a one-to-one correspondence. The central result is the following theorem.

Theorem 15.35 (Lie Group–Lie Algebra Correspondence). *There is a one-to-one correspondence between isomorphism classes of finite-dimensional Lie algebras and isomorphism classes of simply connected Lie groups, given by associating each simply connected Lie group with its Lie algebra.*

The proof of this theorem will tie together all the work we have done so far on Lie groups and their Lie algebras, together with one deep algebraic result that we will state without proof. Let us begin by describing the algebraic result.

Let $\mathfrak{g}$ be a Lie algebra. A *representation* of $\mathfrak{g}$ is a Lie algebra homomorphism $\rho\colon \mathfrak{g} \to \mathfrak{gl}(n,\mathbb{R})$ for some n. If ρ is injective, it is said to be a *faithful* representation. In this case, it is easy to see that $\mathfrak{g}$ is isomorphic to $\rho(\mathfrak{g})$, which is a Lie subalgebra of $\mathfrak{gl}(n,\mathbb{R})$.

Theorem 15.36 (Ado's Theorem). *Every Lie algebra has a faithful representation.*

The proof is long, hard, and very algebraic. It can be found in [Var84].

Proof of Theorem 15.35. We need to show that the association that sends a simply connected Lie group to its Lie algebra is both surjective and injective up to isomorphism. Injectivity is precisely the content of Corollary 15.33. To prove surjectivity, suppose $\mathfrak{g}$ is any finite-dimensional Lie algebra. Replacing $\mathfrak{g}$ with its isomorphic image under a faithful representation $\rho\colon \mathfrak{g} \to \mathfrak{gl}(n,\mathbb{R})$, we may assume that $\mathfrak{g}$ is a Lie subalgebra of $\mathfrak{gl}(n,\mathbb{R})$. By Theorem 15.31, there is a unique Lie subgroup $G \subset \mathrm{GL}(n,\mathbb{R})$ that has $\mathfrak{g}$ as its Lie algebra. Letting $\widetilde{G}$ be the universal covering group of G, Proposition 15.14 shows that $\operatorname{Lie}(\widetilde{G}) \cong \operatorname{Lie}(G) \cong \mathfrak{g}$. □

What happens when we allow non-simply-connected groups? Because every Lie group has a simply connected covering group, the answer in the connected case is easy to describe. (For disconnected groups, the answer is described in Problem 15-4.)

Theorem 15.37. *Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. The connected Lie groups whose Lie algebras are isomorphic to $\mathfrak{g}$ are (up to isomorphism) precisely those of the form G/Γ, where G is the simply connected Lie group with Lie algebra $\mathfrak{g}$, and Γ is a discrete normal subgroup of G.*

Proof. Given $\mathfrak{g}$, let G be a simply connected Lie group with Lie algebra isomorphic to $\mathfrak{g}$. Suppose H is any other Lie group whose Lie algebra is isomorphic to $\mathfrak{g}$, and let $\varphi\colon \operatorname{Lie}(G) \to \operatorname{Lie}(H)$ be a Lie algebra isomorphism. Theorem 15.32 guarantees that there is a Lie group homomorphism $\Phi\colon G \to H$ such that $\Phi_* = \varphi$. Because φ is an isomorphism, Φ is a local diffeomorphism, and therefore $\Gamma = \operatorname{Ker}\Phi$ is a discrete normal subgroup of G. Since Φ is a surjective Lie homomorphism with kernel Γ, H is isomorphic to G/Γ by Problem 7-4. □

Exercise 15.6. Show that two connected Lie groups are locally isomorphic if and only if they have isomorphic universal covering groups.

Problems

15-1. Let G be a connected Lie group, and $U \subset G$ any neighborhood of the identity. Show that U generates G, i.e., that every element of G can be written as a finite product of elements of U.

15-2. Compute the exponential maps of the abelian Lie groups $\mathbb{R}^n$ and $\mathbb{T}^n$.

15-3. Consider $\mathbb{S}^3$ as the unit sphere in $\mathbb{R}^4$ with coordinates (w, x, y, z). Show that there is a Lie group structure on $\mathbb{S}^3$ in which the vector fields

$$\begin{aligned}
X_1 &= -x\frac{\partial}{\partial w} + w\frac{\partial}{\partial x} - z\frac{\partial}{\partial y} + y\frac{\partial}{\partial z},\\
X_2 &= -y\frac{\partial}{\partial w} + z\frac{\partial}{\partial x} + w\frac{\partial}{\partial y} - x\frac{\partial}{\partial z},\\
X_3 &= -z\frac{\partial}{\partial w} - y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}.
\end{aligned}$$

form a left-invariant frame. [See Problems 7-8 and 3-6. It was shown in 1958 by Raoul Bott and John Milnor [BM58] using more advanced methods that the only spheres that are parallelizable are $\mathbb{S}^1$, $\mathbb{S}^3$, and $\mathbb{S}^7$. Thus these are the only spheres that can possibly admit Lie group structures. The first two do, as we have seen; it turns out that $\mathbb{S}^7$ has no Lie group structure.]

15-4. Let $\mathfrak{g}$ be a finite-dimensional Lie algebra, and let G be the simply connected Lie group whose Lie algebra is $\mathfrak{g}$. Describe all Lie groups whose Lie algebra is $\mathfrak{g}$ in terms of G and discrete groups.

15-5. Let G be a Lie group.

(a) Show that the images of one-parameter subgroups in G are precisely the connected Lie subgroups of dimension less than or equal to 1.

(b) If $H \subset G$ is the image of a one-parameter subgroup, show that H is Lie isomorphic to one of the following: the trivial group $\{e\}$, $\mathbb{R}$, or $\mathbb{S}^1$.

15-6. Prove that there is exactly one nonabelian 2-dimensional Lie algebra up to isomorphism.

15-7. Let A and B be the following elements of $\mathfrak{gl}(2, \mathbb{R})$:

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Compute the one-parameter subgroups of $\mathrm{GL}(2, \mathbb{R})$ generated by A and B.

15-8. Let $\mathrm{GL}^+(n,\mathbb{R})$ be the subgroup of $\mathrm{GL}(n,\mathbb{R})$ consisting of matrices with positive determinant. (By Proposition 7.26, it is the identity component of $\mathrm{GL}(n,\mathbb{R})$.)

(a) Suppose $A \in \mathrm{GL}^+(n,\mathbb{R})$ is of the form e^B for some $B \in \mathfrak{gl}(n,\mathbb{R})$. Show that A has a square root, i.e., a matrix $C \in \mathrm{GL}^+(n,\mathbb{R})$ such that $C^2 = A$.

(b) Let

$$A = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}.$$

Show that the exponential map $\exp\colon \mathfrak{gl}(2,\mathbb{R}) \to \mathrm{GL}^+(2,\mathbb{R})$ is not surjective, by showing that A is not in its image.

15-9. Let $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$ denote the standard basis of $\mathbb{R}^3$, and let $\mathbb{H} = \mathbb{R} \times \mathbb{R}^3$, with basis $\{1,\mathbf{i},\mathbf{j},\mathbf{k}\}$. Define a bilinear multiplication $\mathbb{H} \times \mathbb{H} \to \mathbb{H}$ by setting

$$\begin{aligned} 1q &= q1 = q \text{ for all } q \in \mathbb{H}, \\ \mathbf{ij} &= -\mathbf{ji} = \mathbf{k}, \\ \mathbf{jk} &= -\mathbf{kj} = \mathbf{i}, \\ \mathbf{ki} &= -\mathbf{ik} = \mathbf{j}, \\ \mathbf{i}^2 &= \mathbf{j}^2 = \mathbf{k}^2 = -1, \end{aligned}$$

and extending bilinearly. With this multiplication, $\mathbb{H}$ is called the ring of *quaternions*.

(a) Show that quaternionic multiplication is associative.

(b) Show that the set $\mathcal{S}$ of unit quaternions (with respect to the Euclidean metric) is a Lie group under quaternionic multiplication, and is Lie isomorphic to $\mathrm{SU}(2)$.

(c) For any point $q \in \mathbb{H}$, show that the quaternions $\mathbf{i}q$, $\mathbf{j}q$, and $\mathbf{k}q$ are orthogonal to q. Use this to define a left-invariant frame on $\mathcal{S}$, and show that it corresponds under the isomorphism of (b) to the one defined in Problem 15-3.

15-10. Look up the Cayley numbers, and prove that $\mathbb{S}^7$ is parallelizable by mimicking as much as you can of Problem 15-9. Why do the unit Cayley numbers not form a Lie group?

15-11. Let $A \subset \mathcal{T}(\mathbb{R}^3)$ be the subspace with basis $\{X, Y, Z\}$, where

$$X = y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}, \qquad Y = z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}, \qquad Z = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}.$$

Show that A is a Lie subalgebra of $\mathfrak{T}(\mathbb{R}^3)$, which is Lie algebra isomorphic to $\mathbb{R}^3$ with the cross product, and also to the Lie algebra $\mathfrak{o}(3)$ of $\mathrm{O}(3)$.

15-12. Let G be a connected Lie group and $\mathfrak{g}$ its Lie algebra.

(a) If X, $Y \in \mathfrak{g}$, show that $[X, Y] = 0$ if and only if

$$\exp tX \exp sY = \exp sY \exp tX \text{ for all } s,\ t \in \mathbb{R}.$$

(b) Show that G is abelian if and only if $\mathfrak{g}$ is abelian.

(c) Give a counterexample when G is not connected.

15-13. Show that every connected abelian Lie group is Lie isomorphic to $\mathbb{R}^k \times \mathbb{T}^l$ for some nonnegative integers k and l.

15-14. Define a map $\beta\colon \mathrm{GL}(n,\mathbb{C}) \to \mathrm{GL}(2n,\mathbb{R})$ by identifying $(x^1 + iy^1, \dots, x^n + iy^n) \in \mathbb{C}^n$ with $(x^1, y^1, \dots, x^n, y^n) \in \mathbb{R}^{2n}$.

(a) Show that β is an injective Lie group homomorphism, so that we can identify $\mathrm{GL}(n,\mathbb{C})$ with the Lie subgroup of $\mathrm{GL}(2n,\mathbb{R})$ consisting of matrices built up out of 2×2 blocks of the form $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$.

(b) Under our usual (vector space) isomorphisms $T_I \,\mathrm{GL}(n,\mathbb{C}) \cong \mathfrak{gl}(n,\mathbb{C})$ and $T_I \,\mathrm{GL}(n,\mathbb{R}) \cong \mathfrak{gl}(n,\mathbb{R})$, show that the induced Lie algebra homomorphism $\beta_*\colon \mathrm{Lie}(\mathrm{GL}(n,\mathbb{C})) \to \mathrm{Lie}(\mathrm{GL}(n,\mathbb{R}))$ induces an injective Lie algebra homomorphism $\mathfrak{gl}(n,\mathbb{C}) \to \mathfrak{gl}(n,\mathbb{R})$ (considering both as Lie algebras with the commutator bracket). Conclude that $\mathrm{Lie}(\mathrm{GL}(n,\mathbb{C}))$ is Lie algebra isomorphic to the matrix algebra $\mathfrak{gl}(n,\mathbb{C})$.

(c) Determine the Lie algebras $\mathfrak{sl}(n,\mathbb{C})$, $\mathfrak{u}(n)$, and $\mathfrak{su}(n)$ of $\mathrm{SL}(n,\mathbb{C})$, $\mathrm{U}(n)$, and $\mathrm{SU}(n)$, respectively, as matrix subalgebras of $\mathfrak{gl}(n,\mathbb{C})$.

15-15. Show by giving an explicit isomorphism that $\mathfrak{su}(2)$ and $\mathfrak{o}(3)$ are isomorphic Lie algebras.

Appendix
Review of Prerequisites

The essential prerequisites for reading this book are a thorough acquaintance with abstract linear algebra, advanced multivariable calculus, and basic topology. The topological prerequisites include basic properties of topological spaces, topological manifolds, the fundamental group, and covering spaces; these are covered fully in my book *Introduction to Topological Manifolds* [Lee00]. In this appendix, we summarize the most important facts from linear algebra and advanced calculus that are used throughout this book.

Linear Algebra

For the basic properties of vector spaces and linear maps, you can consult almost any linear algebra book that treats vector spaces abstractly, such as [FIS97]. Here we just summarize the main points, with emphasis on those aspects that will prove most important in the study of smooth manifolds.

Vector Spaces

Let $\mathbb{R}$ denote the field of real numbers. A *vector space* over $\mathbb{R}$ (or *real vector space*) is a set V endowed with two operations: *vector addition* $V \times V \to V$, denoted by $(X, Y) \mapsto X + Y$, and *scalar multiplication* $\mathbb{R} \times V \to V$, denoted by $(a, X) \mapsto aX$; the operations are required to satisfy

(i) V is an abelian group under vector addition.

(ii) Scalar multiplication satisfies the following identities:

$$\begin{aligned} a(bX) &= (ab)X && \text{for all } X \in V \text{ and } a, b \in \mathbb{R}; \\ 1X &= X && \text{for all } X \in V. \end{aligned}$$

(iii) Scalar multiplication and vector addition are related by the following distributive laws:

$$\begin{aligned} (a+b)X &= aX + bX && \text{for all } X \in V \text{ and } a, b \in \mathbb{R}; \\ a(X+Y) &= aX + aY && \text{for all } X, Y \in V \text{ and } a \in \mathbb{R}. \end{aligned}$$

The elements of V are usually called *vectors*. When necessary to distinguish them from vectors, real numbers are sometimes called *scalars*.

This definition can be generalized in two directions. First, replacing $\mathbb{R}$ by an arbitrary field $\mathbb{F}$ everywhere, we obtain the definition of a vector space over $\mathbb{F}$. Second, if $\mathbb{R}$ is replaced by a ring $\mathcal{R}$, this becomes the definition of a *module* over $\mathcal{R}$. We will be concerned almost exclusively with real vector spaces, but it is useful to be aware of these more general definitions. Unless we specify otherwise, all vector spaces will be assumed to be real.

If V is a vector space, a subset $W \subset V$ that is closed under vector addition and scalar multiplication is itself a vector space, and is called a *subspace* of V.

Let V be a vector space. A finite sum of the form $\sum_{i=1}^{k} a^i X_i$, where $a^i \in \mathbb{R}$ and $X_i \in V$, is called a *linear combination* of the vectors $X_1, \dots, X_k$. (The reason we write the coefficients a^i with superscripts instead of subscripts is to be consistent with the Einstein summation convention, which is explained in Chapter 1.) If S is an arbitrary subset of V, the set of all linear combinations of elements of S is called the *span* of S and is denoted by $\operatorname{span}(S)$; it is easily seen to be the smallest subspace of V containing S. If $V = \operatorname{span}(S)$, we say S spans V. By convention, a linear combination of no elements is considered to sum to zero, and the span of the empty set is $\{0\}$.

Bases and Dimension

A subset S is said to be *linearly dependent* if there exists a linear relation of the form $\sum_{i=1}^{k} a^i X_i = 0$, where $X_1, \dots, X_k$ are distinct elements of S and at least one of the coefficients a^i is nonzero; it is said to be *linearly independent* otherwise. In other words, S is linearly independent if and only if the only linear combination of distinct elements of S that sums to zero is the one in which all the scalar coefficients are zero. Note that any set containing the zero vector is linearly dependent. By convention, the empty set is considered to be linearly independent.

Exercise A.1. Let V be a vector space and $S \subset V$.

(a) If S is linearly independent, show that any subset of S is linearly independent.

(b) If S is linearly dependent or spans V, show that any subset of V that properly contains S is linearly dependent.

(c) Show that S is linearly dependent if and only if some element $X \in S$ can be expressed as a linear combination of elements of $S \smallsetminus \{X\}$.

(d) If $(X_1, \dots, X_m)$ is a finite, ordered, linearly dependent subset of V, show that some X_i can be written as a linear combination of the *preceding* vectors $(X_1, \dots, X_{i-1})$.

A *basis* for V is a subset $S \subset V$ that is linearly independent and spans V. If S is a basis for V, every element of V has a *unique* expression as a linear combination of elements of S. If V has a finite basis, then V is said to be *finite-dimensional*, and otherwise it is *infinite-dimensional*. The trivial vector space $\{0\}$ (which we denote by $\mathbb{R}^0$) is finite-dimensional, because it has the empty set as a basis.

Lemma A.1. *Let V be a finite-dimensional vector space. If V is spanned by n vectors, then every subset of V containing more than n vectors is linearly dependent.*

Proof. Suppose the vectors $\{X_1, \dots, X_n\}$ span V. To prove the lemma, it clearly suffices to show that any $n+1$ vectors $\{Y_1, \dots, Y_{n+1}\}$ are dependent. Suppose not. By Exercise A.1(b), the set $\{Y_1, X_1, \dots, X_n\}$ is dependent. This means there is an equation of the form $b^1 Y_1 + \sum_i a^i X_i = 0$ in which not all of the coefficients are equal to zero. If b^1 is the only nonzero coefficient, then $Y_1 = 0$ and clearly the set of Y_is is dependent. Otherwise, some a^i is nonzero; renumbering the X_is if necessary, we may assume it is a^1. Since we can solve for X_1 in terms of Y_1 and the other X_is, the set $\{Y_1, X_2, \dots, X_n\}$ still spans V.

Now suppose by induction that $\{Y_1, Y_2, \dots, Y_{k-1}, X_k, \dots, X_n\}$ spans V. As before, the set $\{Y_1, Y_2, \dots, Y_{k-1}, Y_k, X_k, \dots, X_n\}$ is dependent, so there is a relation of the form

$$\sum_{i=1}^{k} b^i Y_i + \sum_{i=k}^{n} a^i X_i = 0$$

with not all coefficients equal to zero. If all the a^is are zero, the Y_is are clearly dependent, so we may assume at least one of the a^is is nonzero, and after reordering we may assume it is a^k. Solving for X_k as before, we conclude that the set $\{Y_1, Y_2, \dots, Y_k, X_{k+1}, \dots, X_n\}$ still spans V. Continuing by induction, we conclude that the vectors $\{Y_1, \dots, Y_n\}$ span V, which means that $\{Y_1, \dots, Y_{n+1}\}$ are dependent by Exercise A.1(b). $\square$

Proposition A.2. *If V is a finite-dimensional vector space, all bases of V contain the same number of elements.*

Proof. If $\{E_1, \ldots, E_n\}$ is a basis for V, then Lemma A.1 implies that any set containing more than n elements is dependent, so no basis can have more than n elements. Conversely, if there were a basis containing fewer than n elements, then Lemma A.1 would imply that $\{E_1, \ldots, E_n\}$ is dependent, which is a contradiction. □

Because of the preceding proposition, it makes sense to define the *dimension* of a finite-dimensional vector space to be the number of elements in a basis.

Exercise A.2. Suppose V is a finite-dimensional vector space.

(a) Show that every set that spans V contains a basis, and every linearly independent subset of V is contained in a basis.

(b) If $S \subset V$ is a subspace, show that S is finite-dimensional and $\dim S \leq \dim V$, with equality if and only if $S = V$.

An *ordered basis* of a finite-dimensional vector space is a basis endowed with a specific ordering of the basis vectors. For most purposes, ordered bases are more useful than bases, so we will assume without comment that each basis comes with a given ordering. We will denote an ordered basis by a notation such as $(E_1, \ldots, E_n)$ or (E_i).

If $(E_1, \ldots, E_n)$ is an (ordered) basis for V, each vector $X \in V$ has a unique expression as a linear combination of basis vectors:

$$X = \sum_{i=1}^{n} X^i E_i.$$

The numbers X^i are called the *components* of X with respect to this basis, and the ordered n-tuple $(X^1, \ldots, X^n)$ is called its *basis representation.* (Here is an example of a definition that requires an ordered basis.)

The fundamental example of a finite-dimensional vector space is of course $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$, which we call *$n$-dimensional Euclidean space.* It is a vector space under the usual operations of vector addition and scalar multiplication. We will denote a point in $\mathbb{R}^n$ by any of the notations x or (x^i) or $(x^1, \ldots, x^n)$; the numbers x^i are called the *coordinates* of x. They are also the components of x with respect to the *standard basis* $(e_1, \ldots, e_n)$, where $e_i = (0, \ldots, 1, \ldots, 0)$ is the vector with a 1 in the ith place and zeros elsewhere.

Notice that we always write the coordinates of a point $(x^1, \ldots, x^n) \in \mathbb{R}^n$ with upper indices, not subscripts as is usually done in linear algebra and calculus books, so as to be consistent with the Einstein summation convention (see Chapter 1).

If S and T are subspaces of a vector space V, the notation $S + T$ denotes the set of all vectors of the form $X + Y$, where $X \in S$ and $Y \in T$. It is easily seen to be a subspace, and in fact is the subspace spanned by $S \cup T$.

If $S+T=V$ and $S\cap T=\{0\}$, V is said to be the *direct sum* of S and T, and we write $V=S\oplus T$.

If S is any subspace of V, another subspace $T\subset V$ is said to be *complementary* to S if $V=S\oplus T$. In this case, it is easy to check that every vector in V has a *unique* expression as a sum of an element of S plus an element of T.

Exercise A.3. Suppose S and T are subspaces of a finite-dimensional vector space V.

(a) Show that $S\cap T$ is a subspace of V.

(b) Show that $\dim(S+T)=\dim S+\dim T-\dim(S\cap T)$.

(c) If $V=S+T$, show that $V=S\oplus T$ if and only if $\dim V=\dim S+\dim T$.

Exercise A.4. Let V be a finite-dimensional vector space. Show that every subspace $S\subset V$ has a complementary subspace in V. In fact, if $(E_1,\dots,E_n)$ is any basis for V, show that there is some subset $\{i_1,\dots,i_k\}$ of the integers $\{1,\dots,n\}$ such that $\operatorname{span}(E_{i_1},\dots,E_{i_k})$ is a complement to S. [Hint: Choose a basis $(F_1,\dots,F_m)$ for S, and apply Exercise A.1(d) to the ordered $(m+n)$-tuple $(F_1,\dots,F_m,E_1,\dots,E_n)$.]

Suppose $S\subset V$ is a subspace. For any vector $x\in V$, the *coset* of S determined by x is the set

$$x+S=\{x+y:y\in S\}.$$

A coset is also sometimes called an *affine subspace* of V parallel to S. The set V/S of cosets of S is called the *quotient* of V by S.

Exercise A.5. Suppose V is a vector space and S is a subspace of V. Define vector addition and scalar multiplication of cosets by

$$\begin{aligned}(x+S)+(y+S)&=(x+y)+S;\\ c(x+S)&=(cx)+S.\end{aligned}$$

(a) Show that the quotient V/S is a vector space under these operations.

(b) If V is finite-dimensional, show that $\dim V/S=\dim V-\dim S$.

Linear Maps

Let V and W be vector spaces. A map $T\colon V\to W$ is *linear* if $T(aX+bY)=aTX+bTY$ for all vectors $X,Y\in V$ and all scalars a,b. The *kernel* of T, denoted by $\operatorname{Ker}T$, is the set $\{X\in V:TX=0\}$, and its *image*, denoted by $\operatorname{Im}T$ or $T(V)$, is $\{Y\in W:Y=TX \text{ for some } X\in V\}$.

One simple but important example of a linear map arises in the following way. Given a subspace $S\subset V$ and a complementary subspace T, there is a unique linear map $\pi\colon V\to S$ defined by

$$\pi(X+Y)=X \text{ for } X\in S,\ Y\in T.$$

This map is called the *projection onto S with kernel T*.

A bijective linear map $T\colon V \to W$ is called an *isomorphism*. In this case, there is a unique inverse map $T^{-1}\colon W \to V$, and the following computation shows that T^{-1} is also linear:

$$\begin{aligned} aT^{-1}X + bT^{-1}Y &= T^{-1}T(aT^{-1}X + bT^{-1}Y) \\ &= T^{-1}(aTT^{-1}X + bTT^{-1}Y) \qquad \text{(by linearity of } T\text{)} \\ &= T^{-1}(aX + bY). \end{aligned}$$

For this reason, a bijective linear map is also said to be *invertible*. If there exists an isomorphism $T\colon V \to W$, then V and W are said to be *isomorphic*. Isomorphism is easily seen to be an equivalence relation.

Example A.3. Let V be any n-dimensional vector space, and let $(E_1, \dots, E_n)$ be any ordered basis for V. Define a map $E\colon \mathbb{R}^n \to V$ by

$$E(x^1, \dots, x^n) = x^1 E_1 + \dots x^n E_n.$$

Then E is bijective, so it is an isomorphism, called the *basis isomorphism* determined by this basis. Thus every n-dimensional vector space is isomorphic to $\mathbb{R}^n$.

Exercise A.6. Let V and W be vector spaces, and suppose $(E_1, \dots, E_n)$ is a basis for V. For any n elements $X_1, \dots, X_n \in W$, show that there is a unique linear map $T\colon V \to W$ satisfying $T(E_i) = X_i$ for $i = 1, \dots, n$.

Exercise A.7. Let $S\colon V \to W$ and $T\colon W \to X$ be linear maps.

(a) Show that $\operatorname{Ker} S$ and $\operatorname{Im} S$ are subspaces of V and W, respectively.

(b) Show that S is injective if and only if $\operatorname{Ker} S = \{0\}$.

(c) If S is an isomorphism, show that $\dim V = \dim W$ (in the sense that these dimensions are either both infinite or both finite and equal).

(d) If S and T are both injective or both surjective, show that $T \circ S$ has the same property.

(e) If $T \circ S$ is surjective, show that T is surjective; give an example to show that S may not be.

(f) If $T \circ S$ is injective, show that S is injective; give an example to show that T may not be.

(g) If V and W are finite-dimensional vector spaces of the same dimension and S is either injective or surjective, show that it is an isomorphism.

Exercise A.8. Suppose V is a vector space and S is a subspace of V, and let $\pi\colon V \to V/S$ denote the projection onto the quotient space. If $T\colon V \to W$ is a linear map, show that there exists a linear map $\widetilde{T}\colon V/S \to W$ such that $\widetilde{T} \circ \pi = T$ if and only if $S \subset \operatorname{Ker} T$.

Now suppose V and W are finite-dimensional vector spaces with ordered bases $(E_1, \dots, E_n)$ and $(F_1, \dots, F_m)$, respectively. If $T\colon V \to W$ is a linear map, the *matrix of* T with respect to these bases is the $m \times n$ matrix

$$A = (A_i^j) = \begin{pmatrix} A_1^1 & \dots & A_n^1 \\ \vdots & \ddots & \vdots \\ A_1^m & \dots & A_n^m \end{pmatrix}$$

whose ith column consists of the components of TE_i with respect to the basis (F_j):

$$TE_i = \sum_{j=1}^{m} A_i^j F_j.$$

By linearity, the action of T on any vector $X = \sum_i X^i E_i$ is then given by

$$T\left(\sum_{i=1}^{n} X^i E_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} A_i^j X^i F_j.$$

If we write the components of a vector with respect to a basis as a column matrix, then the matrix representation of $Y = TX$ is given by matrix multiplication:

$$\begin{pmatrix} Y^1 \\ \vdots \\ Y^m \end{pmatrix} = \begin{pmatrix} A_1^1 & \dots & A_n^1 \\ \vdots & \ddots & \vdots \\ A_1^m & \dots & A_n^m \end{pmatrix} \begin{pmatrix} X^1 \\ \vdots \\ X^n \end{pmatrix},$$

or, more succinctly,

$$Y^j = \sum_{i=1}^{n} A_i^j X^i.$$

It is straightforward to check that the composition of two linear maps is represented by the product of their matrices, and the identity tranformation of any n-dimensional vector space V is represented with respect to any basis by the $n \times n$ *identity matrix*, which we denote by I_n; it is the matrix with ones on the main diagonal and zeros elsewhere.

The set $\mathrm{M}(m \times n, \mathbb{R})$ of all $m \times n$ real matrices is easily seen to be a real vector space of dimension mn. In fact, by stringing out the matrix entries in a single row, we can identify it in a natural way with $\mathbb{R}^{mn}$. Similarly, because $\mathbb{C}$ is a real vector space of dimension 2, the set $\mathrm{M}(m \times n, \mathbb{C})$ of $m \times n$ complex matrices is a real vector space of dimension $2mn$.

Suppose A is an $n \times n$ matrix. If there is a matrix B such that $AB = BA = I_n$, then A is said to be *invertible* or *nonsingular*; it is *singular* otherwise.

Exercise A.9. Suppose A is a nonsingular matrix.

(a) Show that there is a *unique* matrix B such that $AB = BA = I_n$. This matrix is denoted by A^{-1} and is called the *inverse* of A.

(b) If A is the matrix of an invertible linear map $T\colon V \to W$ with respect to some bases for V and W, show that A is invertible and A^{-1} is the matrix of T^{-1} with respect to the same bases.

Because $\mathbb{R}^n$ comes endowed with the canonical basis (e_i), we can unambiguously identify linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ with $m \times n$ matrices, and we will often do so without further comment.

In this book, we often need to be concerned with how various objects transform when we change bases. Suppose (E_i) and $(\widetilde{E}_j)$ are two bases for a finite-dimensional vector space V. Then each basis can be written uniquely in terms of the other, so there is an invertible matrix B, called the *transition matrix* between the two bases, such that

$$\begin{aligned} E_i &= \sum_{j=1}^{n} B_i^j \widetilde{E}_j, \\ \widetilde{E}_j &= \sum_{i=1}^{n} (B^{-1})_j^i E_i. \end{aligned} \tag{A.1}$$

Now suppose V and W are finite-dimensional vector spaces and $T\colon V \to W$ is a linear map. With respect to bases (E_i) for V and (F_j) for W, T is represented by some matrix $A = (A_i^j)$. If $(\widetilde{E}_i)$ and $(\widetilde{F}_j)$ are any other choices of bases for V and W, respectively, let B and C denote the transition matrices satisfying (A.1) and

$$\begin{aligned} F_i &= \sum_{j=1}^{m} C_i^j \widetilde{F}_j, \\ \widetilde{F}_j &= \sum_{i=1}^{m} (C^{-1})_j^i F_i. \end{aligned}$$

Then a straightforward computation shows that the matrix $\widetilde{A}$ representing T with respect to the new bases is related to A by

$$\widetilde{A}_i^j = \sum_{k,l} C_l^j A_k^l (B^{-1})_i^k,$$

or, in matrix form,

$$\widetilde{A} = CAB^{-1}.$$

In particular, if T is a map from V to itself, we usually use the same basis in the domain and the range. In this case, if A denotes the matrix of T

with respect to (E_i), and $\widetilde{A}$ with respect to $(\widetilde{E}_i)$, we have

$$\widetilde{A} = BAB^{-1}.$$

If V and W are vector spaces, the set $L(V,W)$ of linear maps from V to W is a vector space under the operations

$$(S+T)X = SX + TX; \qquad (cT)X = c(TX).$$

If $\dim V = n$ and $\dim W = m$, then any choice of bases for V and W gives us a map $L(V,W) \to \mathrm{M}(m \times n, \mathbb{R})$, by sending each linear map to its matrix with respect to the chosen bases. This map is easily seen to be linear and bijective, so $\dim L(V,W) = \dim \mathrm{M}(m \times n, \mathbb{R}) = mn$.

If $T\colon V \to W$ is a linear map between finite-dimensional spaces, the dimension of $\operatorname{Im} T$ is called the *rank* of T, and the dimension of $\operatorname{Ker} T$ is called its *nullity*.

Exercise A.10. Suppose V, W, X are finite-dimensional vector spaces, and let $S\colon V \to W$ and $T\colon W \to X$ be linear maps.

(a) Show that S is injective if and only if $\operatorname{rank} S = \dim V$, and S is surjective if and only if $\operatorname{rank} S = \dim W$.

(b) Show that $\operatorname{rank}(T \circ S) \leq \operatorname{rank} S$, with equality if and only if $\operatorname{Im} S \cap \operatorname{Ker} T = \{0\}$. [Hint: Replace W by the quotient space $W/\operatorname{Ker} T$.]

(c) Show that $\operatorname{rank}(T \circ S) \leq \operatorname{rank} T$, with equality if and only if $\operatorname{Im} S + \operatorname{Ker} T = W$.

(d) If S is an isomorphism, show that $\operatorname{rank}(T \circ S) = \operatorname{rank} T$, and if T is an isomorphism, show that $\operatorname{rank}(T \circ S) = \operatorname{rank} S$.

The following theorem shows that, up to choices of bases, a linear map is completely determined by its rank together with the dimensions of its domain and range.

Theorem A.4 (Canonical Form for a Linear Map). *Suppose V and W are finite-dimensional vector spaces, and $T\colon V \to W$ is a linear map of rank r. Then there are bases for V and W with respect to which T has the following matrix representation (in block form):*

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Proof. Choose any bases $(F_1, \dots, F_r)$ for $\operatorname{Im} T$ and $(K_1, \dots, K_k)$ for $\operatorname{Ker} T$. Extend (F_j) arbitrarily to a basis $(F_1, \dots, F_m)$ for W. By definition of the image, there are vectors $E_1, \dots, E_r \in V$ such that $TE_i = F_i$ for $i = 1, \dots, r$. We will show that $(E_1, \dots, E_r, K_1, \dots, K_k)$ is a basis for V; once we know this, it follows easily that T has the desired matrix representation.

Suppose first that $\sum_i a^i E_i + \sum_j b^j K_j = 0$. Applying T to this equation yields $\sum_{i=1}^r a^i F_i = 0$, which implies that all the coefficients a^i are zero.

Then it follows also that all the b^js are zero because the K_js are independent. Therefore, the vectors $(E_1,\dots,E_r,K_1,\dots,K_k)$ are independent.

To show that they span V, let $X \in V$ be arbitrary. We can express $TX \in \operatorname{Im} T$ as a linear combination of $(F_1,\dots,F_r)$:

$$TX = \sum_{i=1}^{r} c^i F_i.$$

If we put $Y = \sum_i c^i E_i \in V$, it follows that $TY = TX$, so $Z = X - Y \in \operatorname{Ker} T$. Writing $Z = \sum_j d^j K_j$, we obtain

$$X = Y + Z = \sum_{i=1}^{r} c^i E_i + \sum_{j=1}^{k} d^j K_j,$$

so $(E_1,\dots,E_r,K_1,\dots,K_k)$ do indeed span V. □

This theorem says that any linear map can be put into a particularly nice diagonal form by appropriate choices of bases in the domain and range. However, it is important to be aware of what the theorem does *not* say: If $T\colon V \to V$ is a linear map from a finite-dimensional vector space to itself, it may not be possible to choose a *single* basis for V with respect to which the matrix of T is diagonal.

The next result is central in applications of linear algebra to smooth manifold theory; it is a corollary to the proof of the preceding theorem.

Corollary A.5 (Rank-Nullity Law). *Suppose $T\colon V \to W$ is any linear map. Then*

$$\begin{aligned}\dim V &= \operatorname{rank} T + \operatorname{nullity} T\\ &= \dim(\operatorname{Im} T) + \dim(\operatorname{Ker} T).\end{aligned}$$

Proof. The preceding proof showed that V has a basis consisting of $k + r$ elements, where $k = \dim \operatorname{Ker} T$ and $r = \dim \operatorname{Im} T$. □

Suppose A is an $m \times n$ matrix. The *transpose* of A is the $n \times m$ matrix A^T obtained by interchanging the rows and columns of A: $(A^T)^j_i = A^i_j$. (It can be interpreted abstractly as the matrix of the dual map; see Chapter 3.) The matrix A is said to be *symmetric* if $A = A^T$ and *skew-symmetric* if $A = -A^T$.

Exercise A.11. If A and B are matrices of dimensions $m \times n$ and $n \times k$, respectively, show that $(AB)^T = B^T A^T$.

The *rank* of an $m \times n$ matrix A is defined to be its rank as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Because the columns of A, thought of as vectors in $\mathbb{R}^m$, are the images of the standard basis vectors under this linear map, the rank of

A can also be thought of as the dimension of the span of its columns, and is sometimes called its *column rank*. Analogously, we define the *row rank* of A to be the dimension of the span of its rows, thought of similarly as vectors in $\mathbb{R}^n$.

Proposition A.6. *The row rank of any matrix is equal to its column rank.*

Proof. Let A be an $m \times n$ matrix. Because the row rank of A is equal to the column rank of A^T, we must show that $\operatorname{rank} A = \operatorname{rank} A^T$.

Suppose the (column) rank of A is k. Thought of as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, A factors through $\operatorname{Im} A$ as follows:

$$\begin{tikzcd} \mathbb{R}^n \arrow[rr, "A"] \arrow[dr, "\widetilde{A}"'] & & \mathbb{R}^m \\ & \operatorname{Im} A, \arrow[ur, "\iota"'] & \end{tikzcd}$$

where $\widetilde{A}$ is just the map A with its range restricted to $\operatorname{Im} A$, and ι is the inclusion of $\operatorname{Im} A$ into $\mathbb{R}^m$. Choosing a basis for the k-dimensional subspace $\operatorname{Im} A$, we can write this as a matrix equation $A = BC$, where B and C are the matrices of ι and $\widetilde{A}$ with respect to the chosen basis. Taking transposes, we find $A^T = C^T B^T$, from which it follows that $\operatorname{rank} A^T \le \operatorname{rank} B^T$. Since B^T is a $k \times m$ matrix, its column rank is at most k, which shows that $\operatorname{rank} A^T \le \operatorname{rank} A$. Reversing the roles of A and A^T and using the fact that $(A^T)^T = A$, we conclude that $\operatorname{rank} A = \operatorname{rank} A^T$. $\square$

Suppose $A = (A_i^j)$ is an $m \times n$ matrix. By choosing nonempty subsets $\{i_1, \dots, i_k\} \subset \{1, \dots, m\}$ and $\{j_1, \dots, j_l\} \subset \{1, \dots, n\}$, we obtain a $k \times l$ matrix whose entry in the pth row and qth column is $A_{j_q}^{i_p}$:

$$\begin{pmatrix} A_{j_1}^{i_1} & \dots & A_{j_l}^{i_1} \\ \vdots & \ddots & \vdots \\ A_{j_1}^{i_k} & \dots & A_{j_l}^{i_k} \end{pmatrix}.$$

Such a matrix is called a $k \times l$ *minor* of A. Looking at minors gives a convenient criterion for checking the rank of a matrix.

Proposition A.7. *Suppose A is an $m \times n$ matrix. Then* $\operatorname{rank} A \ge k$ *if and only if some $k \times k$ minor of A is nonsingular.*

Proof. We consider A as usual as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. A subspace of $\mathbb{R}^n$ or $\mathbb{R}^m$ spanned by some subset of the standard basis vectors will be called a *coordinate subspace*. Suppose $\{i_1, \dots, i_k\} \subset \{1, \dots, m\}$ and $\{j_1, \dots, j_k\} \subset \{1, \dots, n\}$, and let M denote the $k \times k$ minor of A determined by these subsets. If $P \subset \mathbb{R}^n$ is the coordinate subspace spanned by $(e_{j_1}, \dots, e_{j_k})$, and $Q \subset \mathbb{R}^m$ is the coordinate subspace spanned by

$(e_{i_1}, \dots, e_{i_k})$, it is easy to check that M is the matrix of the composite map

$$P \overset{\iota}{\hookrightarrow} \mathbb{R}^n \xrightarrow{A} \mathbb{R}^m \xrightarrow{\pi} Q,$$

where ι is inclusion of P into $\mathbb{R}^n$ and π is the coordinate projection of $\mathbb{R}^m$ onto Q. Thus to prove the proposition it suffices to show that $\operatorname{rank} A \geq k$ if and only if there are k-dimensional coordinate subspaces P and Q such that $\pi \circ A \circ \iota$ has rank k.

One direction is easy. If there are such subspaces, then $\operatorname{rank} \iota = \operatorname{rank} \pi = k$, and $k = \operatorname{rank}(\pi \circ A \circ \iota) \leq \min(k, \operatorname{rank} A)$, which implies that $\operatorname{rank} A \geq k$.

Conversely, suppose that $\operatorname{rank} A = r \geq k$. By Exercise A.4, there exists a coordinate subspace $\widetilde{P} \subset \mathbb{R}^n$ complementary to $\operatorname{Ker} A$. Since $\dim \widetilde{P} = n - \dim \operatorname{Ker} A = r \geq k$ by the rank-nullity law, we can choose a k-dimensional coordinate subspace $P \subset \widetilde{P}$. Then $A|_P$ is injective, so if $\iota \colon P \to \mathbb{R}^n$ denotes inclusion, it follows that $A \circ \iota$ is injective and has rank k.

Now let $S \subset \mathbb{R}^m$ be any coordinate subspace complementary to $\operatorname{Im}(A \circ \iota)$, and let Q be the coordinate subspace complementary to S (i.e., the span of the remaining basis elements). If $\pi \colon \mathbb{R}^m \to Q$ is the projection onto Q with kernel S, then $\operatorname{Im}(A \circ \iota)$ intersects $\operatorname{Ker} \pi = S$ trivially, so $\operatorname{rank}(\pi \circ A \circ \iota) = \operatorname{rank} A \circ \iota = k$ by Exercise A.10(b). $\square$

The Determinant

There are a number of ways of defining the determinant of a square matrix, each of which has advantages in different contexts. The definition we will give here, while perhaps not pedagogically the most straightforward, is the simplest to state and fits nicely with our treatment of alternating tensors in Chapter 9.

We let S_n denote the group of permutations of the set $\{1, \dots, n\}$, called the *symmetric group* on n elements. The properties of S_n that we will need are summarized in the following lemma; proofs can be found in any good undergraduate algebra text such as [Hun90] or [Her75]. A *transposition* is a permutation obtained by interchanging two elements and leaving all the others fixed. A permutation that can be written as a composition of an even number of transpositions is called *even*, and one that can be written as a composition of an odd number of transpositions is called *odd*.

Lemma A.8 (Properties of the Symmetric Group).

(*a*) *Every element of S_n can be decomposed as a finite sequence of transpositions.*

(*b*) *For any $\sigma \in S_n$, the parity (evenness or oddness) of the number of transpositions in any decomposition of σ as a sequence of transpositions is independent of the choice of decomposition.*

(*c*) *The map* $\operatorname{sgn}\colon S_n \to \{\pm 1\}$ *given by*

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even,} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}$$

is a surjective group homomorphism, where we consider $\{\pm 1\}$ *as a group under multiplication.*

Exercise A.12. Prove (or look up) Lemma A.8.

If $A = (A_i^j)$ is an $n \times n$ (real or complex) matrix, the *determinant* of A is defined by the expression

$$\det A = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) A_1^{\sigma 1} \cdots A_n^{\sigma n}. \tag{A.2}$$

For simplicity, we assume throughout this section that our matrices are real. The statements and proofs, however, hold equally well in the complex case. In our study of Lie groups we will also have occasion to consider determinants of complex matrices.

Although the determinant is defined as a function of matrices, it is also useful to think of it as a function of n vectors in $\mathbb{R}^n$: If $A_1, \dots, A_n \in \mathbb{R}^n$, we interpret $\det(A_1, \dots, A_n)$ to mean the determinant of the matrix whose columns are $(A_1, \dots, A_n)$:

$$\det(A_1, \dots, A_n) = \det \begin{pmatrix} A_1^1 & \dots & A_n^1 \\ \vdots & \ddots & \vdots \\ A_1^n & \dots & A_n^n \end{pmatrix}.$$

It is obvious from the defining formula (A.2) that the function $\det\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ so defined is *multilinear*, which means that it is linear as a function of each vector when all the other vectors are held fixed.

Proposition A.9 (Properties of the Determinant). *Let A be an $n \times n$ matrix.*

(*a*) *If one column of A is multiplied by a scalar c, the determinant is multiplied by the same scalar:*

$$\det(A_1, \dots, cA_i, \dots, A_n) = c \det(A_1, \dots, A_i, \dots, A_n).$$

(*b*) *The determinant changes sign when two columns are interchanged:*

$$\begin{aligned} &\det(A_1, \dots, A_q, \dots, A_p, \dots, A_n) \\ &\qquad = -\det(A_1, \dots, A_p, \dots, A_q, \dots, A_n). \end{aligned} \tag{A.3}$$

(c) *The determinant is unchanged by adding a scalar multiple of one column to any other column:*

$$\det(A_1, \dots, A_i, \dots, A_j + cA_i, \dots, A_n) \\ = \det(A_1, \dots, A_i, \dots, A_j \dots, A_n).$$

(d) *For any scalar c,* $\det(cA) = c^n \det A$*.*

(e) *If any two columns of A are identical, then* $\det A = 0$*.*

(f) $\det A^T = \det A$*.*

(g) *If* $\operatorname{rank} A < n$*, then* $\det A = 0$*.*

Proof. Part (a) is part of the definition of multilinearity, and (d) follows immediately from (a). To prove (b), suppose $p < q$ and let $\tau \in S_n$ be the transposition that interchanges p and q, leaving all other indices fixed. Then the left-hand side of (A.3) is equal to

$$\begin{aligned}
\det(A_1, \dots, A_q, \dots, A_p, \dots, A_n) &= \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) A_1^{\sigma 1} \cdots A_q^{\sigma p} \cdots A_p^{\sigma q} \cdots A_n^{\sigma n} \\
&= \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) A_1^{\sigma 1} \cdots A_p^{\sigma q} \cdots A_q^{\sigma p} \cdots A_n^{\sigma n} \\
&= \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) A_1^{\sigma \tau 1} \cdots A_n^{\sigma \tau n} \\
&= - \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma \tau)) A_1^{\sigma \tau 1} \cdots A_n^{\sigma \tau n} \\
&= - \sum_{\eta \in S_n} (\operatorname{sgn} \eta) A_1^{\eta 1} \cdots A_n^{\eta n} \\
&= - \det(A_1, \dots, A_n),
\end{aligned}$$

where the next-to-last line follows by substituting $\eta = \sigma\tau$ and noting that η runs over all elements of S_n as σ does. Part (e) is then an immediate consequence of (b), and (c) follows by multilinearity:

$$\begin{aligned}
&\det(A_1, \dots, A_i, \dots, A_j + cA_i, \dots, A_n) \\
&\quad = \det(A_1, \dots, A_i, \dots, A_j \dots, A_n) + c \det(A_1, \dots, A_i, \dots, A_i \dots, A_n) \\
&\quad = \det(A_1, \dots, A_i, \dots, A_j \dots, A_n) + 0.
\end{aligned}$$

Part (f) follows directly from the definition of the determinant:

$$\begin{aligned}
\det A^T &= \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) A^1_{\sigma 1}\cdots A^n_{\sigma n}\\
&= \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) A^{\sigma^{-1}\sigma 1}_{\sigma 1}\cdots A^{\sigma^{-1}\sigma n}_{\sigma n}\\
&= \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) A^{\sigma^{-1}1}_{1}\cdots A^{\sigma^{-1}n}_{n} \qquad \text{(multiplication is commutative)}\\
&= \sum_{\eta\in S_n} (\operatorname{sgn}\eta) A^{\eta 1}_{1}\cdots A^{\eta n}_{n} \qquad \text{(substituting } \eta=\sigma^{-1})\\
&= \det A.
\end{aligned}$$

Finally, to prove (g), suppose $\operatorname{rank} A < n$. Then the columns of A are dependent, so at least one column can be written as a linear combination of the others: $A_j = \sum_{i\neq j} c^i A_i$. The result then follows from the multilinearity of det and (e). □

The operations on matrices described in parts (a), (b), and (c) of the preceding proposition (multiplying one column by a scalar, interchanging two columns, and adding a multiple of one column to another) are called *elementary column operations*. Part of the proposition, therefore, describes precisely how a determinant is affected by elementary column operations. If we define *elementary row operations* analogously, the fact that the determinant of A^T is equal to that of A implies that the determinant behaves similarly under elementary row operations.

Each elementary column operation on a matrix A can be realized by multiplying A on the right by a suitable matrix, called an *elementary matrix*. For example, multiplying the ith column by c is achieved by multiplying A by the matrix E_c that is equal to the identity matrix except for a c in the (i,i) position:

$$\begin{pmatrix}
A^1_1 & \dots & A^1_i & \dots & A^1_n\\
\vdots & & \vdots & & \vdots\\
A^j_1 & \dots & A^j_i & \dots & A^j_n\\
\vdots & & \vdots & & \vdots\\
A^n_1 & \dots & A^n_i & \dots & A^n_n
\end{pmatrix}
\begin{pmatrix}
1 & \dots & 0 & \dots & 0\\
 & \ddots & & & \\
 & & c & & \\
 & & & \ddots & \\
0 & \dots & 0 & \dots & 1
\end{pmatrix}
= \begin{pmatrix}
A^1_1 & \dots & cA^1_i & \dots & A^1_n\\
\vdots & & \vdots & & \vdots\\
A^j_1 & \dots & cA^j_i & \dots & A^j_n\\
\vdots & & \vdots & & \vdots\\
A^n_1 & \dots & cA^n_i & \dots & A^n_n
\end{pmatrix}.$$

Observe that $\det E_c = c$.

> **Exercise A.13.** Show that interchanging two columns of a matrix is equivalent to multiplying on the right by a matrix whose determinant is -1, and adding a multiple of one column to another is equivalent to multiplying on the right by a matrix of determinant 1.

> **Exercise A.14.** Suppose A is a nonsingular $n \times n$ matrix.
>
> (a) Show that A can be reduced to the identity I_n by a sequence of elementary column operations.
>
> (b) Show that A is equal to a product of elementary matrices.

Elementary matrices form a key ingredient in the proof of the following theorem, which is arguably the deepest and most important property of the determinant.

Theorem A.10. *If A and B are $n \times n$ matrices, then*

$$\det(AB) = (\det A)(\det B).$$

Proof. If B is singular, then $\operatorname{rank} B < n$, which implies that $\operatorname{rank} AB < n$. Therefore both $\det B$ and $\det AB$ are zero by Proposition A.9(g). On the other hand, parts (a), (b), and (c) of Proposition A.9 combined with Exercise A.13 show that the theorem is true when B is an elementary matrix. If B is an arbitrary nonsingular matrix, then B can be written as a product of elementary matrices by Exercise A.14, and then the result follows by induction. □

Corollary A.11. *If A is a nonsingular $n \times n$ matrix, then* $\det(A^{-1}) = (\det A)^{-1}$.

Proof. Just note that $1 = \det I_n = \det(AA^{-1}) = (\det A)(\det A^{-1})$. □

Corollary A.12. *A square matrix is nonsingular if and only if its determinant is nonzero.*

Proof. One direction follows from Proposition A.9(g); the other from Corollary A.11. □

For actual computations of determinants, the formula in the following proposition is usually more useful than the definition.

Proposition A.13 (Expansion by Minors). *Let A be an $n \times n$ matrix, and for each i, j let M_i^j denote the $(n-1)\times(n-1)$ minor obtained by deleting the ith column and jth row of A. For any fixed i between 1 and n inclusive,*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} A_i^j \det M_i^j. \tag{A.4}$$

Proof. It is useful to consider first a special case: Suppose A is an $n \times n$ matrix that has the block form

$$A = \begin{pmatrix} B & 0 \\ C & 1 \end{pmatrix}, \tag{A.5}$$

where B is an $(n-1) \times (n-1)$ matrix and C is a $1 \times n$ row matrix. Then in the defining formula (A.2) for $\det A$, the factor $A_n^{\sigma n}$ is equal to 1 when $\sigma n = n$ and zero otherwise, so in fact the only terms that are nonzero are those in which $\sigma \in S_{n-1}$, thought of as the subgroup of S_n consisting of elements that permute $\{1, \dots, n-1\}$ and leave n fixed. Thus the determinant of A simplifies to

$$\det A = \sum_{\sigma \in S_{n-1}} (\operatorname{sgn} \sigma) A_1^{\sigma 1} \cdots A_{n-1}^{\sigma(n-1)} = \det B.$$

Now let A be arbitrary, and fix i between 1 and n. For each $j = 1, \dots, n$, let X_i^j denote the matrix obtained by replacing the ith column of A by the basis vector e_j. Since the determinant is a multilinear function of its columns,

$$\begin{aligned}
\det A &= \det \left(A_1, \dots, A_{i-1}, \sum_{j=1}^n A_i^j e_j, A_{i+1}, \dots, A_n \right) \\
&= \sum_{j=1}^n A_i^j \det(A_1, \dots, A_{i-1}, e_j, A_{i+1}, \dots, A_n) \qquad\qquad (A.6) \\
&= \sum_{j=1}^n A_i^j \det X_i^j.
\end{aligned}$$

On the other hand, by interchanging columns $n - i$ times and then interchanging rows $n - j$ times, we can transform X_i^j to a matrix of the form (A.5) with $B = M_i^j$. Therefore, by the observation in the preceding paragraph,

$$\det X_i^j = (-1)^{n-i+n-j} \det M_i^j = (-1)^{i+j} \det M_i^j.$$

Inserting this into (A.6) completes the proof. □

Formula (A.4) is called the *expansion of* $\det A$ *by minors along the* i*th column.* Since $\det A = \det A^T$, there is an analogous expansion along any row. The factor $(-1)^{i+j} \det M_i^j$ multiplying A_i^j in (A.4) is called the *cofactor* of A_i^j, and is denoted by cof_i^j.

Proposition A.14 (Cramer's Rule). *Let A be a nonsingular $n \times n$ matrix. Then A^{-1} is equal to $1/(\det A)$ times the transposed cofactor matrix of A:*

$$(A^{-1})_j^i = \frac{1}{\det A} \operatorname{cof}_i^j = \frac{1}{\det A} (-1)^{i+j} \det M_i^j. \tag{A.7}$$

Proof. Let B^i_j denote the expression on the right-hand side of (A.7). Then

$$\sum_{j=1}^{n} B^i_j A^j_k = \frac{1}{\det A} \sum_{j=1}^{n} (-1)^{i+j} A^j_k \det M^j_i. \tag{A.8}$$

When $k = i$, the summation on the right-hand side is precisely the expansion of $\det A$ by minors along the ith column, so the right-hand side of (A.8) is equal to 1. On the other hand, if $k \neq i$, the summation is equal to the determinant of the matrix obtained by replacing the ith column of A by the kth column. Since this matrix has two identical columns, its determinant is zero. Thus (A.8) is equivalent to the matrix equation $BA = I_n$, where B is the matrix (B^j_i). Multiplying both sides on the right by A^{-1}, we conclude that $B = A^{-1}$. □

A square matrix $A = (A^j_i)$ is said to be *upper triangular* if $A^j_i = 0$ for $j > i$ (i.e., the only nonzero entries are on and above the main diagonal). Determinants of upper triangular matrices are particularly easy to compute.

Proposition A.15 (Determinant of an Upper Triangular Matrix). *If A is an upper triangular $n \times n$ matrix, then the determinant of A is the product of its diagonal entries:*

$$\det A = A^1_1 \cdots A^n_n.$$

Proof. When $n = 1$, this is trivial. So assume the result is true for $(n-1) \times (n-1)$ matrices, and let A be an upper triangular $n \times n$ matrix. In the expansion of $\det A$ by minors along the first column, there is only one nonzero entry, namely $A^1_1 \det M^1_1$. By induction $\det M^1_1 = A^2_2 \cdots A^n_n$, which proves the proposition. □

Suppose X is an $(m+k) \times (m+k)$ matrix. We say X is *block upper triangular* if X has the form

$$X = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \tag{A.9}$$

for some matrices A, B, C of sizes $m \times m$, $m \times k$, and $k \times k$, respectively.

Proposition A.16. *If X is the block upper triangular matrix given by* (A.9)*, then* $\det X = (\det A)(\det C)$.

Proof. If A is singular, then clearly the columns of X are linearly dependent, which implies that $\det X = 0 = (\det A)(\det C)$. So let us assume that A is nonsingular.

Consider first the following special case:

$$X = \begin{pmatrix} I_m & 0 \\ 0 & C \end{pmatrix}.$$

Expanding by minors along the first column, an easy induction shows that $\det X = \det C$ in this case. A similar argument shows that

$$\det\begin{pmatrix} A & 0 \\ 0 & I_k \end{pmatrix} = \det A.$$

The general case follows from these two observations together with the factorization

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I_k \end{pmatrix}\begin{pmatrix} I_m & 0 \\ 0 & C \end{pmatrix}\begin{pmatrix} I_m & A^{-1}B \\ 0 & I_k \end{pmatrix},$$

noting that the last matrix above is upper triangular with ones along the main diagonal. □

Inner Products and Norms

If V is a real vector space, an *inner product* on V is a map from $V \times V \to V$, usually written $(X, Y) \mapsto \langle X, Y\rangle$, that is

(i) Symmetric:

$$\langle X, Y\rangle = \langle Y, X\rangle.$$

(ii) Bilinear:

$$\langle aX + a'X', Y\rangle = a\langle X, Y\rangle + a'\langle X', Y\rangle,$$
$$\langle X, bY + b'Y'\rangle = b\langle X, Y\rangle + b'\langle X, Y'\rangle.$$

(iii) Positive definite:

$$\langle X, X\rangle > 0 \text{ unless } X = 0.$$

A vector space endowed with a specific inner product is called an *inner product space*. The standard example is, of course, $\mathbb{R}^n$ with its *dot product* or *Euclidean inner product*:

$$\langle x, y\rangle = x \cdot y = \sum_{i=1}^{n} x^i y^i.$$

Suppose V is an inner product space. For any $X \in V$, the *length* of X is the number $|X| = \sqrt{\langle X, X\rangle}$. A *unit vector* is a vector of length 1. The *angle* between two nonzero vectors $X, Y \in V$ is defined to be the unique $\theta \in [0, \pi]$ satisfying

$$\cos\theta = \frac{\langle X, Y\rangle}{|X|\,|Y|}.$$

Two vectors $X, Y \in V$ are said to be *orthogonal* if the angle between them is $\pi/2$, or equivalently if $\langle X, Y\rangle = 0$.

Exercise A.15. Let V be an inner product space. Show that the length function associated with the inner product satisfies

$$\begin{aligned} |X| &> 0, && X \in V,\ X \neq 0; \\ |cX| &= |c|\,|X|, && c \in \mathbb{R},\ X \in V; \\ |X+Y| &\le |X| + |Y|, && X, Y \in V. \end{aligned}$$

Suppose V is a finite-dimensional inner product space. A basis $(E_1, \dots, E_n)$ of V is said to be *orthonormal* if each E_i is a unit vector and E_i is orthogonal to E_j when $i \neq j$.

Proposition A.17 (The Gram-Schmidt Algorithm). *Every finite-dimensional inner product space V has an orthonormal basis. In fact, if $(E_1, \dots, E_n)$ is any basis of V, there is an orthonormal basis $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ with the property that*

$$\operatorname{span}(E_1, \dots, E_k) = \operatorname{span}(\widetilde{E}_1, \dots, \widetilde{E}_k) \textit{ for } k = 1, \dots, n. \tag{A.10}$$

Proof. The proof is by induction on $n = \dim V$. If $n = 1$, there is only one basis element E_1, and then $\widetilde{E}_1 = E_1/|E_1|$ is an orthonormal basis.

Suppose the result is true for inner product spaces of dimension $n-1$, and let V have dimension n. Then $W = \operatorname{span}(E_1, \dots, E_{n-1})$ is an $(n-1)$-dimensional inner product space with the inner product restricted from V, so there is an orthonormal basis $(\widetilde{E}_1, \dots, \widetilde{E}_{n-1})$ satisfying (A.10) for $k = 1, \dots, n-1$. Define $\widetilde{E}_n$ by

$$\widetilde{E}_n = \frac{E_n - \sum_{i=1}^{n-1} \langle E_n, \widetilde{E}_i \rangle \widetilde{E}_i}{\left| E_n - \sum_{i=1}^{n-1} \langle E_n, \widetilde{E}_i \rangle \widetilde{E}_i \right|}.$$

A computation shows that $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ is the desired orthonormal basis. □

An isomorphism $T \colon V \to W$ between inner product spaces is called an *isometry* if it takes the inner product of V to that of W:

$$\langle TX, TY \rangle = \langle X, Y \rangle.$$

Exercise A.16. Show that any isometry is a homeomorphism that preserves lengths, angles, and orthogonality, and takes orthonormal bases to orthonormal bases.

Exercise A.17. If (E_i) is any basis for the finite-dimensional vector space V, show that there is a unique inner product on V for which (E_i) is orthonormal.

Exercise A.18. Suppose V is a finite-dimensional inner product space and $E \colon \mathbb{R}^n \to V$ is the basis map determined by any orthonormal basis. Show that E is an isometry, where $\mathbb{R}^n$ is endowed with the Euclidean inner product.

The preceding exercise shows that finite-dimensional inner product spaces are topologically and geometrically indistinguishable from the Euclidean space of the same dimension. Thus any such space automatically inherits all the usual properties of Euclidean space, such as compactness of closed and bounded subsets.

If V is a finite-dimensional inner product space and $S \subset V$ is a subspace, the *orthogonal complement* of S in V is the set

$$V^{\perp} = \{X \in V : \langle X, Y \rangle = 0 \text{ for all } Y \in S.\}.$$

Exercise A.19. Let V be a finite-dimensional inner product space and $S \subset V$ any subspace. Show that $V = S \oplus S^{\perp}$.

Thanks to the result of the preceding exercise, for any subspace S of an inner product space V, there is a natural projection $\pi \colon V \to S$ with kernel $S^{\perp}$. This is called the *orthogonal projection* of V onto S.

A *norm* on a vector space V is a function from V to $\mathbb{R}$, written $X \mapsto |X|$, satisfying

(i) Positivity: $|X| \geq 0$ for every $X \in V$, and $|X| = 0$ if and only if $X = 0$.

(ii) Homogeneity: $|cX| = |c|\,|X|$ for every $c \in \mathbb{R}$ and $X \in V$.

(iii) Triangle inequality: $|X + Y| \leq |X| + |Y|$ for all $X, Y \in V$.

A vector space together with a specific choice of norm is called a *normed linear space.* Exercise A.15 shows that the length function associated with any inner product is a norm.

If V is a normed linear space, then for any $p \in V$ and any $r > 0$ we define the *open ball* and *closed ball* of radius r around p, denoted by $B_r(p)$ and $\overline{B}_r(p)$, respectively, by

$$B_r(p) = \{x \in V : |x - p| < r\},$$
$$\overline{B}_r(p) = \{x \in V : |x - p| \leq r\}.$$

Given a norm on V, the function $d(X, Y) = |X - Y|$ is a metric, yielding a topology on V called the *norm topology.* The set of all open balls is easily seen to be a basis for this topology. Two norms $|\cdot|_1$ and $|\cdot|_2$ are said to be *equivalent* if there are positive constants c, C such that

$$c|X|_1 \leq |X|_2 \leq C|X|_1 \text{ for all } X \in V.$$

Exercise A.20. Show that equivalent norms on a vector space V determine the same topology.

Exercise A.21. Show that any two norms on a finite-dimensional vector space are equivalent. [Hint: first choose an inner product on V, and show that the unit ball in any norm is compact with respect to the topology determined by the inner product.]

If V and W are normed linear spaces, a linear map $T\colon V \to W$ is said to be *bounded* if there exists a positive constant C such that

$$|TX| \le C|X| \text{ for all } X \in V.$$

Exercise A.22. Show that a linear map between normed linear spaces is bounded if and only if it is continuous.

Exercise A.23. Show that every linear map between finite-dimensional normed linear spaces is continuous.

The vector space $\mathrm{M}(m \times n, \mathbb{R})$ of $m \times n$ real matrices has a natural Euclidean inner product, obtained by identifying a matrix with a point in $\mathbb{R}^{mn}$:

$$A \cdot B = \sum_{i,j} A^i_j B^i_j.$$

This yields a Euclidean norm on matrices:

$$|A| = \sqrt{\textstyle\sum_{i,j} (A^j_i)^2}. \tag{A.11}$$

Whenever we use a norm on a space of matrices, it will always be assumed to be this Euclidean norm.

Exercise A.24. For any matrices $A \in \mathrm{M}(m \times n, \mathbb{R})$ and $B \in \mathrm{M}(n \times k, \mathbb{R})$, show that

$$|AB| \le |A|\,|B|.$$

Calculus

In this section, we summarize the main results from multivariable calculus and real analysis that are needed in this book. For details on most of the ideas touched on here, you can consult [Rud76] or [Apo74].

If $U \subset \mathbb{R}^n$ is any subset and $F\colon U \to \mathbb{R}^m$ is any map, we write the components of $F(x)$ as $F(x) = (F^1(x), \dots, F^m(x))$; this defines n functions $F^1, \dots, F^n\colon U \to \mathbb{R}$ called the *component functions* of F. Note that $F^i = \pi^i \circ F$, where $\pi^i\colon \mathbb{R}^m \to \mathbb{R}$ is the projection on the ith coordinate:

$$\pi^i(x^1, \dots, x^n) = x^i.$$

For maps between Euclidean spaces, there are two separate but closely related types of derivatives: partial derivatives and total derivatives. We begin with partial derivatives.

Partial Derivatives

Suppose $U \subset \mathbb{R}^n$ is open and $f\colon U \to \mathbb{R}$ is any real-valued function. For any $a = (a^1, \dots, a^n) \in U$ and any $j = 1, \dots, n$, the jth *partial derivative* of f at a is defined by differentiating with respect to x^j and holding the other variables fixed:

$$\frac{\partial f}{\partial x^j}(a) = \lim_{h \to 0} \frac{f(a^1, \dots, a^j + h, \dots, a^n)}{h} = \lim_{h \to 0} \frac{f(a + he_j)}{h},$$

if the limit exists. For a vector-valued function $F\colon U \to \mathbb{R}^m$, the partial derivatives of F are defined simply to be the partial derivatives $\partial F^i/\partial x^j$ of the component functions of F. The matrix $(\partial F^i/\partial x^j)$ of partial derivatives is called the *Jacobian matrix* of F.

If $F\colon U \to \mathbb{R}^m$ is a map for which each partial derivative exists at each point in U and the functions $\partial F^i/\partial x^j : U \to \mathbb{R}$ so defined are all continuous, then F is said to be of *class* C^1 or *continuously differentiable*. If this is the case, we can differentiate the functions $\partial F^i/\partial x^j$ to obtain *second-order partial derivatives*

$$\frac{\partial^2 F^i}{\partial x^k \partial x^j} = \frac{\partial}{\partial x^k}\left(\frac{\partial F^i}{\partial x^j}\right),$$

if they exist. Continuing this way leads to higher-order partial derivatives—the partial derivatives of F of order k are the partial derivatives of those of order $k-1$, when they exist.

In general, for $k \geq 0$, a function $F\colon U \to \mathbb{R}^m$ is said to be of *class* C^k or k *times continuously differentiable* if all the partial derivatives of F of order less than or equal to k exist and are continuous functions on U. (Thus a function of class C^0 is just a continuous function.) A function that is of class C^k for every $k \geq 0$ is said to be of *class* C^∞, *smooth*, or *infinitely differentiable*. Because existence and continuity of derivatives are local properties, clearly F is C^1 (or C^k or smooth) if and only if it has that property in a neighborhood of each point in U.

We will often be most concerned with real-valued functions, that is, functions whose range is $\mathbb{R}$. If $U \subset \mathbb{R}^n$ is open, the set of all real-valued functions of class C^k on U is denoted by $C^k(U)$, and the set of all smooth real-valued functions by $C^\infty(U)$. By virtue of the following exercise, $C^\infty(U)$ is a vector space under pointwise addition and multiplication by constants:

$$\begin{aligned}(f+g)(x) &= f(x) + g(x)\\ (cf)(x) &= c(f(x)).\end{aligned}$$

In fact, it is also a ring, with multiplication defined pointwise:

$$(fg)(x) = f(x)g(x).$$

Exercise A.25. Let $U \subset \mathbb{R}^n$ be an open set, and suppose $f, g\colon U \to \mathbb{R}^n$ are smooth.

(a) Show that $f+g$ is smooth.

(b) Show that fg is smooth.

(c) If g never vanishes on U, show that f/g is smooth.

The following important result shows that, for most interesting maps, the order in which we take partial derivatives is irrelevant. For a proof, see [Rud76].

Proposition A.18 (Equality of Mixed Partial Derivatives). *If U is an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ is a map of class C^2, then the mixed second-order partial derivatives of F do not depend on the order of differentiation:*

$$\frac{\partial^2 F^i}{\partial x^j \partial x^k} = \frac{\partial^2 F^i}{\partial x^k \partial x^j}.$$

Corollary A.19. *If $F\colon U \to \mathbb{R}^m$ is smooth, then mixed partial derivatives of any order are independent of the order of differentiation.*

Another important property of smooth functions is that integrals of smooth functions can be differentiated under the integral sign. A precise statement is given in the next theorem; this is not the best that can be proved, but it is more than sufficient for our purposes. For a proof, see [Rud76].

Theorem A.20 (Differentiation Under an Integral Sign). *Let $U \subset \mathbb{R}^n$ be an open set, $a, b \in \mathbb{R}$, and let $f\colon U \times [a,b] \to \mathbb{R}$ be a continuous function such that the partial derivative $\partial f/\partial t\colon U \times [a,b] \to \mathbb{R}$ is also continuous. Define $F\colon U \to \mathbb{R}$ by*

$$F(x) = \int_a^b f(x,t)\, dt.$$

Then F is of class C^1, and its partial derivatives can be computed by differentiating under the integral sign:

$$\frac{\partial F}{\partial x^i}(x) = \int_a^b \frac{\partial f}{\partial x^i}(x,t)\, dt.$$

Theorem A.21 (Taylor's Formula with Remainder). *Let $U \subset \mathbb{R}^n$ be a convex open set, and suppose f is a smooth real-valued function on U. For any integer $m \geq 0$, any $a \in U$, and all $v \in \mathbb{R}^n$ small enough that*

$a + v \in U$,

$$f(a+v) = \sum_{k=0}^{m} \sum_{i_1,\ldots,i_k} \frac{1}{k!} \frac{\partial^k f}{\partial x^{i_1} \cdots \partial x^{i_k}}(a) v^{i_1} \cdots v^{i_k}$$
$$+ \sum_{i_1,\ldots,i_{m+1}} \int_0^1 \frac{1}{m!}(1-t)^m \frac{\partial^{m+1} f}{\partial x^{i_1} \cdots \partial x^{i_{m+1}}}(a+tv) v^{i_1} \cdots v^{i_{m+1}}\, dt, \qquad \text{(A.12)}$$

where each index i_j runs from 1 to n.

Proof. The proof is by induction on m. When $m = 0$, it follows from the fundamental theorem of calculus and the chain rule:

$$f(a+v) - f(a) = \int_0^1 \frac{\partial}{\partial t} f(a+tv)\, dt = \sum_{i=1}^{n} \int_0^1 \frac{\partial f}{\partial x^i}(a+tv) v^i\, dt.$$

So suppose the formula holds for some $m \geq 0$. To prove it for $m+1$, we integrate by parts in (A.12), with

$$u = \sum_{i_1,\ldots,i_{m+1}} \frac{\partial^{m+1} f}{\partial x^{i_1} \cdots \partial x^{i_{m+1}}}(a+tv) v^{i_1} \cdots v^{i_{m+1}},$$
$$du = \sum_{i_1,\ldots,i_{m+2}} \frac{\partial^{m+2} f}{\partial x^{i_1} \cdots \partial x^{i_{m+2}}}(a+tv) v^{i_1} \cdots v^{i_{m+2}}\, dt,$$
$$v = -\frac{1}{(m+1)!}(1-t)^{m+1},$$
$$dv = \frac{1}{m!}(1-t)^m\, dt,$$

to obtain

$$\sum_{i_1,\ldots,i_{m+1}} \int_0^1 \frac{1}{m!}(1-t)^m \frac{\partial^{m+1} f}{\partial x^{i_1} \cdots \partial x^{i_{m+1}}}(a+tv) v^{i_1} \cdots v^{i_{m+1}}\, dt$$
$$= \left[-\frac{1}{(m+1)!}(1-t)^{m+1} \sum_{i_1,\ldots,i_{m+1}} \frac{\partial^{m+1} f}{\partial x^{i_1} \cdots \partial x^{i_{m+1}}}(a+tv) v^{i_1} \cdots v^{i_{m+1}} \right]_{t=0}^{t=1}$$
$$+ \int_0^1 \frac{1}{(m+1)!}(1-t)^{m+1} \sum_{i_1,\ldots,i_{m+2}} \frac{\partial^{m+2} f}{\partial x^{i_1} \cdots \partial x^{i_{m+2}}}(a+tv) v^{i_1} \cdots v^{i_{m+2}} dt$$
$$= \sum_{i_1,\ldots,i_{m+1}} \frac{1}{(m+1)!} \frac{\partial^{m+1} f}{\partial x^{i_1} \cdots \partial x^{i_{m+1}}}(a) v^{i_1} \cdots v^{i_{m+1}}$$
$$+ \sum_{i_1,\ldots,i_{m+2}} \int_0^1 \frac{1}{(m+1)!}(1-t)^{m+1} \frac{\partial^{m+2} f}{\partial x^{i_1} \cdots \partial x^{i_{m+2}}}(a+tv) v^{i_1} \cdots v^{i_{m+2}} dt.$$

Inserting this into (A.12) completes the proof. $\square$

Corollary A.22 (First-Order Taylor Formula). *With f, a, and v as above,*

$$f(a+v) = f(a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(a)v^i + \sum_{i=1}^{n} g_i(v)v^i,$$

for some smooth functions $g_1, \dots, g_n$ defined on U.

The notation $O(|x|^k)$ is used to denote any function $G(x)$ defined on a neighborhood of the origin in $\mathbb{R}^n$ which satisfies $|G(x)| \le C|x|^k$ for some constant C and all sufficiently small x.

Corollary A.23 (Second-Order Taylor Formula). *With f, a, and v as above,*

$$f(a+v) = f(a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(a)v^i + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x^i \partial x^j}(a)v^i v^j + O(|v|^3).$$

Exercise A.26. Prove the previous two corollaries.

We will sometimes need to consider smooth maps on subsets of $\mathbb{R}^n$ that are not open. If $A \subset \mathbb{R}^n$ is *any* subset, a map $F\colon A \to \mathbb{R}^m$ is said to be smooth if it extends to a smooth map $U \to \mathbb{R}^k$ on some open neighborhood U of A.

The Total Derivative

For maps between (open subsets of) finite-dimensional vector spaces, there is another very important notion of derivative, called the total derivative.

Let V, W be finite-dimensional vector spaces, which we may assume to be endowed with norms. If $U \subset V$ is an open set, a map $F\colon U \to W$ is said to be *differentiable* at $a \in U$ if there exists a linear map $L\colon V \to W$ such that

$$\lim_{v \to 0} \frac{F(a+v) - F(a) - Lv}{|v|} = 0. \tag{A.13}$$

Because all norms on a finite-dimensional vector space are equivalent, this definition is independent of the choices of norms on V and W.

Exercise A.27. Suppose $F\colon U \to W$ is differentiable. Show that the linear map L satisfying (A.13) is unique.

If F is differentiable at a, the linear map L satisfying (A.13) is denoted by $DF(a)$ and is called the *total derivative* of F at a. Condition (A.13) can also be written

$$F(a+v) = F(a) + DF(a)v + R(v), \tag{A.14}$$

where the remainder term $R(v)$ satisfies $R(v)/|v| \to 0$ as $v \to 0$. One thing that makes the total derivative so powerful is that it makes sense for arbitrary finite-dimensional vector spaces, without the need to choose a basis or even a norm.

Exercise A.28. Suppose V, W are finite-dimensional vector spaces; $U \subset V$ is an open set; $a \in U$; $F, G\colon U \to W$; and $f, g\colon U \to \mathbb{R}$.

(a) If F is differentiable at a, show that F is continuous at a.

(b) If F and G are differentiable at a, show that $F + G$ is also, and

$$D(F+G)(a) = DF(a) + DG(a).$$

(c) If f and g are differentiable at $a \in U$, show that fg is also, and

$$D(fg)(a) = f(a)Dg(a) + g(a)Df(a).$$

(d) If f is differentiable at a and $f(a) \neq 0$, show that $1/f$ is differentiable at a, and

$$D(1/f)(a) = -(1/f(a)^2)Df(a).$$

Proposition A.24 (The Chain Rule for Total Derivatives). *Suppose V, W, X are finite-dimensional vector spaces, $U \subset V$ and $\widetilde{U} \subset W$ are open sets, and $F\colon U \to \widetilde{U}$ and $G\colon \widetilde{U} \to X$ are maps. If F is differentiable at $a \in U$ and G is differentiable at $F(a) \in \widetilde{U}$, then $G \circ F$ is differentiable at a, and*

$$D(G \circ F)(a) = DG(F(a)) \circ DF(a).$$

Proof. Let $A = DF(a)$ and $B = DG(F(a))$. We need to show that

$$\lim_{v \to 0} \frac{G \circ F(a+v) - G \circ F(a) - BAv}{|v|} = 0. \tag{A.15}$$

We can rewrite the quotient in (A.15) as

$$\frac{G(F(a+v)) - G(F(a)) - B(F(a+v) - F(a))}{|v|} \\ + B\left(\frac{F(a+v) - F(a) - Av}{|v|}\right). \tag{A.16}$$

As $v \to 0$, $F(a+v) - F(a) \to 0$ by continuity of F. Therefore the differentiability of G at $F(a)$ implies that, for any $\varepsilon > 0$, we can make

$$|G(F(a+v)) - G(F(a)) - B(F(a+v) - F(a))| \le \varepsilon |F(a+v) - F(a)|$$

as long as $|v|$ lies in a small enough neighborhood of 0. Thus for $|v|$ small (A.16) is bounded by

$$\varepsilon\frac{|F(a+v)-F(a)|}{|v|}+\left|B\left(\frac{F(a+v)-F(a)-Av}{|v|}\right)\right|. \tag{A.17}$$

Restricting $|v|$ to an even smaller neighborhood of 0, we can ensure that

$$|F(a+v)-F(a)-Av|\le\varepsilon|v|,$$

because of the differentiability of F at a. Since the linear map B is continuous, it follows that (A.17) can be made as small as desired by choosing $|v|$ small enough, thus completing the proof. □

Now let us specialize to the case of maps between Euclidean spaces. Suppose $U\subset\mathbb{R}^n$ is open and $F\colon U\to\mathbb{R}^m$ is differentiable at $a\in U$. As a linear map between Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, $DF(a)$ can be identified with an $m\times n$ matrix. The next lemma identifies that matrix as the Jacobian of F.

Lemma A.25. *Let $U\subset\mathbb{R}^n$ be open, and suppose $F\colon U\to\mathbb{R}^m$ is differentiable at $a\in U$. Then all of the partial derivatives of F at a exist, and $DF(a)$ is the linear map whose matrix is the Jacobian of F at a:*

$$DF(a)=\left(\frac{\partial F^j}{\partial x^i}(a)\right).$$

Proof. Let $B=DF(a)$. Applying the definition of differentiability with $v=te_i$, we obtain

$$0=\lim_{t\to 0}\frac{F^j(a+te_i)-F^j(a)-tB_i^j}{|t|}.$$

Considering $t>0$ and $t<0$ separately, we find

$$\begin{aligned}
0&=\lim_{t\searrow 0}\frac{F^j(a+te_i)-F^j(a)-tB_i^j}{t}\\
&=\lim_{t\searrow 0}\frac{F^j(a+te_i)-F^j(a)}{t}-B_i^j.\\
0&=-\lim_{t\nearrow 0}\frac{F^j(a+te_i)-F^j(a)-tB_i^j}{t}\\
&=-\left(\lim_{t\nearrow 0}\frac{F^j(a+te_i)-F^j(a)}{t}-B_i^j\right).
\end{aligned}$$

Combining these results, we obtain $\partial F^j/\partial x^i(a)=B_i^j$ as claimed. □

Exercise A.29. Suppose $U \subset \mathbb{R}^n$ is open. Show that a map $F\colon U \to \mathbb{R}^m$ is differentiable at $a \in U$ if and only if each of its component functions $F^1, \dots, F^m$ is differentiable at a, and

$$DF(a) = \begin{pmatrix} DF^1(a) \\ \vdots \\ DF^m(a) \end{pmatrix}.$$

The next proposition gives the most important sufficient condition for differentiability; in particular, it shows that all of the usual functions of elementary calculus are differentiable. For a proof, see [Rud76].

Proposition A.26. *Let $U \subset \mathbb{R}^n$ be open. If $F\colon U \to \mathbb{R}^m$ is of class C^1, then it is differentiable at each point of U.*

Exercise A.30. If $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, show that T is differentiable at each $a \in \mathbb{R}^n$, with $DT(a) = T$.

In the case of maps between Euclidean spaces, the chain rule can be rephrased in terms of partial derivatives.

Corollary A.27 (The Chain Rule for Partial Derivatives). *Let $U \subset \mathbb{R}^n$ and $\widetilde{U} \subset \mathbb{R}^m$ be open sets, and let $x = (x^1, \dots, x^n)$ denote the coordinates on U and $y = (y^1, \dots, y^m)$ those on $\widetilde{U}$.*

(*a*) *Any composition of C^1 functions $F\colon U \to \widetilde{U}$ and $G\colon \widetilde{U} \to \mathbb{R}^p$ is again of class C^1, with partial derivatives given by*

$$\frac{\partial (G^i \circ F)}{\partial x^j}(x) = \sum_{k=1}^{m} \frac{\partial G^i}{\partial y^k}(F(x)) \frac{\partial F^k}{\partial x^j}(x).$$

(*b*) *If F and G are smooth, then $G \circ F$ is smooth.*

Exercise A.31. Prove Corollary A.27.

From the chain rule and induction one can derive formulas for the higher partial derivatives of a composite map as needed, provided the maps in question are sufficiently differentiable.

Now suppose $f\colon U \to \mathbb{R}$ is a smooth real-valued function on an open set $U \subset \mathbb{R}^n$, and $a \in U$. For any vector $v \in \mathbb{R}^n$, we define the *directional derivative* of f in the direction v at a to be the number

$$D_v f(a) = \left.\frac{d}{dt}\right|_{t=0} f(a + tv). \tag{A.18}$$

(This definition makes sense for any vector v; we do not require v to be a unit vector as one sometimes does in elementary calculus.)

Since $D_v f(a)$ is the ordinary derivative of the composite map $t \mapsto a + tv \mapsto f(a+tv)$, by the chain rule the directional derivative can be written more concretely as

$$D_v f(a) = \sum_{i=1}^{n} v^i \frac{\partial f}{\partial x^i}(a) = Df(a)v.$$

The next result gives an important estimate for the local behavior of a C^1 function in terms of its derivative. If $U \subset \mathbb{R}^n$ is any subset, a function $F\colon U \to \mathbb{R}^m$ is said to be *Lipschitz continuous* on U if there is a constant C such that

$$|F(x) - F(y)| \le C|x - y| \quad \text{for all } x, y \in U. \tag{A.19}$$

Any such C is called a *Lipschitz constant* for F.

Proposition A.28 (Lipschitz Estimate for C^1 Functions). *Let $U \subset \mathbb{R}^n$ be an open set, and let $F\colon U \to \mathbb{R}^m$ be of class C^1. Then F is Lipschitz continuous on any closed ball $B \subset U$, with Lipschitz constant $M = \sup_{a \in B} |DF(a)|$ (where the norm on $DF(a)$ is the Euclidean norm* (A.11) *for matrices).*

Proof. Let $a, b \in B$ be arbitrary, and define $v = b - a$ and $G(t) = F(a+tv)$. Because B is convex, G is defined and of class C^1 for $t \in [0,1]$. The mean-value theorem of one-variable calculus implies

$$G(1) - G(0) = G'(t_0)(1-0) = G'(t_0)$$

for some $t_0 \in [0,1]$. Using the definition of G, the chain rule, and Exercise A.24, this yields

$$\begin{aligned} |F(b) - F(a)| &= |DF(a+t_0 v)v| \\ &\le |DF(a+t_0 v)||v| \\ &\le M|b-a|, \end{aligned}$$

which was to be proved. □

Multiple Integrals

In this section, we give a brief review of some basic facts regarding multiple integrals in $\mathbb{R}^n$. For our purposes, the Riemann integral will be more than sufficient. Readers who are familiar with the theory of Lebesgue integration are free to interpret all of our integrals in the Lebesgue sense, because the two integrals are equal for the types of functions we will consider. For more details on the aspects of integration theory described here, you can consult nearly any text that treats multivariable calculus rigorously, such as [Apo74, Fle77, Mun91, Rud76, Spi65].

A *rectangle* in $\mathbb{R}^n$ (also called a *closed rectangle*) is a product set of the form $[a^1, b^1] \times \cdots \times [a^n, b^n]$, for real numbers $a^i < b^i$. Analogously, an *open rectangle* is the interior of a closed rectangle, a set of the form $(a^1, b^1) \times \cdots \times (a^n, b^n)$. The *volume* of a rectangle A of either type, denoted by $\mathrm{Vol}(A)$, is defined to be the product of the lengths of its component intervals:

$$\mathrm{Vol}(A) = (b^1 - a^1) \cdots (b^n - a^n).$$

A rectangle is called a *cube* if all of its side lengths $|b_i - a_i|$ are equal.

A *partition* of a closed interval $[a, b]$ is a finite set $P = \{a_0, \ldots, a_k\}$ of real numbers such that $a = a_0 < a_1 < \cdots < a_k = b$. Each of the intervals $[a_{i-1}, a_i]$ for $i = 1, \ldots, k$ is called a *subinterval* of the partition. Similarly, a partition P of a rectangle $A = [a^1, b^1] \times \cdots \times [a^n, b^n]$ is an n-tuple $(P_1, \ldots, P_n)$, where each P_i is a partition of $[a^i, b^i]$. Each rectangle of the form $I_1 \times \cdots \times I_n$, where I_j is a subinterval of P_j, is called a *subrectangle* of P. Clearly A is the union of all the subrectangles in any partition, and distinct subrectangles overlap only on their boundaries.

Suppose $A \subset \mathbb{R}^n$ is a closed rectangle and $f\colon A \to \mathbb{R}$ is a bounded function. For any partition P of A, we define the *lower sum* of f with respect to P by

$$L(f, P) = \sum_j (\inf_{R_j} f) \,\mathrm{Vol}(R_j),$$

where the sum is over all the subrectangles R_j of P. Similarly, the *upper sum* is

$$U(f, P) = \sum_j (\sup_{R_j} f) \,\mathrm{Vol}(R_j).$$

The lower sum with respect to P is obviously less than or equal to the upper sum with respect to the same partition. In fact, more is true.

Lemma A.29. *Let $A \subset \mathbb{R}^n$ be a rectangle, and let $f\colon A \to \mathbb{R}$ be a bounded function. For any pair of partitions P and P' of A,*

$$L(f, P) \le U(f, P').$$

Proof. Write $P = (P_1, \ldots, P_n)$ and $P' = (P'_1, \ldots, P'_n)$, and let Q be the partition $Q = (P_1 \cup P'_1, \ldots, P_n \cup P'_n)$. Each subrectangle of P or P' is a union of finitely many subrectangles of Q. An easy computation shows

$$L(f, P) \le L(f, Q) \le U(f, Q) \le U(f, P'),$$

from which the result follows. □

The *lower integral* of f over A is

$$\underline{\int_A} f\,dV = \sup\{L(f,P) : P \text{ is a partition of } A\},$$

and the *upper integral* is

$$\overline{\int_A} f\,dV = \inf\{U(f,P) : P \text{ is a partition of } A\}.$$

Clearly both numbers exist because f is bounded, and Lemma A.29 implies that the lower integral is less than or equal to the upper integral.

If the upper and lower integrals of f are equal, we say that f is (*Riemann*) *integrable*, and their common value, denoted by

$$\int_A f\,dV,$$

is called the *integral* of f over A. The "dV" in this notation, like the "dx" in the notation for single integrals, does not have any meaning in and of itself; it is just a "closing bracket" for the integral sign. Other notations in common use are

$$\int_A f \quad \text{or} \quad \int_A f\,dx^1 \cdots dx^n \quad \text{or} \quad \int_A f(x^1,\dots,x^n)\,dx^1 \cdots dx^n.$$

In $\mathbb{R}^2$, the symbol dV is often replaced by dA.

There is a simple criterion for a bounded function to be Riemann integrable. It is based on the following notion. A subset $A \subset \mathbb{R}^n$ is said to have *measure zero* if for any $\delta > 0$, there exists a countable cover of A by open cubes $\{C_i\}$ such that $\sum_i \operatorname{Vol}(C_i) < \delta$. (For those who are familiar with the theory of Lebesgue measure, this is equivalent to the condition that the Lebesgue measure of A is equal to zero.)

Lemma A.30 (Properties of Measure Zero Sets).

(*a*) *A countable union of sets of measure zero in* $\mathbb{R}^n$ *has measure zero.*

(*b*) *Any subset of a set of measure zero in* $\mathbb{R}^n$ *has measure zero.*

(*c*) *A set of measure zero in* $\mathbb{R}^n$ *can contain no open set.*

(*d*) *Any proper affine subspace of* $\mathbb{R}^n$ *has measure zero in* $\mathbb{R}^n$.

Exercise A.32. Prove Lemma A.30.

Part (d) of this lemma illustrates that having measure zero is a property of a set in relation to a particular Euclidean space containing it, not of a set in and of itself—for example, an open interval in the x-axis has measure zero as a subset of $\mathbb{R}^2$, but not when considered as a subset of $\mathbb{R}^1$. For this reason, we sometimes say a subset of $\mathbb{R}^n$ has *n-dimensional measure zero* if we wish to emphasize that it has measure zero as a subset of $\mathbb{R}^n$.

The following proposition gives a sufficient condition for a function to be integrable. It shows, in particular, that every bounded continuous function is integrable.

Proposition A.31 (Lebesgue's Integrability Criterion). *Let $A \subset \mathbb{R}^n$ be a rectangle, and let $f\colon A \to \mathbb{R}$ be a bounded function. If the set*

$$S = \{x \in A : f \text{ is not continuous at } x\}$$

has measure zero, then f is integrable.

Proof. Let $\varepsilon > 0$ be given. By definition of measure zero sets, S can be covered by a countable collection of open cubes $\{C_i\}$ with total volume less than ε.

For each point $q \in A \smallsetminus S$, since f is continuous at q, there is a cube D_q centered at q such that $|f(x) - f(q)| < \varepsilon$ for all $x \in D_q \cap A$. This implies $\sup_{D_q} f - \inf_{D_q} f \le 2\varepsilon$.

The collection of all open cubes of the form $\operatorname{Int} C_i$ or $\operatorname{Int} D_q$ is an open cover of A. By compactness, finitely many of them cover A. Let us relabel these cubes as $\{C_1, \dots, C_k, D_1, \dots, D_l\}$. Replacing each C_i or D_j by its intersection with A, we may assume that each C_i and each D_j is a rectangle contained in A.

Since there are only finitely many rectangles $\{C_i, D_j\}$, there is a partition P with the property that each C_i or D_j is equal to a union of subrectangles of P. (Just use the union of all the endpoints of the component intervals of the rectangles C_i and D_j to define the partition.) We can divide the subrectangles of P into two disjoint sets $\mathcal{C}$ and $\mathcal{D}$ such that every subrectangle in $\mathcal{C}$ is contained in C_i for some i, and every subrectangle in $\mathcal{D}$ is contained in D_j for some j. Then

$$\begin{aligned}
&U(f,P) - L(f,P)\\
&\qquad = \sum_i (\sup_{R_i} f) \operatorname{Vol}(R_i) - \sum_i (\inf_{R_i} f) \operatorname{Vol}(R_i)\\
&\qquad = \sum_{R_i \in \mathcal{C}} (\sup_{R_i} f - \inf_{R_i} f) \operatorname{Vol}(R_i) + \sum_{R_i \in \mathcal{D}} (\sup_{R_i} f - \inf_{R_i} f) \operatorname{Vol}(R_i)\\
&\qquad \le (\sup_A f - \inf_A f) \sum_{R_i \in \mathcal{C}} \operatorname{Vol}(R_i) + 2\varepsilon \sum_{R_i \in \mathcal{D}} \operatorname{Vol}(R_i)\\
&\qquad \le (\sup_A f - \inf_A f)\varepsilon + 2\varepsilon \operatorname{Vol}(A).
\end{aligned}$$

It follows that

$$\overline{\int_A} f\,dV - \underline{\int_A} f\,dV \le (\sup_A f - \inf_A f)\varepsilon + 2\varepsilon \operatorname{Vol}(A),$$

which can be made as small as desired by taking ε sufficiently small. This implies that the upper and lower integrals of f must be equal, so f is integrable. $\square$

Remark. In fact, the Lebesgue criterion is both necessary and sufficient for Riemann integrability, but we will not need that.

Now suppose $D \subset \mathbb{R}^n$ is any bounded set, and $f\colon D \to \mathbb{R}$ is a bounded function. Let A be any rectangle containing D, and define $f_D\colon A \to \mathbb{R}$ by

$$f_D(x) = \begin{cases} f(x) & x \in D, \\ 0 & x \in A \smallsetminus D. \end{cases} \tag{A.20}$$

If the integral

$$\int_A f_D \, dV \tag{A.21}$$

exists, f is said to be *integrable over* D, and the integral (A.21) is denoted by $\int_D f \, dV$ and called the *integral of* f *over* D. It is easy to check that the value of the integral does not depend on the rectangle chosen.

In practice, we will be interested only in integrals of bounded continuous functions. However, since we will sometimes need to integrate them over domains other than rectangles, it is necessary to consider also integrals of discontinuous functions such as the function f_D defined by (A.20). The main reason for proving Proposition A.31 is that it allows us to give a simple description of domains on which all bounded continuous functions are integrable.

A subset $D \subset \mathbb{R}^n$ will be called a *domain of integration* if D is bounded and ∂D has n-dimensional measure zero. It is easy to check (using Lemma A.30) that any set whose boundary is contained in a finite union of proper affine subspaces is a domain of integration, and finite unions and intersections of domains of integration are again domains of integration. Thus, for example, any finite union of open or closed rectangles is a domain of integration.

Proposition A.32. *If $D \subset \mathbb{R}^n$ is a domain of integration, then every bounded continuous function on D is integrable over D.*

Proof. Let $f\colon D \to \mathbb{R}$ be bounded and continuous, and let A be a rectangle containing D. To prove the theorem, we need only show that the function $f_D\colon A \to \mathbb{R}$ defined by (A.20) is continuous except on a set of measure zero.

If $x \in \operatorname{Int} D$, then $f_D = f$ on a neighborhood of x, so f_D is continous at x. Similarly, if $x \in A \smallsetminus \overline{D}$, then $f_D \equiv 0$ on a neighborhood of x, so again f is continuous at x. Thus the set of points where f_D is discontinuous is contained in ∂D, and therefore has measure zero. $\square$

Of course, if D is compact, then the assumption that f is bounded in the preceding proposition is superfluous.

If D is a domain of integration, the *volume* of D is defined to be

$$\mathrm{Vol}(D) = \int_D 1\, dV.$$

The integral on the right-hand side is often abbreviated $\int_D dV$.

The next two propositions collect some basic facts about volume and integrals of continuous functions.

Proposition A.33 (Properties of Volume). *Let $D \subset \mathbb{R}^n$ be a domain of integration.*

(*a*) $\mathrm{Vol}(D) \geq 0$*, with equality if and only if D has measure zero.*

(*b*) *If $D_1, \dots, D_k$ are domains of integration whose union is D, then*

$$\mathrm{Vol}(D) \leq \mathrm{Vol}(D_1) + \cdots + \mathrm{Vol}(D_k),$$

with equality if and only if $D_i \cap D_j$ has measure zero for each i, j.

(*c*) *If D_1 is a domain of integration contained in D, then* $\mathrm{Vol}(D_1) \leq \mathrm{Vol}(D)$*, with equality if and only if $D \smallsetminus D_1$ has measure zero.*

Proposition A.34 (Properties of Integrals). *Let $D \subset \mathbb{R}^n$ be a domain of integration, and let $f, g\colon D \to \mathbb{R}$ be continuous and bounded.*

(*a*) *For any $a, b \in \mathbb{R}$,*

$$\int_D (af + bg)\, dV = a \int_D f\, dV + b \int_D g\, dV.$$

(*b*) *If D has measure zero, then $\int_D f\, dV = 0$.*

(*c*) *If $D_1, \dots, D_k$ are domains of integration whose union is D and whose pairwise intersections have measure zero, then*

$$\int_D f\, dV = \int_{D_1} f\, dV + \cdots + \int_{D_k} f\, dV.$$

(*d*) *If $f \geq 0$ on D, then $\int_D f\, dV \geq 0$, with equality if and only if $f \equiv 0$ on* $\mathrm{Int}\, D$.

(*e*) $(\inf_D f)\, \mathrm{Vol}(D) \leq \int_D f\, dV \leq (\sup_D f)\, \mathrm{Vol}(D)$.

(*f*) $\left| \int_D f\, dV \right| \leq \int_D |f|\, dV$.

Exercise A.33. Prove Propositions A.33 and A.34.

There are two more fundamental properties of multiple integrals that we will need. The proofs are too involved to be included in this summary, but you can look them up in the references listed at the beginning of this section if you are interested. Each of these theorems can be stated in various ways, some stronger than others. The versions we give here will be quite sufficient for our applications.

Theorem A.35 (Change of Variables). *Suppose D and E are compact domains of integration in $\mathbb{R}^n$, and $G\colon D \to E$ is a continuous map such that $G|_{\operatorname{Int} D}\colon \operatorname{Int} D \to \operatorname{Int} E$ is a bijective C^1 map with C^1 inverse. For any bounded continuous function $f\colon E \to \mathbb{R}$,*

$$\int_E f\,dV = \int_D (f \circ G) \left|\det\left(\frac{\partial G^i}{\partial x^j}\right)\right| dV.$$

Theorem A.36 (Evaluation by Iterated Integration). *Suppose $E \subset \mathbb{R}^n$ is a compact domain of integration and $g_0, g_1\colon E \to \mathbb{R}$ are continuous functions such that $g_0 \le g_1$ everywhere on E. Let $D \subset \mathbb{R}^{n+1}$ be the subset*

$$D = \{(x^1, \dots, x^n, y) \in \mathbb{R}^{n+1} : x \in E \text{ and } g_0(x) \le y \le g_1(x)\}.$$

Then D is a domain of integration, and

$$\int_D f\,dV = \int_E \left(\int_{g_0(x)}^{g_1(x)} f(x,y)\,dy\right) dV.$$

Of course, there is nothing special about the last variable in this formula; an analogous result holds for any domain D that can be expressed as the set on which one variable is bounded between two continuous functions of the remaining variables.

If the domain E in the preceding theorem is also a region between two graphs, the same theorem can be applied again to E. In particular, the following formula for an integral over a rectangle follows easily by induction.

Corollary A.37. *Let $A = [a^1, b^1] \times \cdots \times [a^n, b^n]$ be a closed rectangle in $\mathbb{R}^n$, and let $f\colon A \to \mathbb{R}$ be continuous. Then*

$$\int_A f\,dV = \int_{a^n}^{b^n} \left(\cdots \left(\int_{a^1}^{b^1} f(x^1, \dots, x^n)\,dx^1\right) \cdots\right) dx^n,$$

and the same is true if the variables in the iterated integral on the right-hand side are reordered in any way.

Sequences and Series of Functions

We conclude with a summary of the most important facts about sequences and series of functions on Euclidean spaces.

Let $S \subset \mathbb{R}^n$ be any subset, and for each integer $i \geq 1$ suppose that $f_i \colon S \to \mathbb{R}^m$ is a function on S. The sequence $\{f_i\}$ is said to *converge pointwise* to $f \colon S \to \mathbb{R}^m$ if for each $a \in S$ and each $\varepsilon > 0$, there exists an integer N such that $i \geq N$ implies $|f_i(a) - f(a)| < \varepsilon$. Pointwise convergence is denoted simply by $f_i \to f$. The sequence is said to *converge uniformly* to f if N can be chosen independently of the point a: for each $\varepsilon > 0$ there exists N such that $i \geq N$ implies $|f_i(a) - f(a)| < \varepsilon$ for every $a \in U$. We will usually indicate uniform convergence by writing "$f_i \to f$ uniformly."

Theorem A.38 (Properties of Uniform Convergence). *Let $S \subset \mathbb{R}^n$, and let $f_i \colon S \to \mathbb{R}^m$ for each integer $i \geq 1$.*

(*a*) *If each f_i is continuous and $f_i \to f$ uniformly, then f is continuous.*

(*b*) *If each f_i is continuous and $f_i \to f$ uniformly, then for any closed domain of integration $D \subset S$,*

$$\lim_{i\to\infty} \int_D f_i \, dV = \int_D f \, dV.$$

(*c*) *If S is open, each f_i is of class C^1, $f_i \to f$ pointwise, and the sequence $\{\partial f_i/\partial x^j\}$ converges uniformly on S as $i \to \infty$, then $\partial f/\partial x^j$ exists on S and*

$$\frac{\partial f}{\partial x^j} = \lim_{i\to\infty} \frac{\partial f_i}{\partial x^j}.$$

For a proof, see [Rud76]. An infinite series of functions $\sum_{i=0}^{\infty} f_i$ on $S \subset \mathbb{R}^n$ is said to converge pointwise to a function g if the corresponding sequence of partial sums converges pointwise:

$$g(x) = \lim_{M\to\infty} \sum_{i=0}^{M} f_i(x) \text{ for all } x \in S.$$

The series is said to converge uniformly if the partial sums converge uniformly.

Proposition A.39 (Weierstrass M-test). *Suppose $S \subset \mathbb{R}^n$ is any subset, and $f_i \colon S \to \mathbb{R}^k$ are functions. If there exist positive real numbers M_i such that $\sup_S |f_i| \leq M_i$ and $\sum_i M_i$ converges, then $\sum_i f_i$ converges uniformly on A.*

Exercise A.34. Prove Proposition A.39.

References

[AM78] Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings, Reading, MA, second edition, 1978.

[Apo74] Tom M. Apostol. *Mathematical Analysis*. Addison-Wesley, Reading, Massachusetts, second edition, 1974.

[BM58] Raoul Bott and John Milnor. On the parallelizability of the spheres. *Bull. Amer. Math. Soc.*, 64:87–89, 1958.

[Bre93] Glen E. Bredon. *Topology and Geometry*. Springer-Verlag, New York, 1993.

[DK90] S. K. Donaldson and P. B. Kronheimer. *The geometry of four-manifolds*. Clarendon Press, New York, 1990.

[FIS97] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence. *Linear Algebra*. Prentice Hall, Upper Saddle River, NJ, third edition, 1997.

[Fle77] Wendell Fleming. *Functions of Several Variables*. Springer-Verlag, New York, second edition, 1977.

[FQ90] Michael Freedman and Frank Quinn. *Topology of 4-manifolds*. Princeton University Press, Princeton, 1990.

[Her75] Israel N. Herstein. *Topics in Algebra*. Wiley, New York, second edition, 1975.

[HL82] Reese Harvey and H. Blaine Lawson, Jr. Calibrated geometries. *Acta Math.*, 148:47–157, 1982.

[Hun90] Thomas W. Hungerford. *Abstract Algebra: An Introduction*. Saunders College Publishing, Philadelphia, 1990.

[Ker60] Michel A. Kervaire. A manifold which does not admit any differentiable structure. *Comment. Math. Helv.*, 34:257–270, 1960.

[KM63] Michel A. Kervaire and John W. Milnor. Groups of homotopy spheres: I. *Annals of Math.*, 77:504–537, 1963.

[Lee97] John M. Lee. *Riemannian Manifolds: An Introduction to Curvature*. Springer-Verlag, New York, 1997.

[Lee00] John M. Lee. *Introduction to Topological Manifolds*. Springer-Verlag, New York, 2000.

[Mil65] John W. Milnor. *Topology from the Differentiable Viewpoint*. Princeton University Press, Princeton, 1965.

[Moi77] Edwin E. Moise. *Geometric Topology in Dimensions 2 and 3*. Springer-Verlag, New York, 1977.

[Mos65] Jürgen Moser. On the volume elements on a manifold. *Trans. Amer. Math. Soc.*, 120, 1965.

[Mun60] James R. Munkres. Obstructions to the smoothing of piecewise differentiable homeomorphisms. *Annals of Math.*, 72:521–554, 1960.

[Mun75] James R. Munkres. *Topology: A First Course*. Prentice Hall, Englewood Cliffs, New Jersey, 1975.

[Mun84] James R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, Menlo Park, California, 1984.

[Mun91] James R. Munkres. *Analysis on Manifolds*. Addison-Wesley, Redwood City, California, 1991.

[Rud76] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, third edition, 1976.

[Spa89] Edwin H. Spanier. *Algebraic Topology*. Springer-Verlag, New York, 1989.

[Spi65] Michael Spivak. *Calculus on Manifolds*. W. A. Benjamin, New York, 1965.

[Ste64] Shlomo Sternberg. *Lectures on Differential Geometry*. Prentice Hall, Englewood Cliffs, New Jersey, 1964.

[Var84] V. S. Varadarajan. *Lie Groups, Lie Algebras, and Their Representations*. Springer-Verlag, New York, 1984.

[Wei69] Alan Weinstein. Symplectic structures on Banach manifolds. *Bull. Amer. Math. Soc.*, 75:1040–1041, 1969.

[Wei77] Alan Weinstein. *Lectures on Symplectic Manifolds*. American Mathematical Society, Providence, R.I., 1977. CBMS Regional Conf. Ser. in Math., No. 29.

[Whi36] Hassler Whitney. The word problem and the isomorphism problem for groups. *Ann. of Math. (2)*, 37:645–680, 1936.

[Whi44a] Hassler Whitney. The self-intersections of a smooth n-manifold in $2n$-space. *Ann. of Math. (2)*, 45:220–246, 1944.

[Whi44b] Hassler Whitney. The singularities of a smooth n-manifold in $(2n-1)$-space. *Ann. of Math. (2)*, 45:247–293, 1944.

Index