Appendix A
Review of Topology

This book is written for readers who have already completed a rigorous course in basic topology, including an introduction to the fundamental group and covering maps. A convenient source for this material is [LeeTM], which covers all the topological ideas we need, and uses notations and conventions that are compatible with those in the present book. But almost any other good topology text would do as well, such as [Mun00, Sie92, Mas89]. In this appendix we state the most important definitions and results, with most of the proofs left as exercises. If you have had sufficient exposure to topology, these exercises should be straightforward, although you might want to look a few of them up in the topology texts listed above.

Topological Spaces

We begin with the definitions. Let X be a set. A ***topology on X*** is a collection $\mathcal{T}$ of subsets of X, called ***open subsets***, satisfying

(i) X and $\varnothing$ are open.
(ii) The union of any family of open subsets is open.
(iii) The intersection of any finite family of open subsets is open.

A pair $(X, \mathcal{T})$ consisting of a set X together with a topology $\mathcal{T}$ on X is called a ***topological space***. Ordinarily, when the topology is understood, one omits mention of it and simply says "X is a topological space."

There are a host of constructions and definitions associated with topological spaces. Here we summarize the ones that are most important for this book.

Suppose X is a topological space, $p \in X$, and $S \subseteq X$.

- A ***neighborhood of p*** is an open subset containing p. Similarly, a ***neighborhood of the set S*** is an open subset containing S. (Be warned that some authors use the word "neighborhood" in the more general sense of a subset containing an open subset containing p or S.)
- S is said to be ***closed*** if $X \smallsetminus S$ is open (where $X \smallsetminus S$ denotes the ***set difference*** $\{x \in X : x \notin S\}$).

J.M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics 218,
DOI 10.1007/978-1-4419-9982-5,

- The ***interior of S***, denoted by $\operatorname{Int} S$, is the union of all open subsets of X contained in S.
- The ***exterior of S***, denoted by $\operatorname{Ext} S$, is the union of all open subsets of X contained in $X \smallsetminus S$.
- The ***closure of S***, denoted by $\overline{S}$, is the intersection of all closed subsets of X containing S.
- The ***boundary of S***, denoted by ∂S, is the set of all points of X that are in neither $\operatorname{Int} S$ nor $\operatorname{Ext} S$.
- A point $p \in S$ is said to be an ***isolated point of S*** if p has a neighborhood $U \subseteq X$ such that $U \cap S = \{p\}$.
- A point $p \in X$ (not necessarily in S) is said to be a ***limit point of S*** if every neighborhood of p contains at least one point of S other than p.
- S is said to be ***dense in X*** if $\overline{S} = X$, or equivalently if every nonempty open subset of X contains at least one point of S.
- S is said to be ***nowhere dense in X*** if $\overline{S}$ contains no nonempty open subset.

The most important concepts of topology are continuous maps and convergent sequences, which we define next. Let X and Y be topological spaces.

- A map $F\colon X \to Y$ is said to be ***continuous*** if for every open subset $U \subseteq Y$, the preimage $F^{-1}(U)$ is open in X.
- A continuous bijective map $F\colon X \to Y$ with continuous inverse is called a ***homeomorphism***. If there exists a homeomorphism from X to Y, we say that X and Y are ***homeomorphic***.
- A continuous map $F\colon X \to Y$ is said to be a ***local homeomorphism*** if every point $p \in X$ has a neighborhood $U \subseteq X$ such that $F(U)$ is open in Y and F restricts to a homeomorphism from U to $F(U)$.
- Given a sequence $(p_i)_{i=1}^{\infty}$ of points in X and a point $p \in X$, the sequence is said to ***converge to p*** if for every neighborhood U of p, there exists a positive integer N such that $p_i \in U$ for all $i \geq N$. In this case, we write $p_i \to p$ or $\lim_{i\to\infty} p_i = p$.

► **Exercise A.1.** Let $F\colon X \to Y$ be a map between topological spaces. Prove that each of the following properties is equivalent to continuity of F:

(a) For every subset $A \subseteq X$, $F\big(\overline{A}\big) \subseteq \overline{F(A)}$.
(b) For every subset $B \subseteq Y$, $F^{-1}(\operatorname{Int} B) \subseteq \operatorname{Int} F^{-1}(B)$.

► **Exercise A.2.** Let X, Y, and Z be topological spaces. Show that the following maps are continuous:

(a) The ***identity map*** $\operatorname{Id}_X\colon X \to X$, defined by $\operatorname{Id}_X(x) = x$ for all $x \in X$.
(b) Any ***constant map*** $F\colon X \to Y$ (i.e., a map such that $F(x) = F(y)$ for all $x, y \in X$).
(c) Any composition $G \circ F$ of continuous maps $F\colon X \to Y$ and $G\colon Y \to Z$.

► **Exercise A.3.** Let X and Y be topological spaces. Suppose $F\colon X \to Y$ is continuous and $p_i \to p$ in X. Show that $F(p_i) \to F(p)$ in Y.

The most important examples of topological spaces, from which most of our examples of manifolds are built in one way or another, are described below.

Example A.4 (Discrete Spaces). If X is an arbitrary set, the ***discrete topology*** on X is the topology defined by declaring every subset of X to be open. Any space that has the discrete topology is called a ***discrete space***. //

Example A.5 (Metric Spaces). A ***metric space*** is a set M endowed with a ***distance function*** (also called a ***metric***) $d\colon M \times M \to \mathbb{R}$ (where $\mathbb{R}$ denotes the set of real numbers) satisfying the following properties for all $x, y, z \in M$:

(i) POSITIVITY: $d(x,y) \geq 0$, with equality if and only if $x = y$.
(ii) SYMMETRY: $d(x,y) = d(y,x)$.
(iii) TRIANGLE INEQUALITY: $d(x,z) \leq d(x,y) + d(y,z)$.

If M is a metric space, $x \in M$, and $r > 0$, the ***open ball of radius r around x*** is the set

$$B_r(x) = \{y \in M : d(x,y) < r\},$$

and the ***closed ball of radius r*** is

$$\overline{B}_r(x) = \{y \in M : d(x,y) \leq r\}.$$

The ***metric topology on M*** is defined by declaring a subset $S \subseteq M$ to be open if for every point $x \in S$, there is some $r > 0$ such that $B_r(x) \subseteq S$. //

If M is a metric space and S is any subset of M, the restriction of the distance function to pairs of points in S turns S into a metric space and thus also a topological space. We use the following standard terminology for metric spaces:

- A subset $S \subseteq M$ is ***bounded*** if there exists a positive number R such that $d(x,y) \leq R$ for all $x, y \in S$.
- If S is a nonempty bounded subset of M, the ***diameter of S*** is the number $\operatorname{diam} S = \sup\{d(x,y) : x, y \in S\}$.
- A sequence of points $(x_i)_{i=1}^{\infty}$ in M is a ***Cauchy sequence*** if for every $\varepsilon > 0$, there exists an integer N such that $i, j \geq N$ implies $d(x_i, x_j) < \varepsilon$.
- A metric space M is said to be ***complete*** if every Cauchy sequence in M converges to a point of M.

Example A.6 (Euclidean Spaces). For each integer $n \geq 1$, the set $\mathbb{R}^n$ of ordered n-tuples of real numbers is called ***n-dimensional Euclidean space***. We denote a point in $\mathbb{R}^n$ by $(x^1, \dots, x^n)$, (x^i), or x; the numbers x^i are called the ***components*** or ***coordinates of x***. (When n is small, we often use more traditional names such as (x,y,z) for the coordinates.) Notice that we write the coordinates of a point $(x^1, \dots, x^n) \in \mathbb{R}^n$ with superscripts, not subscripts as is usually done in linear algebra and calculus books, so as to be consistent with the Einstein summation convention, explained in Chapter 1. By convention, $\mathbb{R}^0$ is the one-element set $\{0\}$.

For each $x \in \mathbb{R}^n$, the ***Euclidean norm of x*** is the nonnegative real number

$$|x| = \sqrt{(x^1)^2 + \cdots + (x^n)^2},$$

and for $x, y \in \mathbb{R}^n$, the ***Euclidean distance function*** is defined by

$$d(x, y) = |x - y|.$$

This distance function turns $\mathbb{R}^n$ into a complete metric space. The resulting metric topology on $\mathbb{R}^n$ is called the ***Euclidean topology***. //

Example A.7 (Complex Euclidean Spaces). We also sometimes have occasion to work with complex Euclidean spaces. We consider the set $\mathbb{C}$ of complex numbers, as a set, to be simply $\mathbb{R}^2$, with the complex number $x + iy$ corresponding to $(x, y) \in \mathbb{R}^2$. For any positive integer n, the ***n-dimensional complex Euclidean space*** is the set $\mathbb{C}^n$ of ordered n-tuples of complex numbers. It becomes a topological space when identified with $\mathbb{R}^{2n}$ via the correspondence

$$\left(x^1 + iy^1, \dots, x^n + iy^n\right) \leftrightarrow \left(x^1, y^1, \dots, x^n, y^n\right). \qquad //$$

Example A.8 (Subsets of Euclidean Spaces). Every subset of $\mathbb{R}^n$ or $\mathbb{C}^n$ becomes a metric space, and thus a topological space, when endowed with the Euclidean metric. Whenever we mention such a subset, it is always assumed to have this metric topology unless otherwise specified. It is a complete metric space if and only if it is a closed subset of $\mathbb{R}^n$. Here are some standard subsets of Euclidean spaces that we work with frequently:

- The ***unit interval*** is the subset $I \subseteq \mathbb{R}$ defined by
$$I = [0, 1] = \{x \in \mathbb{R} : 0 \le x \le 1\}.$$
- The **(*open*) *unit ball of dimension n*** is the subset $\mathbb{B}^n \subseteq \mathbb{R}^n$ defined by
$$\mathbb{B}^n = \left\{x \in \mathbb{R}^n : |x| < 1\right\}.$$
- The ***closed unit ball of dimension n*** is the subset $\overline{\mathbb{B}}^n \subseteq \mathbb{R}^n$ defined by
$$\overline{\mathbb{B}}^n = \left\{x \in \mathbb{R}^n : |x| \le 1\right\}.$$
The terms **(*open*) *unit disk*** and ***closed unit disk*** are commonly used for $\mathbb{B}^2$ and $\overline{\mathbb{B}}^2$, respectively.
- For $n \ge 0$, the **(*unit*) *n-sphere*** is the subset $\mathbb{S}^n \subseteq \mathbb{R}^{n+1}$ defined by
$$\mathbb{S}^n = \left\{x \in \mathbb{R}^{n+1} : |x| = 1\right\}.$$
Sometimes it is useful to think of an odd-dimensional sphere $\mathbb{S}^{2n+1}$ as a subset of $\mathbb{C}^{n+1}$, by means of the usual identification of $\mathbb{C}^{n+1}$ with $\mathbb{R}^{2n+2}$.
- The **(*unit*) *circle*** is the 1-sphere $\mathbb{S}^1$, considered either as a subset of $\mathbb{R}^2$ or as a subset of $\mathbb{C}$. //

Hausdorff Spaces

Topological spaces allow us to describe a wide variety of concepts of "spaces." But for the purposes of manifold theory, arbitrary topological spaces are far too general, because they can have some unpleasant properties, as the next exercise illustrates.

► **Exercise A.9.** Let X be any set. Show that $\{X, \varnothing\}$ is a topology on X, called the ***trivial topology***. Show that when X is endowed with this topology, every sequence in X converges to every point of X, and every map from a topological space into X is continuous.

To avoid pathological cases like this, which result when X does not have sufficiently many open subsets, we often restrict our attention to topological spaces satisfying the following special condition. A topological space X is said to be a ***Hausdorff space*** if for every pair of distinct points $p, q \in X$, there exist disjoint open subsets $U, V \subseteq X$ such that $p \in U$ and $q \in V$.

► **Exercise A.10.** Show that every metric space is Hausdorff in the metric topology.

► **Exercise A.11.** Let X be a Hausdorff space. Show that each finite subset of X is closed, and that each convergent sequence in X has a unique limit.

Bases and Countability

Suppose X is a topological space. A collection $\mathscr{B}$ of open subsets of X is said to be a ***basis for the topology of X*** (plural: ***bases***) if every open subset of X is the union of some collection of elements of $\mathscr{B}$.

More generally, suppose X is merely a set, and $\mathscr{B}$ is a collection of subsets of X satisfying the following conditions:

(i) $X = \bigcup_{B \in \mathscr{B}} B$.
(ii) If $B_1, B_2 \in \mathscr{B}$ and $x \in B_1 \cap B_2$, then there exists $B_3 \in \mathscr{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

Then the collection of all unions of elements of $\mathscr{B}$ is a topology on X, called the ***topology generated by*** $\mathscr{B}$, and $\mathscr{B}$ is a basis for this topology.

If X is a topological space and $p \in X$, a ***neighborhood basis at p*** is a collection $\mathscr{B}_p$ of neighborhoods of p such that every neighborhood of p contains at least one $B \in \mathscr{B}_p$.

A set is said to be ***countably infinite*** if it admits a bijection with the set of positive integers, and ***countable*** if it is finite or countably infinite. A topological space X is said to be ***first-countable*** if there is a countable neighborhood basis at each point, and ***second-countable*** if there is a countable basis for its topology. Since a countable basis for X contains a countable neighborhood basis at each point, second-countability implies first-countability.

The next lemma expresses the most important properties of first-countable spaces. To say that a sequence is ***eventually in a subset*** means that all but finitely many terms of the sequence are in the subset.

Lemma A.12 (Sequence Lemma). *Let X be a first-countable space, let $A \subseteq X$ be any subset, and let $x \in X$.*

(a) *$x \in \overline{A}$ if and only if x is a limit of a sequence of points in A.*
(b) *$x \in \operatorname{Int} A$ if and only if every sequence in X converging to x is eventually in A.*

(c) *A is closed in X if and only if A contains every limit of every convergent sequence of points in A.*
(d) *A is open in X if and only if every sequence in X converging to a point of A is eventually in A.*

► **Exercise A.13.** Prove the sequence lemma.

► **Exercise A.14.** Show that every metric space is first-countable.

► **Exercise A.15.** Show that the set of all open balls in $\mathbb{R}^n$ whose radii are rational and whose centers have rational coordinates is a countable basis for the Euclidean topology, and thus $\mathbb{R}^n$ is second-countable.

One of the most important properties of second-countable spaces is expressed in the following proposition. Let X be a topological space. A ***cover of X*** is a collection $\mathcal{U}$ of subsets of X whose union is X; it is called an ***open cover*** if each of the sets in $\mathcal{U}$ is open. A ***subcover of $\mathcal{U}$*** is a subcollection of $\mathcal{U}$ that is still a cover.

Proposition A.16. *Let X be a second-countable topological space. Every open cover of X has a countable subcover.*

Proof. Let $\mathcal{B}$ be a countable basis for X, and let $\mathcal{U}$ be an arbitrary open cover of X. Let $\mathcal{B}' \subseteq \mathcal{B}$ be the collection of basis open subsets $B \in \mathcal{B}$ such that $B \subseteq U$ for some $U \in \mathcal{U}$. For each $B \in \mathcal{B}'$, choose a particular set $U_B \in \mathcal{U}$ containing B. The collection $\{U_B : B \in \mathcal{B}'\}$ is countable, so it suffices to show that it covers X. Given a point $x \in X$, there is some $V \in \mathcal{U}$ containing x, and because $\mathcal{B}$ is a basis there exists $B \in \mathcal{B}$ such that $x \in B \subseteq V$. This implies, in particular, that $B \in \mathcal{B}'$, and therefore $x \in B \subseteq U_B$. □

Subspaces, Products, Disjoint Unions, and Quotients

Subspaces

Probably the simplest way to obtain new topological spaces from old ones is by taking subsets of other spaces. If X is a topological space and $S \subseteq X$ is an arbitrary subset, we define the ***subspace topology on S*** (sometimes called the ***relative topology***) by declaring a subset $U \subseteq S$ to be open in S if and only if there exists an open subset $V \subseteq X$ such that $U = V \cap S$. A subset of S that is open or closed in the subspace topology is sometimes said to be ***relatively open*** or ***relatively closed in S***, to make it clear that we do not mean open or closed as a subset of X. Any subset of X endowed with the subspace topology is said to be a ***subspace of X***. Whenever we treat a subset of a topological space as a space in its own right, we always assume that it has the subspace topology unless otherwise specified.

If X and Y are topological spaces, a continuous injective map $F: X \to Y$ is called a ***topological embedding*** if it is a homeomorphism onto its image $F(X) \subseteq Y$ in the subspace topology.

The most important properties of the subspace topology are summarized in the next proposition.

Proposition A.17 (Properties of the Subspace Topology). *Let X be a topological space and let S be a subspace of X.*

(a) CHARACTERISTIC PROPERTY: *If Y is a topological space, a map $F: Y \to S$ is continuous if and only if the composition $\iota_S \circ F: Y \to X$ is continuous, where $\iota_S: S \hookrightarrow X$ is the inclusion map (the restriction of the identity map of X to S).*
(b) *The subspace topology is the unique topology on S for which the characteristic property holds.*
(c) *A subset $K \subseteq S$ is closed in S if and only if there exists a closed subset $L \subseteq X$ such that $K = L \cap S$.*
(d) *The inclusion map $\iota_S: S \hookrightarrow X$ is a topological embedding.*
(e) *If Y is a topological space and $F: X \to Y$ is continuous, then $F|_S: S \to Y$ (the restriction of F to S) is continuous.*
(f) *If $\mathcal{B}$ is a basis for the topology of X, then $\mathcal{B}_S = \{B \cap S : B \in \mathcal{B}\}$ is a basis for the subspace topology on S.*
(g) *If X is Hausdorff, then so is S.*
(h) *If X is first-countable, then so is S.*
(i) *If X is second-countable, then so is S.*

► **Exercise A.18.** Prove the preceding proposition.

If X and Y are topological spaces and $F: X \to Y$ is a continuous map, part (e) of the preceding proposition guarantees that the restriction of F to every subspace of X is continuous (in the subspace topology). We can also ask the converse question: If we know that the restriction of F to certain subspaces of X is continuous, is F itself continuous? The next two propositions express two somewhat different answers to this question.

Lemma A.19 (Continuity Is Local). *Continuity is a local property, in the following sense: if $F: X \to Y$ is a map between topological spaces such that every point $p \in X$ has a neighborhood U on which the restriction $F|_U$ is continuous, then F is continuous.*

Lemma A.20 (Gluing Lemma for Continuous Maps). *Let X and Y be topological spaces, and suppose one of the following conditions holds:*

(a) *$B_1, \dots, B_n$ are finitely many closed subsets of X whose union is X.*
(b) *$\{B_i\}_{i \in A}$ is a collection of open subsets of X whose union is X.*

Suppose that for all i we are given continuous maps $F_i: B_i \to Y$ that agree on overlaps: $F_i|_{B_i \cap B_j} = F_j|_{B_i \cap B_j}$. Then there exists a unique continuous map $F: X \to Y$ whose restriction to each B_i is equal to F_i.

► **Exercise A.21.** Prove the two preceding lemmas.

► **Exercise A.22.** Let X be a topological space, and suppose X admits a countable open cover $\{U_i\}$ such that each set U_i is second-countable in the subspace topology. Show that X is second-countable.

Product Spaces

Next we consider finite products of topological spaces. If $X_1, \dots, X_k$ are (finitely many) sets, their ***Cartesian product*** is the set $X_1 \times \dots \times X_k$ consisting of all ordered k-tuples of the form $(x_1, \dots, x_k)$ with $x_i \in X_i$ for each i. The ***i th projection map*** is the map $\pi_i : X_1 \times \dots \times X_k \to X_i$ defined by $\pi_i(x_1, \dots, x_k) = x_i$.

Suppose $X_1, \dots, X_k$ are topological spaces. The collection of all subsets of $X_1 \times \dots \times X_k$ of the form $U_1 \times \dots \times U_k$, where each U_i is open in X_i, forms a basis for a topology on $X_1 \times \dots \times X_k$, called the ***product topology***. Endowed with this topology, a finite product of topological spaces is called a ***product space***. Any open subset of the form $U_1 \times \dots \times U_k \subseteq X_1 \times \dots \times X_k$, where each U_i is open in X_i, is called a ***product open subset***. (A slightly different definition is required for products of infinitely many spaces, but we need only the finite case. See [LeeTM] for more about infinite product spaces.)

Proposition A.23 (Properties of the Product Topology). *Suppose $X_1, \dots, X_k$ are topological spaces, and let $X_1 \times \dots \times X_k$ be their product space.*

(a) CHARACTERISTIC PROPERTY: *If B is a topological space, a map $F : B \to X_1 \times \dots \times X_k$ is continuous if and only if each of its component functions $F_i = \pi_i \circ F : B \to X_i$ is continuous.*
(b) *The product topology is the unique topology on $X_1 \times \dots \times X_k$ for which the characteristic property holds.*
(c) *Each projection map $\pi_i : X_1 \times \dots \times X_k \to X_i$ is continuous.*
(d) *Given any continuous maps $F_i : X_i \to Y_i$ for $i = 1, \dots, k$, the **product map** $F_1 \times \dots \times F_k : X_1 \times \dots \times X_k \to Y_1 \times \dots \times Y_k$ is continuous, where*

$$F_1 \times \dots \times F_k(x_1, \dots, x_k) = \big(F_1(x_1), \dots, F_k(x_k)\big).$$

(e) *If S_i is a subspace of X_i for $i = 1, \dots, n$, the product topology and the subspace topology on $S_1 \times \dots \times S_n \subseteq X_1 \times \dots \times X_n$ coincide.*
(f) *For any $i \in \{1, \dots, k\}$ and any choices of points $a_j \in X_j$ for $j \neq i$, the map $x \mapsto (a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_k)$ is a topological embedding of X_i into the product space $X_1 \times \dots \times X_k$.*
(g) *If $\mathcal{B}_i$ is a basis for the topology of X_i for $i = 1, \dots, k$, then the collection*

$$\mathcal{B} = \{B_1 \times \dots \times B_k : B_i \in \mathcal{B}_i\}$$

is a basis for the topology of $X_1 \times \dots \times X_k$.
(h) *Every finite product of Hausdorff spaces is Hausdorff.*
(i) *Every finite product of first-countable spaces is first-countable.*
(j) *Every finite product of second-countable spaces is second-countable.*

► **Exercise A.24.** Prove the preceding proposition.

Disjoint Union Spaces

Another simple way of building new topological spaces is by taking disjoint unions of other spaces. From a set-theoretic point of view, the disjoint union is defined as follows. If $(X_\alpha)_{\alpha\in A}$ is an indexed family of sets, their ***disjoint union*** is the set

$$\coprod_{\alpha\in A} X_\alpha = \big\{(x,\alpha) : \alpha\in A,\ x\in X_\alpha\big\}.$$

For each α, there is a canonical injective map $\iota_\alpha\colon X_\alpha \to \coprod_{\alpha\in A} X_\alpha$ given by $\iota_\alpha(x) = (x,\alpha)$, and the images of these maps for different values of α are disjoint. Typically, we implicitly identify X_α with its image in the disjoint union, thereby viewing X_α as a subset of $\coprod_{\alpha\in A} X_\alpha$. The α in the notation (x,α) should be thought of as a "tag" to indicate which set x comes from, so that the subsets corresponding to different values of α are disjoint, even if some or all of the original sets X_α were identical.

Given an indexed family of topological spaces $(X_\alpha)_{\alpha\in A}$, we define the ***disjoint union topology*** on $\coprod_{\alpha\in A} X_\alpha$ by declaring a subset of $\coprod_{\alpha\in A} X_\alpha$ to be open if and only if its intersection with each X_α is open in X_α.

Proposition A.25 (Properties of the Disjoint Union Topology). *Suppose $(X_\alpha)_{\alpha\in A}$ is an indexed family of topological spaces, and $\coprod_{\alpha\in A} X_\alpha$ is endowed with the disjoint union topology.*

(a) CHARACTERISTIC PROPERTY: *If Y is a topological space, a map*

$$F\colon \coprod_{\alpha\in A} X_\alpha \to Y$$

is continuous if and only if $F\circ\iota_\alpha\colon X_\alpha\to Y$ is continuous for each $\alpha\in A$.

(b) *The disjoint union topology is the unique topology on $\coprod_{\alpha\in A} X_\alpha$ for which the characteristic property holds.*

(c) *A subset of $\coprod_{\alpha\in A} X_\alpha$ is closed if and only if its intersection with each X_α is closed.*

(d) *Each injection $\iota_\alpha\colon X_\alpha\to\coprod_{\alpha\in A} X_\alpha$ is a topological embedding.*

(e) *Every disjoint union of Hausdorff spaces is Hausdorff.*

(f) *Every disjoint union of first-countable spaces is first-countable.*

(g) *Every disjoint union of countably many second-countable spaces is second-countable.*

► **Exercise A.26.** Prove the preceding proposition.

Quotient Spaces and Quotient Maps

If X is a topological space, Y is a set, and $\pi\colon X\to Y$ is a surjective map, the ***quotient topology on Y determined by π*** is defined by declaring a subset $U\subseteq Y$ to be open if and only if $\pi^{-1}(U)$ is open in X. If X and Y are topological spaces, a map $\pi\colon X\to Y$ is called a ***quotient map*** if it is surjective and continuous and Y has the quotient topology determined by π.

The following construction is the most common way of producing quotient maps. A relation $\sim$ on a set X is called an ***equivalence relation*** if it is ***reflexive*** ($x\sim x$ for all $x\in X$), ***symmetric*** ($x\sim y$ implies $y\sim x$), and ***transitive*** ($x\sim y$ and $y\sim z$ imply

$x \sim z$). If $R \subseteq X \times X$ is any relation on X, then the intersection of all equivalence relations on X containing R is an equivalence relation, called the ***equivalence relation generated by*** $\boldsymbol{R}$. If $\sim$ is an equivalence relation on X, then for each $x \in X$, the ***equivalence class of*** $\boldsymbol{x}$, denoted by $[x]$, is the set of all $y \in X$ such that $y \sim x$. The set of all equivalence classes is a ***partition of*** $\boldsymbol{X}$: a collection of disjoint nonempty subsets whose union is X.

Suppose X is a topological space and $\sim$ is an equivalence relation on X. Let $X/{\sim}$ denote the set of equivalence classes in X, and let $\pi\colon X \to X/{\sim}$ be the natural projection sending each point to its equivalence class. Endowed with the quotient topology determined by π, the space $X/{\sim}$ is called the ***quotient space*** (or ***identification space***) ***of*** $\boldsymbol{X}$ ***determined by*** $\boldsymbol{\sim}$. For example, suppose X and Y are topological spaces, $A \subseteq Y$ is a closed subset, and $f\colon A \to X$ is a continuous map. The relation $a \sim f(a)$ for all $a \in A$ generates an equivalence relation on $X \amalg Y$, whose quotient space is denoted by $X \cup_f Y$ and called an ***adjunction space***. It is said to be formed by ***attaching Y to X along f***.

If $\pi\colon X \to Y$ is a map, a subset $U \subseteq X$ is said to be ***saturated with respect to*** $\boldsymbol{\pi}$ if U is the entire preimage of its image: $U = \pi^{-1}\big(\pi(U)\big)$. Given $y \in Y$, the ***fiber of*** $\boldsymbol{\pi}$ ***over*** $\boldsymbol{y}$ is the set $\pi^{-1}(y)$. Thus, a subset of X is saturated if and only if it is a union of fibers.

Theorem A.27 (Properties of Quotient Maps). *Let $\pi\colon X \to Y$ be a quotient map.*

(a) CHARACTERISTIC PROPERTY: *If B is a topological space, a map $F\colon Y \to B$ is continuous if and only if $F \circ \pi\colon X \to B$ is continuous.*
(b) *The quotient topology is the unique topology on Y for which the characteristic property holds.*
(c) *A subset $K \subseteq Y$ is closed if and only if $\pi^{-1}(K)$ is closed in X.*
(d) *If π is injective, then it is a homeomorphism.*
(e) *If $U \subseteq X$ is a saturated open or closed subset, then the restriction $\pi|_U\colon U \to \pi(U)$ is a quotient map.*
(f) *Any composition of π with another quotient map is again a quotient map.*

► **Exercise A.28.** Prove the preceding theorem.

► **Exercise A.29.** Let X and Y be topological spaces, and suppose that $F\colon X \to Y$ is a surjective continuous map. Show that the following are equivalent:

(a) F is a quotient map.
(b) F takes saturated open subsets to open subsets.
(c) F takes saturated closed subsets to closed subsets.

The next two properties of quotient maps play important roles in topology, and have equally important generalizations in smooth manifold theory (see Chapter 4).

Theorem A.30 (Passing to the Quotient). *Suppose $\pi\colon X \to Y$ is a quotient map, B is a topological space, and $F\colon X \to B$ is a continuous map that is constant on the fibers of π (i.e., $\pi(p) = \pi(q)$ implies $F(p) = F(q)$). Then there exists a unique continuous map $\widetilde{F}\colon Y \to B$ such that $F = \widetilde{F} \circ \pi$.*

Proof. The existence and uniqueness of $\widetilde{F}$ follow from set-theoretic considerations, and its continuity is an immediate consequence of the characteristic property of the quotient topology. □

Theorem A.31 (Uniqueness of Quotient Spaces). *If $\pi_1 \colon X \to Y_1$ and $\pi_2 \colon X \to Y_2$ are quotient maps that are constant on each other's fibers (i.e., $\pi_1(p) = \pi_1(q)$ if and only if $\pi_2(p) = \pi_2(q)$), then there exists a unique homeomorphism $\varphi \colon Y_1 \to Y_2$ such that $\varphi \circ \pi_1 = \pi_2$.*

Proof. Applying the preceding theorem to the quotient map $\pi_1 \colon X \to Y_1$, we see that π_2 passes to the quotient, yielding a continuous map $\widetilde{\pi}_2 \colon Y_1 \to Y_2$ satisfying $\widetilde{\pi}_2 \circ \pi_1 = \pi_2$. Applying the same argument with the roles of π_1 and π_2 reversed, there is a continuous map $\widetilde{\pi}_1 \colon Y_2 \to Y_1$ satisfying $\widetilde{\pi}_1 \circ \pi_2 = \pi_1$. Together, these identities imply that $\widetilde{\pi}_2 \circ \widetilde{\pi}_1 \circ \pi_2 = \pi_2$. Applying Theorem A.30 again with π_2 playing the roles of both π and F, we see that both $\widetilde{\pi}_2 \circ \widetilde{\pi}_1$ and Id_{Y_2} are obtained from π_2 by passing to the quotient, so the uniqueness assertion of Theorem A.30 implies that $\widetilde{\pi}_2 \circ \widetilde{\pi}_1 = \mathrm{Id}_{Y_2}$. A similar argument shows that $\widetilde{\pi}_1 \circ \widetilde{\pi}_2 = \mathrm{Id}_{Y_1}$, so that $\widetilde{\pi}_2$ is the desired homeomorphism. □

Open and Closed Maps

A map $F \colon X \to Y$ (continuous or not) is said to be an ***open map*** if for every open subset $U \subseteq X$, the image set $F(U)$ is open in Y, and a ***closed map*** if for every closed subset $K \subseteq X$, the image $F(K)$ is closed in Y. Continuous maps may be open, closed, both, or neither, as can be seen by examining simple examples involving subsets of the plane.

► **Exercise A.32.** Suppose $X_1, \dots, X_k$ are topological spaces. Show that each projection $\pi_i \colon X_1 \times \dots \times X_k \to X_i$ is an open map.

► **Exercise A.33.** Let $(X_\alpha)_{\alpha \in A}$ be an indexed family of topological spaces. Show that each injection $\iota_\alpha \colon X_\alpha \to \coprod_{\alpha \in A} X_\alpha$ is both open and closed.

► **Exercise A.34.** Show that every local homeomorphism is an open map.

► **Exercise A.35.** Show that every bijective local homeomorphism is a homeomorphism.

► **Exercise A.36.** Suppose $q \colon X \to Y$ is an open quotient map. Prove that Y is Hausdorff if and only if the set $\mathcal{R} = \{(x_1, x_2) : q(x_1) = q(x_2)\}$ is closed in $X \times X$.

► **Exercise A.37.** Let X and Y be topological spaces, and let $F \colon X \to Y$ be a map. Prove the following:

(a) F is closed if and only if for every $A \subseteq X$, $F\big(\overline{A}\big) \supseteq \overline{F(A)}$.
(b) F is open if and only if for every $B \subseteq Y$, $F^{-1}(\operatorname{Int} B) \supseteq \operatorname{Int} F^{-1}(B)$.

The most important classes of continuous maps in topology are the homeomorphisms, quotient maps, and topological embeddings. Obviously, it is necessary for a map to be bijective in order for it to be a homeomorphism, surjective for it to be a quotient map, and injective for it to be a topological embedding. However, even when a continuous map is known to satisfy one of these necessary set-theoretic conditions, it is not always easy to tell whether it has the desired topological property. One simple sufficient condition is that it be either an open or a closed map, as the next theorem shows.

Theorem A.38. *Suppose X and Y are topological spaces, and $F: X \to Y$ is a continuous map that is either open or closed.*

(a) *If F is surjective, then it is a quotient map.*
(b) *If F is injective, then it is a topological embedding.*
(c) *If F is bijective, then it is a homeomorphism.*

Proof. Suppose first that F is surjective. If it is open, it certainly takes saturated open subsets to open subsets. Similarly, if it is closed, it takes saturated closed subsets to closed subsets. Thus it is a quotient map by Exercise A.29.

Now suppose F is open and injective. Then $F: X \to F(X)$ is bijective, so $F^{-1}: F(X) \to X$ exists by elementary set-theoretic considerations. If $U \subseteq X$ is open, then $\left(F^{-1}\right)^{-1}(U) = F(U)$ is open in Y by hypothesis, and therefore is also open in $F(X)$ by definition of the subspace topology on $F(X)$. This proves that F^{-1} is continuous, so that F is a homeomorphism onto its image. If F is closed, the same argument goes through with "open" replaced by "closed" (using the characterization of closed subsets of $F(X)$ given in Proposition A.17(c)). This proves part (b), and part (c) is just the special case of (b) in which $F(X) = Y$. □

Connectedness and Compactness

A topological space X is said to be ***disconnected*** if it has two disjoint nonempty open subsets whose union is X, and it is ***connected*** otherwise. Equivalently, X is connected if and only if the only subsets of X that are both open and closed are $\varnothing$ and X itself. If X is any topological space, a ***connected subset of X*** is a subset that is a connected space when endowed with the subspace topology. For example, the nonempty connected subsets of $\mathbb{R}$ are the ***singletons*** (one-element sets) and the ***intervals***, which are the subsets $J \subseteq \mathbb{R}$ containing more than one point and having the property that whenever $a, b \in J$ and $a < c < b$, it follows that $c \in J$ as well.

A maximal connected subset of X (i.e., a connected subset that is not properly contained in any larger connected subset) is called a ***component*** (or ***connected component***) ***of X***.

Proposition A.39 (Properties of Connected Spaces). *Let X and Y be topological spaces.*

(a) *If $F: X \to Y$ is continuous and X is connected, then $F(X)$ is connected.*
(b) *Every connected subset of X is contained in a single component of X.*

(c) *A union of connected subspaces of* X *with a point in common is connected.*
(d) *The components of* X *are disjoint nonempty closed subsets whose union is* X*, and thus they form a partition of* X*.*
(e) *If* S *is a subset of* X *that is both open and closed, then* S *is a union of components of* X*.*
(f) *Every finite product of connected spaces is connected.*
(g) *Every quotient space of a connected space is connected.*

► **Exercise A.40.** Prove the preceding proposition.

Closely related to connectedness is *path connectedness*. If X is a topological space and $p, q \in X$, a ***path in*** X ***from*** p ***to*** q is a continuous map $f : I \to X$ (where $I = [0, 1]$) such that $f(0) = p$ and $f(1) = q$. If for every pair of points $p, q \in X$ there exists a path in X from p to q, then X is said to be ***path-connected***. The ***path components of*** X are its maximal path-connected subsets.

Proposition A.41 (Properties of Path-Connected Spaces).

(a) *Proposition A.39 holds with "connected" replaced by "path-connected" and "component" by "path component" throughout.*
(b) *Every path-connected space is connected.*

► **Exercise A.42.** Prove the preceding proposition.

For most topological spaces we treat in this book, including all manifolds, connectedness and path connectedness turn out to be equivalent. The link between the two concepts is provided by the following notion. A topological space is said to be ***locally path-connected*** if it admits a basis of path-connected open subsets.

Proposition A.43 (Properties of Locally Path-Connected Spaces). *Let* X *be a locally path-connected topological space.*

(a) *The components of* X *are open in* X*.*
(b) *The path components of* X *are equal to its components.*
(c) X *is connected if and only if it is path-connected.*
(d) *Every open subset of* X *is locally path-connected.*

► **Exercise A.44.** Prove the preceding proposition.

A topological space X is said to be ***compact*** if every open cover of X has a finite subcover. A ***compact subset*** of a topological space is one that is a compact space in the subspace topology. For example, it is a consequence of the Heine–Borel theorem that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proposition A.45 (Properties of Compact Spaces). *Let* X *and* Y *be topological spaces.*

(a) *If* $F : X \to Y$ *is continuous and* X *is compact, then* $F(X)$ *is compact.*
(b) *If* X *is compact and* $f : X \to \mathbb{R}$ *is continuous, then* f *is bounded and attains its maximum and minimum values on* X*.*

(c) *Any union of finitely many compact subspaces of X is compact.*
(d) *If X is Hausdorff and K and L are disjoint compact subsets of X, then there exist disjoint open subsets $U, V \subseteq X$ such that $K \subseteq U$ and $L \subseteq V$.*
(e) *Every closed subset of a compact space is compact.*
(f) *Every compact subset of a Hausdorff space is closed.*
(g) *Every compact subset of a metric space is bounded.*
(h) *Every finite product of compact spaces is compact.*
(i) *Every quotient of a compact space is compact.*

► **Exercise A.46.** Prove the preceding proposition.

For maps between metric spaces, there are several variants of continuity that are useful, especially in the context of compact spaces. Suppose (M_1, d_1) and (M_2, d_2) are metric spaces, and $F\colon M_1 \to M_2$ is a map. Then F is said to be ***uniformly continuous*** if for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in M_1$, $d_1(x, y) < \delta$ implies $d_2\big(F(x), F(y)\big) < \varepsilon$. It is said to be ***Lipschitz continuous*** if there is a constant C such that $d_2\big(F(x), F(y)\big) \le C d_1(x, y)$ for all $x, y \in M_1$. Any such C is called a ***Lipschitz constant for F***. We say that F is ***locally Lipschitz continuous*** if every point $x \in M_1$ has a neighborhood on which F is Lipschitz continuous. (To emphasize the distinction, Lipschitz continuous functions are sometimes called ***uniformly*** or ***globally Lipschitz continuous***.)

► **Exercise A.47.** For maps between metric spaces, show that Lipschitz continuous $\Rightarrow$ uniformly continuous $\Rightarrow$ continuous, and Lipschitz continuous $\Rightarrow$ locally Lipschitz continuous $\Rightarrow$ continuous. (Exercise A.49 below shows that these implications are not reversible.)

Proposition A.48. *Suppose (M_1, d_1) and (M_2, d_2) are metric spaces and $F\colon M_1 \to M_2$ is a map. Let K be any compact subset of M_1.*

(a) *If F is continuous, then $F|_K$ is uniformly continuous.*
(b) *If F is locally Lipschitz continuous, then $F|_K$ is Lipschitz continuous.*

Proof. First we prove (a). Assume F is continuous, and let $\varepsilon > 0$ be given. For each $x \in K$, by continuity there is a positive number $\delta(x)$ such that $d_1(x, y) < 2\delta \Rightarrow d_2\big(F(x), F(y)\big) < \varepsilon/2$. Because the open balls $\{B_{\delta(x)}(x) : x \in K\}$ cover K, by compactness there are finitely many points $x_1, \dots, x_n \in K$ such that $K \subseteq B_{\delta(x_1)}(x_1) \cup \dots \cup B_{\delta(x_n)}(x_n)$. Let $\delta = \min\{\delta(x_1), \dots, \delta(x_n)\}$. Suppose $x, y \in K$ satisfy $d_1(x, y) < \delta$. There is some i such that $x \in B_{\delta(x_i)}(x_i)$, and then the triangle inequality implies that x and y both lie in $B_{2\delta(x_i)}(x_i)$. It follows that

$$d_2\big(F(x), F(y)\big) \le d_2\big(F(x), F(x_i)\big) + d_2\big(F(x_i), F(y)\big) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Next we prove (b). Assume F is locally Lipschitz continuous. Because F is continuous, Proposition A.45 shows that $F(K)$ is compact and therefore bounded. Let $D = \operatorname{diam} F(K)$. For each $x \in K$, there is a positive number $\delta(x)$ such that F is Lipschitz continuous on $B_{2\delta(x)}(x)$, with Lipschitz constant $C(x)$. By compactness, there are points $x_1, \dots, x_n \in K$ such that $K \subseteq B_{\delta(x_1)}(x_1) \cup \dots \cup B_{\delta(x_n)}(x_n)$.

Let $C = \max\{C(x_1), \dots, C(x_n)\}$ and $\delta = \min\{\delta(x_1), \dots, \delta(x_n)\}$, and let $x, y \in K$ be arbitrary. On the one hand, if $d_1(x, y) < \delta$, then by the same argument as in the preceding paragraph, x and y lie in one of the balls on which F is Lipschitz continuous, so $d_2\big(F(x), F(y)\big) \le C d_1(x, y)$. On the other hand, if $d_1(x, y) \ge \delta$, then $d_2\big(F(x), F(y)\big) \le D \le (D/\delta) d_1(x, y)$. Therefore, $\max\{C, D/\delta\}$ is a Lipschitz constant for F on K. □

► **Exercise A.49.** Let $f, g\colon [0, \infty) \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$ and $g(x) = x^2$. Show that f is uniformly continuous but not locally or globally Lipschitz continuous, and g is locally Lipschitz continuous but not uniformly continuous or globally Lipschitz continuous.

For manifolds, subsets of manifolds, and most other spaces we work with, there are two other equivalent formulations of compactness that are frequently useful. Proofs of the next proposition can be found in [LeeTM, Chap. 4], [Mun00, Chap. 3], and [Sie92, Chap. 7].

Proposition A.50 (Equivalent Formulations of Compactness). *Suppose M is a second-countable Hausdorff space or a metric space. The following are equivalent.*

(a) *M is compact.*
(b) *Every infinite subset of M has a limit point in M.*
(c) *Every sequence in M has a convergent subsequence in M.*

► **Exercise A.51.** Show that every compact metric space is complete.

The next lemma expresses one of the most useful properties of compact spaces.

Lemma A.52 (Closed Map Lemma). *Suppose X is a compact space, Y is a Hausdorff space, and $F\colon X \to Y$ is a continuous map.*

(a) *F is a closed map.*
(b) *If F is surjective, it is a quotient map.*
(c) *If F is injective, it is a topological embedding.*
(d) *If F is bijective, it is a homeomorphism.*

Proof. By virtue of Theorem A.38, the last three assertions follow from the first, so we need only prove that F is closed. Suppose $K \subseteq X$ is a closed subset. Then part (e) of Proposition A.45 implies that K is compact; part (a) of that proposition implies that $F(K)$ is compact; and part (f) implies that $F(K)$ is closed in Y. □

If X and Y are topological spaces, a map $F\colon X \to Y$ (continuous or not) is said to be ***proper*** if for every compact set $K \subseteq Y$, the preimage $F^{-1}(K)$ is compact. Here are some useful sufficient conditions for a map to be proper.

Proposition A.53 (Sufficient Conditions for Properness). *Suppose X and Y are topological spaces, and $F\colon X \to Y$ is a continuous map.*

(a) *If X is compact and Y is Hausdorff, then F is proper.*
(b) *If F is a closed map with compact fibers, then F is proper.*
(c) *If F is a topological embedding with closed image, then F is proper.*

(d) *If Y is Hausdorff and F has a continuous left inverse* (*i.e.*, *a continuous map* $G\colon Y \to X$ *such that* $G \circ F = \mathrm{Id}_X$), *then F is proper.*
(e) *If F is proper and $A \subseteq X$ is a subset that is saturated with respect to F, then* $F|_A\colon A \to F(A)$ *is proper.*

► **Exercise A.54.** Prove the preceding proposition.

Locally Compact Hausdorff Spaces

In general, the topological spaces whose properties are most familiar are those whose topologies are induced by metrics; such a topological space is said to be ***metrizable***. However, when studying manifolds, it is often quite inconvenient to exhibit a metric that generates a manifold's topology. Fortunately, as shown in Chapter 1, manifolds belong to another class of spaces with similarly nice properties, the *locally compact Hausdorff spaces*. In this section, we review some of the properties of these spaces.

A topological space X is said to be ***locally compact*** if every point has a neighborhood contained in a compact subset of X. If X is Hausdorff, this property has two equivalent formulations that are often more useful, as the next exercise shows. A subset of X is said to be ***precompact in X*** if its closure in X is compact.

► **Exercise A.55.** For a Hausdorff space X, show that the following are equivalent:

(a) X is locally compact.
(b) Each point of X has a precompact neighborhood.
(c) X has a basis of precompact open subsets.

► **Exercise A.56.** Prove that every open or closed subspace of a locally compact Hausdorff space is itself a locally compact Hausdorff space.

The next result can be viewed as a generalization of the closed map lemma (Lemma A.52).

Theorem A.57 (Proper Continuous Maps Are Closed). *Suppose X is a topological space and Y is a locally compact Hausdorff space. Then every proper continuous map $F\colon X \to Y$ is closed.*

Proof. Let $K \subseteq X$ be a closed subset. To show that $F(K)$ is closed in Y, we show that it contains all of its limit points. Let y be a limit point of $F(K)$, and let U be a precompact neighborhood of y. Then y is also a limit point of $F(K) \cap \bar{U}$. Because F is proper, $F^{-1}(\bar{U})$ is compact, which implies that $K \cap F^{-1}(\bar{U})$ is compact. Because F is continuous, $F\big(K \cap F^{-1}(\bar{U})\big) = F(K) \cap \bar{U}$ is compact and therefore closed in Y. In particular, $y \in F(K) \cap \bar{U} \subseteq F(K)$, so $F(K)$ is closed. □

Here is an important property of locally compact Hausdorff spaces, which is also shared by complete metric spaces. For a proof, see [LeeTM, Chap. 4].

Theorem A.58 (Baire Category Theorem). *In a locally compact Hausdorff space or a complete metric space, every countable union of nowhere dense sets has empty interior.*

Corollary A.59. *In a locally compact Hausdorff space or a complete metric space, every nonempty countable closed subset contains at least one isolated point.*

Proof. Assume X is such a space. Let $A \subseteq X$ be a nonempty countable closed subset, and assume that A has no isolated points. The fact that A is closed in X means that A itself is either a locally compact Hausdorff space or a complete metric space. For each $a \in A$, the singleton $\{a\}$ is nowhere dense in A: it is closed in A because A is Hausdorff, and it contains no nonempty open subset because A has no isolated points. Since A is a countable union of singletons, the Baire category theorem implies that A has empty interior in A, which is a contradiction. □

If we add the hypothesis of second-countability to a locally compact Hausdorff space, we can prove even more. A sequence $(K_i)_{i=1}^{\infty}$ of compact subsets of a topological space X is called an ***exhaustion of X by compact sets*** if $X = \bigcup_i K_i$ and $K_i \subseteq \operatorname{Int} K_{i+1}$ for each i.

Proposition A.60. *A second-countable, locally compact Hausdorff space admits an exhaustion by compact sets.*

Proof. Let X be such a space. Because X is a locally compact Hausdorff space, it has a basis of precompact open subsets; since it is second-countable, it is covered by countably many such sets. Let $(U_i)_{i=1}^{\infty}$ be such a countable cover. Beginning with $K_1 = \bar{U}_1$, assume by induction that we have constructed compact sets $K_1, \dots, K_k$ satisfying $U_j \subseteq K_j$ for each j and $K_{j-1} \subseteq \operatorname{Int} K_j$ for $j \geq 2$. Because K_k is compact, there is some m_k such that $K_k \subseteq U_1 \cup \cdots \cup U_{m_k}$. If we let $K_{k+1} = \bar{U}_1 \cup \cdots \cup \bar{U}_{m_k}$, then K_{k+1} is a compact set whose interior contains K_k. Moreover, by increasing m_k if necessary, we may assume that $m_k \geq k+1$, so that $U_{k+1} \subseteq K_{k+1}$. By induction, we obtain the required exhaustion. □

Homotopy and the Fundamental Group

If X and Y are topological spaces and $F_0, F_1 \colon X \to Y$ are continuous maps, a ***homotopy from F_0 to F_1*** is a continuous map $H \colon X \times I \to Y$ satisfying

$$H(x,0) = F_0(x),$$
$$H(x,1) = F_1(x),$$

for all $x \in X$. If there exists a homotopy from F_0 to F_1, we say that ***F_0 and F_1 are homotopic***, and write $F_0 \simeq F_1$. If the homotopy satisfies $H(x,t) = F_0(x) = F_1(x)$ for all $t \in I$ and all x in some subset $A \subseteq X$, the maps F_0 and F_1 are said to be ***homotopic relative to A***. Both "homotopic" and "homotopic relative to A" are equivalence relations on the set of all continuous maps from X to Y.

The most important application of homotopies is to paths. Suppose X is a topological space. Two paths $f_0, f_1 \colon I \to X$ are said to be ***path-homotopic***, denoted symbolically by $f_0 \sim f_1$, if they are homotopic relative to $\{0,1\}$. Explicitly, this means that there is a continuous map $H\colon I \times I \to X$ satisfying

$$\begin{aligned} H(s,0) &= f_0(s), & s &\in I; \\ H(s,1) &= f_1(s), & s &\in I; \\ H(0,t) &= f_0(0) = f_1(0), & t &\in I; \\ H(1,t) &= f_0(1) = f_1(1), & t &\in I. \end{aligned}$$

For any given points $p, q \in X$, path homotopy is an equivalence relation on the set of all paths from p to q. The equivalence class of a path f is called its ***path class***, and is denoted by $[f]$.

Given two paths $f, g\colon I \to X$ such that $f(1) = g(0)$, their ***product*** is the path $f \cdot g\colon I \to X$ defined by

$$f \cdot g(s) = \begin{cases} f(2s), & 0 \le s \le \frac{1}{2}; \\ g(2s-1), & \frac{1}{2} \le s \le 1. \end{cases}$$

If $f \sim f'$ and $g \sim g'$, it is not hard to show that $f \cdot g \sim f' \cdot g'$. Therefore, it makes sense to define the product of the path classes $[f]$ and $[g]$ by $[f] \cdot [g] = [f \cdot g]$. Although multiplication of paths is not associative, it is associative up to path homotopy: $([f] \cdot [g]) \cdot [h] = [f] \cdot ([g] \cdot [h])$. When we need to consider products of three or more actual paths (as opposed to path classes), we adopt the convention that such products are to be evaluated from left to right: $f \cdot g \cdot h = (f \cdot g) \cdot h$.

If X is a topological space and q is a point in X, a ***loop in X based at q*** is a path in X from q to q, that is, a continuous map $f\colon I \to X$ such that $f(0) = f(1) = q$. The set of path classes of loops based at q is denoted by $\pi_1(X, q)$. Equipped with the product described above, it is a group, called the ***fundamental group of X based at q***. The identity element of this group is the path class of the ***constant path*** $c_q(s) \equiv q$, and the inverse of $[f]$ is the path class of the ***reverse path*** $\bar{f}(s) = f(1-s)$.

It can be shown that for path-connected spaces, the fundamental groups based at different points are isomorphic. If X is path-connected and for some (hence every) $q \in X$, the fundamental group $\pi_1(X, q)$ is the trivial group consisting of $[c_q]$ alone, we say that X is ***simply connected***. This means that every loop is path-homotopic to a constant path.

► **Exercise A.61.** Let X be a path-connected topological space. Show that X is simply connected if and only if every pair of paths in X with the same starting and ending points are path-homotopic.

A key feature of the homotopy relation is that it is preserved by composition, as the next proposition shows.

Proposition A.62. *If $F_0, F_1\colon X \to Y$ and $G_0, G_1\colon Y \to Z$ are continuous maps with $F_0 \simeq F_1$ and $G_0 \simeq G_1$, then $G_0 \circ F_0 \simeq G_1 \circ F_1$. Similarly, if $f_0, f_1\colon I \to X$ are path-homotopic and $F\colon X \to Y$ is a continuous map, then $F \circ f_0 \sim F \circ f_1$.*

► **Exercise A.63.** Prove the preceding proposition.

Thus if $F: X \to Y$ is a continuous map, for each $q \in X$ we obtain a well-defined map $F_*: \pi_1(X,q) \to \pi_1\big(Y, F(q)\big)$ by setting

$$F_*[f] = [F \circ f].$$

Proposition A.64. *If X and Y are topological spaces and $F: X \to Y$ is a continuous map, then $F_*: \pi_1(X,q) \to \pi_1\big(Y, F(q)\big)$ is a group homomorphism, known as the* ***homomorphism induced by F****.*

Proposition A.65 (Properties of the Induced Homomorphism).

(a) *Let $F: X \to Y$ and $G: Y \to Z$ be continuous maps. Then for each $q \in X$, $(G \circ F)_* = G_* \circ F_*: \pi_1(X,q) \to \pi_1\big(Z, G(F(q))\big)$.*
(b) *For each space X and each $q \in X$, the homomorphism induced by the identity map $\mathrm{Id}_X: X \to X$ is the identity map of $\pi_1(X,q)$.*
(c) *If $F: X \to Y$ is a homeomorphism, then $F_*: \pi_1(X,q) \to \pi_1\big(Y, F(q)\big)$ is an isomorphism. Thus, homeomorphic spaces have isomorphic fundamental groups.*

► **Exercise A.66.** Prove the two preceding propositions.

► **Exercise A.67.** A subset $U \subseteq \mathbb{R}^n$ is said to be ***star-shaped*** if there is a point $c \in U$ such that for each $x \in U$, the line segment from c to x is contained in U. Show that every star-shaped set is simply connected.

Proposition A.68 (Fundamental Groups of Spheres).

(a) *$\pi_1\big(\mathbb{S}^1, (1,0)\big)$ is the infinite cyclic group generated by the path class of the loop $\omega: I \to \mathbb{S}^1$ given by $\omega(s) = (\cos 2\pi s, \sin 2\pi s)$.*
(b) *If $n > 1$, $\mathbb{S}^n$ is simply connected.*

Proposition A.69 (Fundamental Groups of Product Spaces). *Suppose $X_1, \dots, X_k$ are topological spaces, and let $p_i: X_1 \times \cdots \times X_k \to X_i$ denote the ith projection map. For any points $q_i \in X_i$, $i = 1, \dots, k$, define a map*

$$P: \pi_1\big(X_1 \times \cdots \times X_k, (q_1, \dots, q_k)\big) \to \pi_1(X_1, q_1) \times \cdots \times \pi_1(X_k, q_k)$$

by

$$P[f] = \big(p_{1*}[f], \dots, p_{k*}[f]\big).$$

Then P is an isomorphism.

► **Exercise A.70.** Prove the two preceding propositions.

A continuous map $F: X \to Y$ between topological spaces is said to be a ***homotopy equivalence*** if there is a continuous map $G: Y \to X$ such that $F \circ G \simeq \mathrm{Id}_Y$ and $G \circ F \simeq \mathrm{Id}_X$. Such a map G is called a ***homotopy inverse for F***. If there exists a homotopy equivalence between X and Y, the two spaces are said to be ***homotopy equivalent***. For example, the inclusion map $\iota: \mathbb{S}^{n-1} \hookrightarrow \mathbb{R}^n \smallsetminus \{0\}$ is a homotopy equivalence with homotopy inverse $r(x) = x/|x|$, because $r \circ \iota = \mathrm{Id}_{\mathbb{S}^{n-1}}$ and $\iota \circ r$ is homotopic to the identity map of $\mathbb{R}^n \smallsetminus \{0\}$ via the straight-line homotopy $H(x,t) = tx + (1-t)x/|x|$.

Theorem A.71 (Homotopy Invariance). *If $F: X \to Y$ is a homotopy equivalence, then for each $p \in X$, $F_*: \pi_1(X, p) \to \pi_1\big(Y, F(p)\big)$ is an isomorphism.*

For a proof, see any of the topology texts mentioned at the beginning of this appendix.

Covering Maps

Suppose E and X are topological spaces. A map $\pi: E \to X$ is called a ***covering map*** if E and X are connected and locally path-connected, π is surjective and continuous, and each point $p \in X$ has a neighborhood U that is ***evenly covered by*** $\boldsymbol{\pi}$, meaning that each component of $\pi^{-1}(U)$ is mapped homeomorphically onto U by π. In this case, X is called the ***base of the covering***, and E is called a ***covering space of*** $\boldsymbol{X}$. If U is an evenly covered subset of X, the components of $\pi^{-1}(U)$ are called the ***sheets of the covering over*** $\boldsymbol{U}$.

Some immediate consequences of the definition should be noted. First, it follows from Proposition A.43 that E and X are actually path-connected. Second, suppose $U \subseteq X$ is any evenly covered open subset. Because $\pi^{-1}(U)$ is open in E, it is locally path-connected, and therefore its components are open subsets of $\pi^{-1}(U)$ and thus also of E. Because U is the homeomorphic image of any one of the components of $\pi^{-1}(U)$, each of which is path-connected, it follows that evenly covered open subsets are path-connected.

► **Exercise A.72.** Show that every covering map is a local homeomorphism, an open map, and a quotient map.

► **Exercise A.73.** Show that an injective covering map is a homeomorphism.

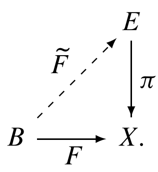
► **Exercise A.74.** Show that all fibers of a covering map have the same cardinality, called the ***number of sheets of the covering***.

► **Exercise A.75.** Show that a covering map is a proper map if and only if it is finite-sheeted.

► **Exercise A.76.** Show that every finite product of covering maps is a covering map.

The main properties of covering maps that we need are summarized in the next four propositions. For proofs, you can consult [LeeTM, Chaps. 11 and 12], [Mun00, Chaps. 9 and 13], or [Sie92, Chap. 14].

If $\pi: E \to X$ is a covering map and $F: B \to X$ is a continuous map, a ***lift of*** $\boldsymbol{F}$ is a continuous map $\widetilde{F}: B \to E$ such that $\pi \circ \widetilde{F} = F$:



Proposition A.77 (Lifting Properties of Covering Maps). *Suppose $\pi\colon E \to X$ is a covering map.*

(a) UNIQUE LIFTING PROPERTY: *If B is a connected space and $F\colon B \to X$ is a continuous map, then any two lifts of F that agree at one point are identical.*
(b) PATH LIFTING PROPERTY: *If $f\colon I \to X$ is a path, then for any point $e \in E$ such that $\pi(e) = f(0)$, there exists a unique lift $\widetilde{f}_e\colon I \to E$ of f such that $\widetilde{f}(0) = e$.*
(c) MONODROMY THEOREM: *If $f, g\colon I \to X$ are path-homotopic paths and $\widetilde{f}_e, \widetilde{g}_e\colon I \to E$ are their lifts starting at the same point $e \in E$, then $\widetilde{f}_e$ and $\widetilde{g}_e$ are path-homotopic and $\widetilde{f}_e(1) = \widetilde{g}_e(1)$.*

Proposition A.78 (Lifting Criterion). *Suppose $\pi\colon E \to X$ is a covering map, Y is a connected and locally path-connected space, and $F\colon Y \to X$ is a continuous map. Let $y \in Y$ and $e \in E$ be such that $\pi(e) = F(y)$. Then there exists a lift $\widetilde{F}\colon Y \to E$ of F satisfying $\widetilde{F}(y) = e$ if and only if $F_*\big(\pi_1(Y, y)\big) \subseteq \pi_*\big(\pi_1(E, e)\big)$.*

Proposition A.79 (Coverings of Simply Connected Spaces). *If X is a simply connected space, then every covering map $\pi\colon E \to X$ is a homeomorphism.*

A topological space is said to be ***locally simply connected*** if it admits a basis of simply connected open subsets.

Proposition A.80 (Existence of a Universal Covering Space). *If X is a connected and locally simply connected topological space, there exists a simply connected topological space $\widetilde{X}$ and a covering map $\pi\colon \widetilde{X} \to X$. If $\widehat{\pi}\colon \widehat{X} \to X$ is any other simply connected covering of X, there is a homeomorphism $\varphi\colon \widetilde{X} \to \widehat{X}$ such that $\widehat{\pi} \circ \varphi = \pi$.*

The simply connected covering space $\widetilde{X}$ whose existence and uniqueness (up to homeomorphism) are guaranteed by this proposition is called the ***universal covering space of X***.

Appendix B
Review of 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 [FIS03]. Here we just summarize the main points, with emphasis on those aspects that are most important for 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 $(v, w) \mapsto v + w$, and ***scalar multiplication*** $\mathbb{R} \times V \to V$, denoted by $(a, v) \mapsto av$, satisfying the following properties:

(i) V is an abelian group under vector addition.
(ii) Scalar multiplication satisfies the following identities:

$$\begin{aligned} a(bv) &= (ab)v && \text{for all } v \in V \text{ and } a, b \in \mathbb{R}; \\ 1v &= v && \text{for all } v \in V. \end{aligned}$$

(iii) Scalar multiplication and vector addition are related by the following distributive laws:

$$\begin{aligned} (a+b)v &= av + bv && \text{for all } v \in V \text{ and } a, b \in \mathbb{R}; \\ a(v+w) &= av + aw && \text{for all } v, w \in V \text{ and } a \in \mathbb{R}. \end{aligned}$$

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*** $\mathbb{F}$. In particular, we sometimes have occasion to consider vector spaces over $\mathbb{C}$, called ***complex vector spaces***. Unless we specify otherwise, all vector spaces are assumed to be real.

Second, if $\mathbb{R}$ is replaced by a commutative ring $\mathcal{R}$, this becomes the definition of a ***module over*** $\mathcal{R}$ (or ***$\mathcal{R}$-module***). For example, if $\mathbb{Z}$ denotes the ring of integers,

J.M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics 218,
DOI 10.1007/978-1-4419-9982-5,

it is straightforward to check that modules over $\mathbb{Z}$ are just abelian groups under addition.

The elements of a vector space are usually called ***vectors***. When it is necessary to distinguish them from vectors, elements of the underlying field (which is $\mathbb{R}$ unless otherwise specified) are called ***scalars***.

Let V be a vector space. A subset $W \subseteq V$ that is closed under vector addition and scalar multiplication is itself a vector space with the same operations, and is called a ***subspace of V***. To avoid confusion with the use of the word "subspace" in topology, we sometimes use the term ***linear subspace*** for a subspace of a vector space in this sense, and ***topological subspace*** for a subset of a topological space endowed with the subspace topology.

A finite sum of the form $\sum_{i=1}^{k} a^i v_i$, where a^i are scalars and $v_i \in V$, is called a ***linear combination of the vectors*** $\boldsymbol{v_1, \dots, v_k}$. (The reason we write the coefficients a^i with superscripts instead of subscripts is to be consistent with the Einstein summation convention, 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 that ***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 $\{0\}$.

If p and q are points of V, the ***line segment from p to q*** is the set $\{(1-t)p + tq : 0 \le t \le 1\}$. A subset $B \subseteq V$ is said to be ***convex*** if for every two points $p, q \in B$, the line segment from p to q is contained in B.

Bases and Dimension

Suppose V is a vector space. A subset $S \subseteq V$ is said to be ***linearly dependent*** if there exists a linear relation of the form $\sum_{i=1}^{k} a^i v_i = 0$, where $v_1, \dots, v_k$ are distinct elements of S and at least one of the coefficients a^i is nonzero; S 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 every set containing the zero vector is linearly dependent. By convention, the empty set is considered to be linearly independent.

It is frequently important to work with ordered k-tuples of vectors in V; such a k-tuple is denoted by $(v_1, \dots, v_k)$ or (v_i), with parentheses instead of braces to distinguish it from the (unordered) set of elements $\{v_1, \dots, v_k\}$. When we consider ordered k-tuples, linear dependence takes on a slightly different meaning. We say that $(v_1, \dots, v_k)$ is a ***linearly dependent k-tuple*** if there are scalars $(a^1, \dots, a^k)$, not all zero, such that $\sum_{i=1}^{k} a^i v_i = 0$; it is a ***linearly independent k-tuple*** otherwise. The only difference between a *linearly independent set* and a *linearly independent k-tuple* is that the latter cannot have repeated vectors. For example if $v \in V$ is a nonzero vector, the ordered pair (v, v) is linearly dependent, while the *set* $\{v, v\} = \{v\}$ is linearly independent. On the other hand, if $(v_1, \dots, v_k)$ is any linearly independent k-tuple, then the set $\{v_1, \dots, v_k\}$ is also linearly independent.

► **Exercise B.1.** Let V be a vector space. Prove the following statements.

(a) If $S \subseteq V$ is linearly independent, then every subset of S is linearly independent.
(b) If $S \subseteq V$ is linearly dependent or spans V, then every subset of V that properly contains S is linearly dependent.
(c) A subset $S \subseteq V$ containing more than one element is linearly dependent if and only if some element $v \in S$ can be expressed as a linear combination of elements of $S \smallsetminus \{v\}$.
(d) If $(v_1, \dots, v_k)$ is a linearly dependent k-tuple in V with $v_1 \neq 0$, then some v_i can be expressed as a linear combination of the *preceding* vectors $(v_1, \dots, v_{i-1})$.

A ***basis for V*** (plural: ***bases***) is a subset $S \subseteq V$ 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 $\{0\}$ is finite-dimensional, because it has the empty set as a basis.

If V is finite-dimensional, an ***ordered basis for V*** is a basis endowed with a specific ordering of the basis vectors, or equivalently a linearly independent n-tuple (E_i) that spans V. For most purposes, ordered bases are more useful than unordered bases, so we always assume, often without comment, that each basis comes with a given ordering.

If $(E_1, \dots, E_n)$ is an (ordered) basis for V, each vector $v \in V$ has a unique expression as a linear combination of basis vectors:

$$v = \sum_{i=1}^{n} v^i E_i.$$

The numbers v^i are called the ***components of v*** with respect to this basis, and the ordered n-tuple $\left(v^1, \dots, v^n\right)$ is called its ***basis representation***. (Here is an example of a definition that requires an ordered basis.)

Lemma B.2. *Let V be a vector space. If V is spanned by a set of n vectors, then every subset of V containing more than n vectors is linearly dependent.*

Proof. Suppose $\{v_1, \dots, v_n\}$ is an n-element set that spans V. To prove the lemma, it suffices to show that every set containing exactly $n+1$ vectors is linearly dependent. Let $\{w_1, \dots, w_{n+1}\}$ be such a set. If any of the w_i's is zero, then clearly the set is dependent, so we might as well assume they are all nonzero. By Exercise B.1(b), the set $\{w_1, v_1, \dots, v_n\}$ is linearly dependent, and thus so is the ordered $(n+1)$-tuple $(w_1, v_1, \dots, v_n)$. By Exercise B.1(d), one of the vectors v_j can be written as a linear combination of $\{w_1, v_1, \dots, v_{j-1}\}$, and thus the set $\{w_1, v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$ still spans V. Renumbering the v_i's if necessary, we may assume that the set $\{w_1, v_2, \dots, v_n\}$ spans V.

Now suppose by induction that $\{w_1, w_2, \dots, w_{k-1}, v_k, \dots, v_n\}$ spans V. As before, the $(n+1)$-tuple $(w_1, w_2, \dots, w_{k-1}, w_k, v_k, \dots, v_n)$ is linearly dependent, so one of the vectors in this list can be written as a linear combination of the preceding ones. If one of the w_i's can be so written, then the set $\{w_1, \dots, w_{n+1}\}$ is dependent and we are done. Otherwise, one of the v_j's can be so written, and after reordering we may assume that $\{w_1, w_2, \dots, w_k, v_{k+1}, \dots, v_n\}$ still spans V. Continuing by induction, by the time we get to $k = n$, if we have not already shown that the w_i's are dependent, we conclude that the set $\{w_1, \dots, w_n\}$ spans V. But this means that the set $\{w_1, \dots, w_{n+1}\}$ is linearly dependent by Exercise B.1(b). □

Proposition B.3. *If V is a finite-dimensional vector space, all bases for V contain the same number of elements.*

Proof. If $\{E_1, \dots, E_n\}$ is a basis for V with n elements, then Lemma B.2 implies that every set containing more than n elements is linearly dependent, so no basis can have more than n elements. On the other hand, if there were a basis containing fewer than n elements, then Lemma B.2 would imply that $\{E_1, \dots, E_n\}$ is linearly dependent, which is a contradiction. □

Because of the preceding proposition, if V is a finite-dimensional vector space, it makes sense to define the ***dimension of V***, denoted by $\dim V$, to be the number of elements in any basis.

► **Exercise B.4.** 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) Show that every subspace $S \subseteq V$ is finite-dimensional and satisfies $\dim S \leq \dim V$, with equality if and only if $S = V$.
(c) Show that $\dim V = 0$ if and only if $V = \{0\}$.

► **Exercise B.5.** Suppose V is an infinite-dimensional vector space.

(a) Use Zorn's lemma to show that every linearly independent subset of V is contained in a basis.
(b) Show that any two bases for V have the same cardinality. [Hint: assume that S and T are bases such that S has larger cardinality than T. Each element of T can be expressed as a linear combination of elements of S, and the hypothesis guarantees that some element of S does not appear in any of the expressions for elements of T. Show that this element can be expressed as a linear combination of other elements of S, contradicting the hypothesis that S is linearly independent.]

If S is a subspace of a finite-dimensional vector space V, we define the ***codimension of S in V*** to be $\dim V - \dim S$. By virtue of Exercise B.4(b), the codimension of S is always nonnegative, and is zero if and only if $S = V$. A **(*linear*) *hyperplane*** is a linear subspace of codimension 1.

Example B.6 (Euclidean Spaces). For each integer $n \geq 0$, $\mathbb{R}^n$ is a real vector space under the usual operations of vector addition and scalar multiplication:

$$\begin{aligned}\left(x^1, \dots, x^n\right) + \left(y^1, \dots, y^n\right) &= \left(x^1 + y^1, \dots, x^n + y^n\right),\\ a\left(x^1, \dots, x^n\right) &= \left(ax^1, \dots, ax^n\right).\end{aligned}$$

There is a natural basis $(e_1, \dots, e_n)$ for $\mathbb{R}^n$, called the ***standard basis***, where $e_i = (0, \dots, 1, \dots, 0)$ is the vector with a 1 in the ith place and zeros elsewhere; thus

$\mathbb{R}^n$ has dimension n, as one would expect. Any element $x \in \mathbb{R}^n$ can be written $\left(x^1,\dots,x^n\right) = \sum_{i=1}^n x^i e_i$, so its components with respect to the standard basis are just its coordinates $\left(x^1,\dots,x^n\right)$. //

Example B.7 (Complex Euclidean Spaces). With scalar multiplication and vector addition defined just as in the real case, the n-dimensional complex Euclidean space $\mathbb{C}^n$ becomes a complex vector space. Because the vectors $(e_1,\dots,e_n)$, defined as above, form a basis for $\mathbb{C}^n$ over $\mathbb{C}$, it follows that $\mathbb{C}^n$ has dimension n as a complex vector space.

By restricting scalar multiplication to real scalars, we can also consider $\mathbb{C}^n$ as a *real vector space*. In this case, it is straightforward to check that the vectors $(e_1, ie_1,\dots,e_n, ie_n)$ form a basis for $\mathbb{C}^n$ over $\mathbb{R}$, so $\mathbb{C}^n$ has dimension $2n$ when considered as a real vector space. //

If S and T are subspaces of a vector space V, the notation $S+T$ denotes the set of all vectors of the form $v+w$, where $v \in S$ and $w \in T$. It is easily seen to be a subspace of V, and in fact is the subspace spanned by $S \cup T$. If $S+T=V$ and $S \cap T = \{0\}$, then V is said to be the ***(internal) direct sum of S and T***, and we write $V = S \oplus T$. Two linear subspaces $S,T \subseteq V$ are said to be ***complementary subspaces*** if $V = S \oplus T$. In this case, every vector in V has a *unique* expression as a sum of an element of S plus an element of T.

► **Exercise B.8.** 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) Suppose $V = S+T$. Show that $V = S \oplus T$ if and only if $\dim V = \dim S + \dim T$.

► **Exercise B.9.** Let V be a finite-dimensional vector space. Show that every subspace $S \subseteq V$ has a complementary subspace in V. In fact, given an arbitrary basis $(E_1,\dots,E_n)$ 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 B.1(d) to the ordered $(m+n)$-tuple $(F_1,\dots,F_m,E_1,\dots,E_n)$.]

Suppose $S \subseteq V$ is a linear subspace. Any subset of V of the form

$$v + S = \{v + w : w \in S\}$$

for some fixed $v \in V$ is called an ***affine subspace of V parallel to S***. If S is a linear hyperplane, then any affine subspace parallel to S is called an ***affine hyperplane***.

► **Exercise B.10.** Let V be a vector space, and let $v+S$ be an affine subspace of V parallel to S.

(a) Show that $v+S$ is a linear subspace if an only if it contains 0, which is true if and only if $v \in S$.
(b) Show that $v + S = \widetilde{v} + \widetilde{S}$ if and only if $S = \widetilde{S}$ and $v - \widetilde{v} \in S$.

Because of part (b) of the preceding exercise, we can unambiguously define the ***dimension of $v + S$*** to be the dimension of S.

For each vector $v \in V$, the affine subspace $v + S$ is also called the ***coset of S determined by v***. The set V/S of cosets of S is called the ***quotient of V by S***.

► **Exercise B.11.** Suppose V is a vector space and S is a linear subspace of V. Define vector addition and scalar multiplication of cosets by

$$(v + S) + (w + S) = (v + w) + S,$$
$$c(v + S) = (cv) + S.$$

(a) Show that the quotient V/S is a vector space under these operations.
(b) Show that if V is finite-dimensional, then $\dim V/S = \dim V - \dim S$.

Linear Maps

Let V and W be real vector spaces. A map $T: V \to W$ is ***linear*** if $T(av + bw) = aTv + bTw$ for all vectors $v, w \in V$ and all scalars a, b. (Because of the close connection between linear maps and matrix multiplication described below, we generally write the action of a linear map T on a vector v as Tv without parentheses, unless parentheses are needed for grouping.) In the special case $W = \mathbb{R}$, a linear map from V to $\mathbb{R}$ is usually called a ***linear functional on V***.

If $T: V \to W$ is a linear map, the ***kernel*** or ***null space of T***, denoted by $\operatorname{Ker} T$ or $T^{-1}(0)$, is the set $\{v \in V : Tv = 0\}$, and the ***image of T***, denoted by $\operatorname{Im} T$ or $T(V)$, is the set $\{w \in W : w = Tv \text{ for some } v \in V\}$.

One simple but important example of a linear map arises in the following way. Given a subspace $S \subseteq V$ and a complementary subspace T, there is a unique linear map $\pi: V \to S$ defined by

$$\pi(v + w) = v \quad \text{for } v \in S,\ w \in T.$$

This map is called the ***projection onto S with kernel T***.

If V and W are vector spaces, a bijective linear map $T: V \to W$ is called an ***isomorphism***. In this case, there is a unique inverse map $T^{-1}: W \to V$, and the following computation shows that T^{-1} is also linear:

$$\begin{aligned} aT^{-1}v + bT^{-1}w &= T^{-1}T\left(aT^{-1}v + bT^{-1}w\right) \\ &= T^{-1}\left(aTT^{-1}v + bTT^{-1}w\right) \quad \text{(by linearity of } T\text{)} \\ &= T^{-1}(av + bw). \end{aligned}$$

For this reason, a bijective linear map is also said to be ***invertible***. If there exists an isomorphism $T: V \to W$, then V and W are said to be ***isomorphic***.

Example B.12. Let V be an n-dimensional real vector space, and $(E_1, \dots, E_n)$ be an ordered basis for V. Define a map $E: \mathbb{R}^n \to V$ by

$$E\left(x^1, \dots, x^n\right) = x^1 E_1 + \dots + x^n E_n.$$

Then E is linear and bijective, so it is an isomorphism, called the ***basis isomorphism*** determined by this basis. Thus, every n-dimensional real vector space is isomorphic to $\mathbb{R}^n$. //

► **Exercise B.13.** Let V and W be vector spaces, and let $(E_1, \dots, E_n)$ be a basis for V. For any n elements $w_1, \dots, w_n \in W$, show that there is a unique linear map $T: V \to W$ satisfying $T(E_i) = w_i$ for $i = 1, \dots, n$.

► **Exercise B.14.** Let $S: V \to W$ and $T: W \to X$ 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) Show that if S is an isomorphism, then $\dim V = \dim W$ (in the sense that these dimensions are either both infinite or both finite and equal).
(d) Show that if S and T are both injective or both surjective, then $T \circ S$ has the same property.
(e) Show that if $T \circ S$ is surjective, then T is surjective; give an example to show that S might not be.
(f) Show that if $T \circ S$ is injective, then S is injective; give an example to show that T might not be.

► **Exercise B.15.** Suppose V is a vector space and S is a subspace of V, and let $\pi: V \to V/S$ denote the projection defined by $\pi(v) = v + S$.

(a) Show that π is a surjective linear map with kernel equal to S.
(b) Given a linear map $T: V \to W$, show that there exists a linear map $\widetilde{T}: V/S \to W$ such that $\widetilde{T} \circ \pi = T$ if and only if $S \subseteq \operatorname{Ker} T$.

If V and W are vector spaces, a map $F: V \to W$ is called an ***affine map*** if it can be written in the form $F(v) = w + Tv$ for some linear map $T: V \to W$ and some fixed $w \in W$.

► **Exercise B.16.** Suppose $F: V \to W$ is an affine map. Show that $F(V)$ is an affine subspace of W, and the sets $F^{-1}(z)$ for $z \in W$ are parallel affine subspaces of V.

► **Exercise B.17.** Suppose V is a finite-dimensional vector space. Show that every affine subspace of V is of the form $F^{-1}(z)$ for some affine map $F: V \to W$ and some $z \in W$.

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 $T: V \to W$ is a linear map, the ***matrix of T*** with respect to these bases is the $m \times n$ matrix

$$A = \left(A^i_j\right) = \begin{pmatrix} A^1_1 & \dots & A^1_n \\ \vdots & \ddots & \vdots \\ A^m_1 & \dots & A^m_n \end{pmatrix}$$

whose jth column consists of the components of TE_j with respect to the basis (F_i):

$$TE_j = \sum_{i=1}^{m} A^i_j F_i.$$

By linearity, the action of T on an arbitrary vector $v = \sum_j v^j E_j$ is then given by

$$T\left(\sum_{j=1}^{n} v^j E_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_j^i v^j F_i.$$

If we write the components of a vector with respect to a basis as a column matrix, then the matrix representation of $w = Tv$ is given by matrix multiplication:

$$\begin{pmatrix} w^1 \\ \vdots \\ w^m \end{pmatrix} = \begin{pmatrix} A_1^1 & \dots & A_n^1 \\ \vdots & \ddots & \vdots \\ A_1^m & \dots & A_n^m \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix},$$

or, more succinctly,

$$w^i = \sum_{j=1}^{n} A_j^i v^j.$$

Insofar as possible, we denote the row index of a matrix by a superscript and the column index by a subscript, so that A_j^i represents the element in the ith row and jth column. Thus the entry in the ith row and jth column of a matrix product AB is given by

$$(AB)_j^i = \sum_{k=1}^{n} A_k^i B_j^k.$$

The composition of two linear maps is represented by the product of their matrices. Provided we use the same basis for both the domain and the codomain, the identity map on an n-dimensional vector space is represented by the ***n* × *n* identity matrix**, which we denote by I_n; it is the matrix with ones on the main diagonal (where the row number equals the column number) 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 $2mn$. When $m = n$, we abbreviate the spaces of $n \times n$ square real and complex matrices by $\mathrm{M}(n, \mathbb{R})$ and $\mathrm{M}(n, \mathbb{C})$, respectively. In this case, matrix multiplication gives these spaces additional algebraic structure. If V, W, and Z are vector spaces, a map $B\colon V \times W \to Z$ is said to be ***bilinear*** if it is linear in each variable separately when the other is held fixed:

$$B(a_1 v_1 + a_2 v_2, w) = a_1 B(v_1, w) + a_2 B(v_2, w),$$
$$B(v, a_1 w_1 + a_2 w_2) = a_1 B(v, w_1) + a_2 B(v, w_2).$$

An ***algebra*** (over $\mathbb{R}$) is a real vector space V endowed with a bilinear product map $V \times V \to V$. The algebra is said to be ***commutative*** or ***associative*** if the bilinear product has that property.

► **Exercise B.18.** Show that matrix multiplication turns both $\mathrm{M}(n,\mathbb{R})$ and $\mathrm{M}(n,\mathbb{C})$ into associative algebras over $\mathbb{R}$. Show that they are noncommutative unless $n = 1$.

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 B.19.** Suppose A is an $n \times n$ matrix. Prove the following statements.

(a) If A is nonsingular, then there is a *unique* $n \times n$ 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 a linear map $T\colon V \to W$ with respect to some bases for V and W, then T is invertible if and only if A is invertible, in which case A^{-1} is the matrix of T^{-1} with respect to the same bases.
(c) If B is an $n \times n$ matrix such that either $AB = I_n$ or $BA = I_n$, then A is nonsingular and $B = A^{-1}$.

Because $\mathbb{R}^n$ comes equipped with the standard basis (e_i), we can unambiguously identify linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ with $m \times n$ real matrices, and we often do so without further comment.

Change of Basis

In this book we often need to be concerned with how various objects transform when we change bases. Suppose (E_i) and $\big(\widetilde{E}_j\big)$ are two bases for a finite-dimensional real 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

$$E_i = \sum_{j=1}^{n} B_i^j \widetilde{E}_j, \qquad \widetilde{E}_j = \sum_{i=1}^{n} \left(B^{-1}\right)_j^i E_i. \tag{B.1}$$

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 the domain V and (F_j) for the codomain W, the map T is represented by some matrix $A = \big(A_j^i\big)$. If $\big(\widetilde{E}_i\big)$ and $\big(\widetilde{F}_j\big)$ are any other choices of bases for V and W, respectively, let B and C denote the transition matrices satisfying (B.1) and

$$F_i = \sum_{j=1}^{m} C_i^j \widetilde{F}_j, \qquad \widetilde{F}_j = \sum_{i=1}^{m} \left(C^{-1}\right)_j^i F_i.$$

Then a straightforward computation shows that the matrix $\widetilde{A}$ representing T with respect to the new bases is related to A by

$$\widetilde{A}_j^i = \sum_{k,l} C_l^i A_k^l \left(B^{-1}\right)_j^k,$$

or, in matrix notation,

$$\widetilde{A} = CAB^{-1}.$$

In particular, if T is a map from V to itself, we usually use the same basis for the domain and the codomain. In this case, if A denotes the matrix of T with respect to (E_i), and $\widetilde{A}$ is its matrix with respect to $\left(\widetilde{E}_i\right)$, we have

$$\widetilde{A} = BAB^{-1}. \tag{B.2}$$

If V and W are real vector spaces, the set $\mathrm{L}(V;W)$ of linear maps from V to W is a real vector space under the operations

$$(S+T)v = Sv + Tv; \qquad (cT)v = c(Tv).$$

If $\dim V = n$ and $\dim W = m$, then each choice of bases for V and W gives us a map $\mathrm{L}(V;W) \to \mathrm{M}(m \times n, \mathbb{R})$, by sending every linear map to its matrix with respect to the chosen bases. This map is easily seen to be linear and bijective, so $\dim \mathrm{L}(V;W) = \dim \mathrm{M}(m \times n, \mathbb{R}) = mn$.

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***. 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 codomain.

Theorem B.20 (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 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^j's are zero because the K_j's are linearly independent. Therefore, the $(r+k)$-tuple $(E_1, \dots, E_r, K_1, \dots, K_k)$ is linearly independent.

To show that these vectors span V, let $v \in V$ be arbitrary. We can express $Tv \in \operatorname{Im} T$ as a linear combination of $(F_1, \dots, F_r)$:

$$Tv = \sum_{i=1}^r c^i F_i.$$

If we put $w = \sum_i c^i E_i \in V$, it follows that $Tw = Tv$, so $z = v - w \in \operatorname{Ker} T$. Writing $z = \sum_j d^j K_j$, we obtain

$$v = w + z = \sum_{i=1}^r c^i E_i + \sum_{j=1}^k d^j K_j,$$

so the $(r+k)$-tuple $(E_1, \dots, E_r, K_1, \dots, K_k)$ does indeed span V. □

This theorem says that every linear map can be put into a particularly nice diagonal form by appropriate choices of bases in the domain and codomain. However, it is important to be aware of what the theorem does *not* say: if $T: V \to V$ is a linear map from a finite-dimensional vector space to itself, it might not be possible to represent T by a diagonal matrix with respect to the same basis for the domain and codomain.

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 B.21 (Rank-Nullity Law). *Suppose $T: V \to W$ is a linear map between finite-dimensional vector spaces. Then*

$$\dim V = \operatorname{rank} T + \operatorname{nullity} T = \dim(\operatorname{Im} T) + \dim(\operatorname{Ker} T).$$

Proof. The preceding proof showed that V has a basis consisting of $k + r$ elements, where $k = \dim(\operatorname{Ker} T)$ and $r = \dim(\operatorname{Im} T)$. □

► **Exercise B.22.** Suppose V, W, X are finite-dimensional vector spaces, and $S: V \to W$ and $T: W \to X$ are linear maps. Prove the following statements.

(a) $\operatorname{rank} S \le \dim V$, with equality if and only if S is injective.
(b) $\operatorname{rank} S \le \dim W$, with equality if and only if S is surjective.
(c) If $\dim V = \dim W$ and S is either injective or surjective, then it is an isomorphism.
(d) $\operatorname{rank}(T \circ S) \le \operatorname{rank} S$, with equality if and only if $\operatorname{Im} S \cap \operatorname{Ker} T = \{0\}$.
(e) $\operatorname{rank}(T \circ S) \le \operatorname{rank} T$, with equality if and only if $\operatorname{Im} S + \operatorname{Ker} T = W$.
(f) If S is an isomorphism, then $\operatorname{rank}(T \circ S) = \operatorname{rank} T$.
(g) If T is an isomorphism, then $\operatorname{rank}(T \circ S) = \operatorname{rank} S$.

Let A be 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: $\left(A^T\right)_i^j = A_j^i$. A square matrix A is said to be ***symmetric*** if $A = A^T$ and ***skew-symmetric*** if $A = -A^T$.

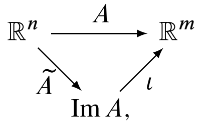
► **Exercise B.23.** Show that if A and B are matrices of dimensions $m \times n$ and $n \times k$, respectively, then $(AB)^T = B^T A^T$.

The ***rank*** of an $m \times n$ matrix A is defined to be the rank of the corresponding 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 B.24. *The row rank of a 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:



where $\widetilde{A}$ is just the map A with its codomain 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 standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$ and the chosen basis in $\operatorname{Im} A$. Taking transposes, we obtain $A^T = C^T B^T$, from which it follows that $\operatorname{rank} A^T \leq \operatorname{rank} B^T$. Since B^T is a $k \times m$ matrix, its column rank is at most k, which shows that $\operatorname{rank} A^T \leq \operatorname{rank} A$. Reversing the roles of A and A^T and using the fact that $(A^T)^T = A$, we conclude that $\operatorname{rank} A = \operatorname{rank} A^T$. □

Suppose $A = \left(A^i_j\right)$ is an $m \times n$ matrix. If we choose integers $1 \leq i_1 < \cdots < i_k \leq m$ and $1 \leq j_1 < \cdots < j_l \leq n$, we obtain a $k \times l$ matrix whose entry in the pth row and qth column is $A^{i_p}_{j_q}$:

$$\begin{pmatrix} A^{i_1}_{j_1} & \dots & A^{i_1}_{j_l} \\ \vdots & \ddots & \vdots \\ A^{i_k}_{j_1} & \dots & A^{i_k}_{j_l} \end{pmatrix}.$$

Such a matrix is called a ***submatrix of A***. Looking at submatrices gives a convenient criterion for checking the rank of a matrix.

Proposition B.25. *Suppose A is an $m \times n$ matrix. Then* $\operatorname{rank} A \geq k$ *if and only if some $k \times k$ submatrix of A is nonsingular.*

Proof. By definition, $\operatorname{rank} A \geq k$ if and only if A has at least k linearly independent columns, which is equivalent to A having some $m \times k$ submatrix with rank k. But by Proposition B.24, an $m \times k$ submatrix has rank k if and only if it has k linearly independent rows. Thus A has rank at least k if and only if it has an $m \times k$ submatrix with k linearly independent rows, if and only if it has a $k \times k$ submatrix that is nonsingular. □

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 give here, while not particularly intuitive, is the simplest to state and fits nicely with our treatment of alternating tensors in Chapter 14.

If X is a set, a ***permutation of X*** is a bijective map from X to itself. The set of all permutations of X is a group under composition. A ***transposition*** is a permutation that interchanges two elements and leaves all the others fixed.

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 need are summarized in the following proposition; proofs can be found in any good undergraduate algebra text such as [Hun97] or [Her75].

Proposition B.26 (Properties of the Symmetric Group).

(a) *Every element of S_n can be expressed as a composition of finitely many transpositions.*
(b) *For each $\sigma \in S_n$, the parity (evenness or oddness) of the number of factors in any decomposition of σ as a product of transpositions is independent of the choice of decomposition. We say that σ is an **even permutation** if every such decomposition has an even number of factors, and an **odd permutation** otherwise.*
(c) *For each $\sigma \in S_n$, define the **sign of σ** to be the number*

$$\operatorname{sgn}\sigma = \begin{cases} +1 & \text{if } \sigma \text{ is even}, \\ -1 & \text{if } \sigma \text{ is odd}. \end{cases}$$

If $n \geq 2$, $\operatorname{sgn}\colon S_n \to \{\pm 1\}$ is a surjective group homomorphism, where we consider $\{\pm 1\}$ as a group under multiplication.

► **Exercise B.27.** Prove (or look up) Proposition B.26.

If $A = \left(A^i_j\right)$ is an $n \times n$ (real or complex) matrix, the ***determinant of A*** is defined by the expression

$$\det A = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) A_1^{\sigma(1)} \cdots A_n^{\sigma(n)}. \tag{B.3}$$

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 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 (B.3) 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 B.28 (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*:

$$\det(A_1, \dots, A_q, \dots, A_p, \dots, A_n) = -\det(A_1, \dots, A_p, \dots, A_q, \dots, A_n). \quad \text{(B.4)}$$

(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 every scalar c*, $\det(cA) = c^n \det A$.
(e) *If any two columns of A are identical*, *then* $\det A = 0$.
(f) *If A has a column of zeros*, *then* $\det A = 0$.
(g) $\det A^T = \det A$.
(h) $\det I_n = 1$.
(i) *If A is singular, then* $\det A = 0$.

Proof. Part (a) is part of the definition of multilinearity, and (d) follows immediately from (a). Part (f) also follows from (a), because a matrix with a column of zeros is unchanged when that column is multiplied by zero, so $\det A = 0(\det A) = 0$. 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 (B.4) is equal to

$$\begin{aligned}
&\det(A_1, \dots, A_q, \dots, A_p, \dots, A_n) \\
&\quad = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) A_1^{\sigma(1)} \cdots A_q^{\sigma(p)} \cdots A_p^{\sigma(q)} \cdots A_n^{\sigma(n)} \\
&\quad = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) A_1^{\sigma(1)} \cdots A_p^{\sigma(q)} \cdots A_q^{\sigma(p)} \cdots A_n^{\sigma(n)} \\
&\quad = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) A_1^{\sigma(\tau(1))} \cdots A_n^{\sigma(\tau(n))} \\
&\quad = -\sum_{\sigma \in S_n} \big(\operatorname{sgn}(\sigma\tau)\big) A_1^{\sigma(\tau(1))} \cdots A_n^{\sigma(\tau(n))} \\
&\quad = -\sum_{\eta \in S_n} (\operatorname{sgn} \eta) A_1^{\eta(1)} \cdots A_n^{\eta(n)} \\
&\quad = -\det(A_1, \dots, A_p, \dots, A_q, \dots, A_n),
\end{aligned}$$

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 (g) follows directly from the definition of the determinant:

$$\begin{aligned}\det A^T &= \sum_{\sigma\in S_n}(\operatorname{sgn}\sigma)A^1_{\sigma(1)}\cdots A^n_{\sigma(n)}\\ &= \sum_{\sigma\in S_n}(\operatorname{sgn}\sigma)A^{\sigma^{-1}(\sigma(1))}_{\sigma(1)}\cdots A^{\sigma^{-1}(\sigma(n))}_{\sigma(n)}\\ &= \sum_{\sigma\in S_n}(\operatorname{sgn}\sigma)A_1^{\sigma^{-1}(1)}\cdots A_n^{\sigma^{-1}(n)}\\ &= \sum_{\eta\in S_n}(\operatorname{sgn}\eta)A_1^{\eta(1)}\cdots A_n^{\eta(n)} = \det A.\end{aligned}$$

In the third line we have used the fact that multiplication is commutative, and the numbers $\{A^{\sigma^{-1}(\sigma(1))}_{\sigma(1)},\dots,A^{\sigma^{-1}(\sigma(n))}_{\sigma(n)}\}$ are just $\{A_1^{\sigma^{-1}(1)},\dots,A_n^{\sigma^{-1}(n)}\}$ in a different order; and the fourth line follows by substituting $\eta=\sigma^{-1}$ and noting that $\operatorname{sgn}\sigma^{-1}=\operatorname{sgn}\sigma$. Similarly, (h) follows from the definition, because when A is the identity matrix, for each σ except the identity permutation there is some j such that $A_j^{\sigma(j)}=0$.

Finally, to prove (i), suppose A is singular. Then, as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$, A has rank less than n by parts (a) and (b) of Exercise B.22. Thus the columns of A are linearly dependent, so at least one column can be written as a linear combination of the others: $A_j=\sum_{i\neq j}c^iA_i$. The result then follows from (e) and the multilinearity of det. □

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.

Since the columns of an $n\times n$ matrix A are the images of the standard basis vectors under the linear map from $\mathbb{R}^n$ to itself that A defines, elementary column operations correspond to changes of basis in the domain. Thus each elementary column operation on a matrix A can be realized by multiplying A on the right by a suitable 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:

$$\begin{pmatrix}A_1^1 & \dots & A_i^1 & \dots & A_n^1\\ \vdots & & \vdots & & \vdots\\ A_1^j & \dots & A_i^j & \dots & A_n^j\\ \vdots & & \vdots & & \vdots\\ A_1^n & \dots & A_i^n & \dots & A_n^n\end{pmatrix}\begin{pmatrix}1 & \dots & 0 & \dots & 0\\ & \ddots & & & \\ & & c & & \\ & & & \ddots & \\ 0 & \dots & 0 & \dots & 1\end{pmatrix}=\begin{pmatrix}A_1^1 & \dots & cA_i^1 & \dots & A_n^1\\ \vdots & & \vdots & & \vdots\\ A_1^j & \dots & cA_i^j & \dots & A_n^j\\ \vdots & & \vdots & & \vdots\\ A_1^n & \dots & cA_i^n & \dots & A_n^n\end{pmatrix}.$$

► **Exercise B.29.** Show that replacing one column of a matrix by c times that same column is equivalent to multiplying on the right by a matrix whose determinant is c; interchanging two columns 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. Matrices of these three types are called ***elementary matrices***.

► **Exercise B.30.** 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 B.31. *If A and B are $n \times n$ matrices, then*

$$\det(AB) = (\det A)(\det B).$$

Proof. If B is singular, then rank $B < n$, which implies that rank $AB < n$. Therefore both $\det B$ and $\det AB$ are zero by Proposition B.28(i). On the other hand, parts (a), (b), and (c) of Proposition B.28 combined with Exercise B.29 show that the theorem is true when B is an elementary matrix. Finally, if B is an arbitrary nonsingular matrix, then B can be written as a product of elementary matrices by Exercise B.30, and the result follows by induction on the number of elementary matrices in such a product. □

Corollary B.32. *If A is a nonsingular $n \times n$ matrix, then* $\det A \neq 0$ *and* $\det\left(A^{-1}\right) = (\det A)^{-1}$.

Proof. Just note that $1 = \det I_n = \det\left(AA^{-1}\right) = (\det A)\left(\det A^{-1}\right)$. □

Corollary B.33. *A square matrix is singular if and only if its determinant is zero.*

Proof. One direction follows from Proposition B.28(i), and the other from Corollary B.32. □

Corollary B.34. *Suppose A and B are $n \times n$ matrices and B is nonsingular. Then* $\det\left(BAB^{-1}\right) = \det A$.

Proof. This is just a computation using Theorem B.31 and Corollary B.32:

$$\begin{aligned}\det\left(BAB^{-1}\right) &= (\det B)(\det A)\left(\det B^{-1}\right)\\ &= (\det B)(\det A)(\det B)^{-1}\\ &= \det A.\end{aligned}$$
□

The last corollary allows us to extend the definition of the determinant to linear maps on arbitrary finite-dimensional vector spaces. Suppose V is an n-dimensional vector space and $T: V \to V$ is a linear map. With respect to a choice of basis for V, T is represented by an $n \times n$ matrix. As we observed above, the matrices

A and $\widetilde{A}$ representing T with respect to two different bases are related by $\widetilde{A} = BAB^{-1}$ for some nonsingular matrix B (see (B.2)). It follows from Corollary B.34, therefore, that $\det \widetilde{A} = \det A$. Thus, we can make the following definition: for each linear map $T: V \to V$ from a finite-dimensional vector space to itself, we define the ***determinant of T*** to be the determinant of any matrix representation of T (using the same basis for the domain and codomain).

For actual computations of determinants, the formula in the following proposition is usually more useful than the definition.

Proposition B.35 (Expansion by Minors). *Let A be an $n \times n$ matrix, and for each i, j let M_i^j denote the $(n-1) \times (n-1)$ submatrix obtained by deleting the ith column and jth row of A. For any fixed i between 1 and n inclusive,*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} A_i^j \det M_i^j. \tag{B.5}$$

Proof. It is useful to consider first a special case: suppose A is an $n \times n$ matrix that has the block form

$$A = \begin{pmatrix} B & 0 \\ C & 1 \end{pmatrix}, \tag{B.6}$$

where B is an $(n-1) \times (n-1)$ matrix and C is a $1 \times (n-1)$ row matrix. Then in the defining formula (B.3) 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 $i \in \{1, \dots, n\}$. 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,

$$\begin{aligned} \det A &= \det\left(A_1, \dots, A_{i-1}, \sum_{j=1}^{n} A_i^j e_j, A_{i+1}, \dots, A_n \right) \\ &= \sum_{j=1}^{n} A_i^j \det(A_1, \dots, A_{i-1}, e_j, A_{i+1}, \dots, A_n) \\ &= \sum_{j=1}^{n} A_i^j \det X_i^j. \end{aligned} \tag{B.7}$$

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 (B.6) 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 (B.7) completes the proof. □

Each determinant $\det M_i^j$ is called a ***minor of A***, and (B.5) 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 (B.5) is called the ***cofactor of*** $\boldsymbol{A_i^j}$, and is denoted by cof_i^j.

Proposition B.36 (Cramer's Rule). *If A is a nonsingular $n \times n$ matrix, then A^{-1} is equal to* $1/(\det A)$ *times the transposed cofactor matrix of A. Thus, the entry in the i th row and j th column of A^{-1} is*

$$\left(A^{-1}\right)_j^i = \frac{1}{\det A} \operatorname{cof}_i^j = \frac{1}{\det A} (-1)^{i+j} \det M_i^j . \tag{B.8}$$

Proof. Let B_j^i denote the expression on the right-hand side of (B.8). Then

$$\sum_{j=1}^n B_j^i A_k^j = \frac{1}{\det A} \sum_{j=1}^n (-1)^{i+j} A_k^j \det M_i^j . \tag{B.9}$$

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 (B.9) 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 (B.9) is equivalent to the matrix equation $BA = I_n$, where B is the matrix (B_j^i). By Exercise B.19(c), therefore, $B = A^{-1}$. □

A square matrix $A = \left(A_j^i\right)$ is said to be ***upper triangular*** if $A_j^i = 0$ for $i > j$ (i.e., the only nonzero entries are on and above the main diagonal). Determinants of upper triangular matrices are particularly easy to compute.

Proposition B.37. *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 that X is ***block upper triangular*** if X has the form

$$X = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \tag{B.10}$$

for some matrices A, B, C of sizes $m \times m$, $m \times k$, and $k \times k$, respectively.

Proposition B.38. *If X is the block upper triangular matrix given by* (B.10), *then* $\det X = (\det A)(\det C)$.

Proof. If A is singular, then the columns of both A and 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 and using induction on m, we conclude easily that $\det X = \det C$ in this case. A similar argument shows that

$$\det \begin{pmatrix} A & 0 \\ 0 & I_k \end{pmatrix} = \det A.$$

In the general case, a straightforward computation yields the factorization

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I_k \end{pmatrix} \begin{pmatrix} I_m & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} I_m & A^{-1}B \\ 0 & I_k \end{pmatrix}. \tag{B.11}$$

By the preceding observations, the determinants of the first two factors are equal to $\det A$ and $\det C$, respectively; and the third factor is upper triangular, so its determinant is 1 by Proposition B.37. The result then follows from Theorem B.31. □

Inner Products and Norms

If V is a real vector space, an ***inner product on V*** is a map $V \times V \to \mathbb{R}$, usually written $(v, w) \mapsto \langle v, w \rangle$, that satisfies the following conditions:

(i) SYMMETRY:

$$\langle v, w \rangle = \langle w, v \rangle;$$

(ii) BILINEARITY:

$$\langle av + a'v', w \rangle = a\langle v, w \rangle + a'\langle v', w \rangle,$$
$$\langle v, bw + b'w' \rangle = b\langle v, w \rangle + b'\langle v, w' \rangle;$$

(iii) POSITIVE DEFINITENESS:

$$\langle v, v \rangle \geq 0, \quad \text{with equality if and only if } v = 0.$$

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 ***Euclidean dot product***:

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x^i y^i.$$

Suppose V is an inner product space. For each $v \in V$, the ***length of v*** is the nonnegative real number $|v| = \sqrt{\langle v, v\rangle}$. A ***unit vector*** is a vector of length 1. If $v, w \in V$ are nonzero vectors, the ***angle between v and w*** is defined to be the unique $\theta \in [0, \pi]$ satisfying

$$\cos\theta = \frac{\langle v, w\rangle}{|v|\,|w|}.$$

Two vectors $v, w \in V$ are said to be ***orthogonal*** if $\langle v, w\rangle = 0$; this means that either one of the vectors is zero, or the angle between them is $\pi/2$.

► **Exercise B.39.** Let V be an inner product space. Show that the length function associated with the inner product satisfies

$$\begin{aligned} |v| &> 0, && v \in V,\ v \neq 0,\\ |cv| &= |c||v|, && c \in \mathbb{R},\ v \in V,\\ |v + w| &\le |v| + |w|, && v, w \in V, \end{aligned}$$

and the ***Cauchy–Schwarz inequality***:

$$|\langle v, w\rangle| \le |v||w|, \quad v, w \in V.$$

Suppose V is a finite-dimensional inner product space. A basis $(E_1, \dots, E_n)$ for 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 B.40 (The Gram–Schmidt Algorithm). *Let V be an inner product space of dimension $n \ge 1$. Then V has an orthonormal basis. In fact, if $(E_1, \dots, E_n)$ is an arbitrary basis for V, there is an orthonormal basis $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ with the property that*

$$\operatorname{span}(\widetilde{E}_1, \dots, \widetilde{E}_k) = \operatorname{span}(E_1, \dots, E_k) \quad \textit{for } k = 1, \dots, n. \tag{B.12}$$

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 (B.12) for $k = 1, \dots, n-1$. Define $\widetilde{E}_n$ by

$$\widetilde{E}_n = \frac{E_n - \sum_{i=1}^{n-1}\langle E_n, \widetilde{E}_i\rangle \widetilde{E}_i}{\left|E_n - \sum_{i=1}^{n-1}\langle E_n, \widetilde{E}_i\rangle \widetilde{E}_i\right|}. \tag{B.13}$$

A computation shows that $(\widetilde{E}_1, \dots, \widetilde{E}_n)$ is the desired orthonormal basis for V. □

► **Exercise B.41.** For $w = (w^1, \dots, w^n)$ and $z = (z^1, \dots, z^n) \in \mathbb{C}^n$, define the ***Hermitian dot product*** by $w \cdot z = \sum_{j=1}^n w^j \overline{z^j}$, where, for any complex number $z = x + iy$, the notation $\overline{z}$ denotes the ***complex conjugate***: $\overline{z} = x - iy$. A basis $(E_1, \dots, E_n)$ for $\mathbb{C}^n$ (over $\mathbb{C}$) is said to be *orthonormal* if $E_i \cdot E_i = 1$ and $E_i \cdot E_j = 0$ for $i \neq j$. Show that the statement and proof of Proposition B.40 hold for the Hermitian dot product.

An isomorphism $T: V \to W$ between inner product spaces is called a ***linear isometry*** if it takes the inner product of V to that of W:

$$\langle Tv, Tw \rangle = \langle v, w \rangle \quad \text{for all } v, w \in V.$$

► **Exercise B.42.** Show that every linear isometry between inner product spaces is a homeomorphism that preserves lengths, angles, and orthogonality, and takes orthonormal bases to orthonormal bases.

► **Exercise B.43.** Given any basis (E_i) for a finite-dimensional vector space V, show that there is a unique inner product on V for which (E_i) is orthonormal.

► **Exercise B.44.** Suppose V is a finite-dimensional inner product space and $E: \mathbb{R}^n \to V$ is the basis map determined by some orthonormal basis. Show that E is a linear isometry when $\mathbb{R}^n$ is endowed with the Euclidean inner product.

The preceding exercise shows that finite-dimensional inner product spaces are geometrically indistinguishable from the Euclidean space of the same dimension.

If V is a finite-dimensional inner product space and $S \subseteq V$ is a subspace, the ***orthogonal complement of S in V*** is the set

$$S^{\perp} = \{v \in V : \langle v, w \rangle = 0 \text{ for all } w \in S\}.$$

► **Exercise B.45.** Let V be a finite-dimensional inner product space and let $S \subseteq V$ be a subspace. Show that $S^{\perp}$ is a subspace and $V = S \oplus S^{\perp}$.

Thanks to the result of the preceding exercise, for any subspace S of an inner product space V, there is a natural projection $\pi: V \to S$ with kernel $S^{\perp}$. This is called the ***orthogonal projection of V onto S***.

Norms

If V is a real vector space, a ***norm on V*** is a function from V to $\mathbb{R}$, written $v \mapsto |v|$, satisfying the following properties.

(i) POSITIVITY: $|v| \geq 0$ for all $v \in V$, with equality if and only if $v = 0$.
(ii) HOMOGENEITY: $|cv| = |c|\,|v|$ for all $c \in \mathbb{R}$ and $v \in V$.
(iii) TRIANGLE INEQUALITY: $|v + w| \leq |v| + |w|$ for all $v, w \in V$.

A vector space together with a specific choice of norm is called a ***normed linear space***. Exercise B.39 shows that the length function associated with any inner product is a norm; thus, in particular, every finite-dimensional vector space possesses many norms. Given a norm on V, the distance function $d(v, w) = |v - w|$ turns V into a metric space, yielding a topology on V called the ***norm topology***.

Example B.46 (Euclidean Spaces). Endowed with the ***Euclidean norm*** defined by

$$|x| = \sqrt{x \cdot x}, \tag{B.14}$$

$\mathbb{R}^n$ is a normed linear space, whose norm topology is exactly the Euclidean topology described in Appendix A. //

Example B.47 (The Frobenius Norm on Matrices). 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 norm on matrices, called the ***Frobenius norm***:

$$|A| = \sqrt{\sum_{i,j} \left(A^i_j\right)^2}. \tag{B.15}$$

Whenever we use a norm on matrices, it is always this one. //

► **Exercise B.48.** For any matrices $A \in \mathrm{M}(m \times n, \mathbb{R})$ and $B \in \mathrm{M}(n \times k, \mathbb{R})$, show that

$$|AB| \le |A|\,|B|.$$

Two norms $|\cdot|_1$ and $|\cdot|_2$ on a vector space V are said to be ***equivalent*** if there are positive constants c, C such that

$$c|v|_1 \le |v|_2 \le C|v|_1 \quad \text{for all } v \in V.$$

► **Exercise B.49.** Show that equivalent norms determine the same topology.

► **Exercise B.50.** Show that any two norms on a finite-dimensional vector space are equivalent. [Hint: first do the case in which $V = \mathbb{R}^n$ and one of the norms is the Euclidean norm, and consider the restriction of the other norm to the unit sphere.]

The preceding exercise shows that finite-dimensional normed linear spaces of the same dimension are topologically indistinguishable from one another. Thus, any such space automatically inherits all the usual topological properties of Euclidean space, such as compactness of closed and bounded subsets.

If V and W are normed linear spaces, a linear map $T: V \to W$ is said to be ***bounded*** if there exists a positive constant C such that

$$|Tv| \le C|v| \quad \text{for all } v \in V.$$

► **Exercise B.51.** Show that a linear map between normed linear spaces is continuous if and only if it is bounded. [Hint: to show that continuity of T implies boundedness, first show that there exists $\delta > 0$ such that $|x| < \delta \Rightarrow |T(x)| < 1$.]

► **Exercise B.52.** Show that every linear map between finite-dimensional normed linear spaces is bounded and therefore continuous.

Direct Products and Direct Sums

If $V_1, \dots, V_k$ are real vector spaces, their ***direct product*** is the vector space whose underlying set is the Cartesian product $V_1 \times \dots \times V_k$, with addition and scalar multiplication defined componentwise:

$$\left(v_1, \dots, v_k\right) + \left(v_1', \dots, v_k'\right) = \left(v_1 + v_1', \dots, v_k + v_k'\right),$$

$$c(v_1,\dots,v_k) = (cv_1,\dots,cv_k).$$

The basic example is the Euclidean space $\mathbb{R}^n = \mathbb{R}\times\cdots\times\mathbb{R}$.

For some applications (chiefly in our treatment of de Rham cohomology in Chapters 17 and 18), it is important to generalize this to an infinite number of vector spaces. For this discussion, we turn to the general setting of modules over a commutative ring $\mathcal{R}$. Linear maps between $\mathcal{R}$-modules are defined exactly as for vector spaces: if V and W are $\mathcal{R}$-modules, a map $F\colon V\to W$ is said to be ***$\mathcal{R}$-linear*** if $F(av+bw) = aF(v)+bF(w)$ for all $a,b\in\mathcal{R}$ and $v,w\in V$. If V is an $\mathcal{R}$-module, a subset $S\subseteq V$ is called a ***submodule of V*** if it is closed under addition and scalar multiplication, so it is itself an $\mathcal{R}$-module. Throughout the rest of this section we assume that $\mathcal{R}$ is a fixed commutative ring. In all of our applications, $\mathcal{R}$ will be either the field $\mathbb{R}$ of real numbers, in which case the modules are real vector spaces and the linear maps are the usual ones, or the ring of integers $\mathbb{Z}$, in which case the modules are abelian groups and the linear maps are group homomorphisms.

If $(V_\alpha)_{\alpha\in A}$ is an arbitrary indexed family of sets, their ***Cartesian product***, denoted by $\prod_{\alpha\in A} V_\alpha$, is defined as the set of functions $v\colon A\to\bigcup_{\alpha\in A} V_\alpha$ with the property that $v(\alpha)\in V_\alpha$ for each α. Thanks to the axiom of choice, the Cartesian product of a nonempty indexed family of nonempty sets is nonempty. If v is an element of the Cartesian product, we usually denote the value of v at $\alpha\in A$ by v_α instead of $v(\alpha)$; the element v itself is usually denoted by $(v_\alpha)_{\alpha\in A}$, or just (v_α) if the index set is understood. This can be thought of as an indexed family of elements of the sets V_α, or an "A-tuple." For each $\beta\in A$, we have a canonical ***projection map*** $\pi_\beta\colon \prod_{\alpha\in A} V_\alpha\to V_\beta$, defined by

$$\pi_\beta\big((v_\alpha)_{\alpha\in A}\big) = v_\beta.$$

Now suppose that $(V_\alpha)_{\alpha\in A}$ is an indexed family of $\mathcal{R}$-modules. The ***direct product*** of the family is the set $\prod_{\alpha\in A} V_\alpha$, made into an $\mathcal{R}$-module by defining addition and scalar multiplication as follows:

$$(v_\alpha) + \big(v'_\alpha\big) = \big(v_\alpha + v'_\alpha\big),$$
$$c(v_\alpha) = (cv_\alpha).$$

The zero element of this module is the A-tuple with $v_\alpha = 0$ for every α. It is easy to check that each projection map π_β is $\mathcal{R}$-linear.

Proposition B.53 (Characteristic Property of the Direct Product). *Let $(V_\alpha)_{\alpha\in A}$ be an indexed family of $\mathcal{R}$-modules. Given an $\mathcal{R}$-module W and a family of $\mathcal{R}$-linear maps $G_\alpha\colon W\to V_\alpha$, there exists a unique $\mathcal{R}$-linear map $G\colon W\to\prod_{\alpha\in A} V_\alpha$ such that $\pi_\alpha\circ G = G_\alpha$ for each $\alpha\in A$.*

► **Exercise B.54.** Prove the preceding proposition.

Complementary to direct products is the notion of direct sums. Given an indexed family $(V_\alpha)_{\alpha\in A}$ as above, we define the ***direct sum*** of the family to be the submodule of their direct product consisting of A-tuples $(v_\alpha)_{\alpha\in A}$ with the property that $v_\alpha = 0$

for all but finitely many α. The direct sum is denoted by $\bigoplus_{\alpha\in A} V_\alpha$, or in the case of a finite family by $V_1 \oplus \cdots \oplus V_k$. For finite families of modules, the direct product and the direct sum are identical.

For each $\beta \in A$, there is a canonical $\mathcal{R}$-linear injection $\iota_\beta\colon V_\beta \to \bigoplus_{\alpha\in A} V_\alpha$, defined by letting $\iota_\beta(v)$ be the A-tuple $(v_\alpha)_{\alpha\in A}$ with $v_\beta = v$ and $v_\alpha = 0$ for $\alpha \neq \beta$. In the case of a finite direct sum, this just means $\iota_\beta(v) = (0,\dots,0,v,0,\dots,0)$, with v in position β.

Proposition B.55 (Characteristic Property of the Direct Sum). *Let $(V_\alpha)_{\alpha\in A}$ be an indexed family of $\mathcal{R}$-modules. Given an $\mathcal{R}$-module W and a family of $\mathcal{R}$-linear maps $G_\alpha\colon V_\alpha \to W$, there exists a unique $\mathcal{R}$-linear map $G\colon \bigoplus_{\alpha\in A} V_\alpha \to W$ such that $G \circ \iota_\alpha = G_\alpha$ for each $\alpha \in A$.*

► **Exercise B.56.** Prove the preceding proposition.

If W is an $\mathcal{R}$-module and $(V_\alpha)_{\alpha\in A}$ is a family of subspaces of W, then the characteristic property applied to the inclusions $\iota_\alpha\colon V_\alpha \hookrightarrow W$ guarantees the existence of a canonical $\mathcal{R}$-linear map $\bigoplus_\alpha V_\alpha \to W$ that restricts to inclusion on each V_α. This map is an isomorphism precisely when the V_α's are chosen so that every element of W has a unique expression as a finite linear combination $\sum_\alpha c_\alpha v_\alpha$ with $v_\alpha \in V_\alpha$ for each α. In this case, we can naturally *identify* W with $\bigoplus_\alpha V_\alpha$, and we say that W is the ***internal direct sum*** of the submodules $\{V_\alpha\}$, extending the terminology we introduced earlier for two complementary subspaces of a vector space. A direct sum of an abstract family of modules is sometimes called their ***external direct sum*** to distinguish it from an internal direct sum.

If V and W are $\mathcal{R}$-modules, the set $\operatorname{Hom}_{\mathcal{R}}(V,W)$ of all $\mathcal{R}$-linear maps from V to W is an $\mathcal{R}$-module under pointwise addition and scalar multiplication:

$$(F+G)(v) = F(v) + G(v),$$
$$(aF)(v) = aF(v).$$

(If V and W are real vector spaces, then $\operatorname{Hom}_{\mathcal{R}}(V,W)$ is just the space $\mathrm{L}(V;W)$ of $\mathbb{R}$-linear maps; if they are abelian groups, then $\mathbb{Z}$-linear maps are group homomorphisms, and we usually write $\operatorname{Hom}(V,W)$ instead of $\operatorname{Hom}_{\mathbb{Z}}(V,W)$.) Our last proposition is used in the proof of the de Rham theorem in Chapter 18.

Proposition B.57. *Let $(V_\alpha)_{\alpha\in A}$ be an indexed family of $\mathcal{R}$-modules. For each $\mathcal{R}$-module W, there is a canonical isomorphism*

$$\operatorname{Hom}_{\mathcal{R}}\Big(\bigoplus_{\alpha\in A} V_\alpha, W\Big) \cong \prod_{\alpha\in A} \operatorname{Hom}_{\mathcal{R}}(V_\alpha, W).$$

Proof. Define a map $\Phi\colon \operatorname{Hom}_{\mathcal{R}}\big(\bigoplus_{\alpha\in A} V_\alpha, W\big) \to \prod_{\alpha\in A} \operatorname{Hom}_{\mathcal{R}}(V_\alpha, W)$ by setting $\Phi(F) = (F_\alpha)_{\alpha\in A}$, where $F_\alpha = F \circ \iota_\alpha$.

To prove that Φ is surjective, suppose $(F_\alpha)_{\alpha\in A}$ is an arbitrary element of $\prod_{\alpha\in A} \operatorname{Hom}_{\mathcal{R}}(V_\alpha, W)$. This just means that for each α, F_α is an $\mathcal{R}$-linear map from V_α to W. The characteristic property of the direct sum then guarantees the existence of an $\mathcal{R}$-linear map $F\colon \bigoplus_{\alpha\in A} V_\alpha \to W$ satisfying $F \circ \iota_\alpha = F_\alpha$ for each α, which is equivalent to $\Phi(F) = (F_\alpha)_{\alpha\in A}$.

To prove that Φ is injective, suppose that $\Phi(F) = (F_\alpha)_{\alpha\in A} = 0$. By definition of the zero element of the direct product, this means that $F_\alpha = F \circ \iota_\alpha$ is the zero homomorphism for each α. By the uniqueness assertion in Proposition B.55, this implies that F itself is the zero homomorphism. $\square$

Appendix C
Review of Calculus

In this appendix 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 [Apo74], [Rud76], or [Str00].

Total and Partial Derivatives

For maps between (open subsets of) finite-dimensional vector spaces, the most general notion of derivative is the total derivative.

Let V, W be finite-dimensional vector spaces, which we may assume to be endowed with norms. If $U \subseteq V$ is an open subset and $a \in U$, a map $F\colon U \to W$ is said to be ***differentiable at a*** if there exists a linear map $L\colon V \to W$ such that

$$\lim_{v \to 0} \frac{|F(a+v) - F(a) - Lv|}{|v|} = 0. \tag{C.1}$$

The norm in the numerator of this expression is that of W, while the norm in the denominator is that of V. Because all norms on a finite-dimensional vector space are equivalent (Exercise B.49), the definition is independent of both choices of norms.

► **Exercise C.1.** Suppose $F\colon U \to W$ is differentiable at $a \in U$. Show that the linear map L satisfying (C.1) is unique.

If F is differentiable at a, the linear map L satisfying (C.1) is denoted by $DF(a)$ and is called the ***total derivative of F at a***. Condition (C.1) can also be written

$$F(a+v) = F(a) + DF(a)v + R(v), \tag{C.2}$$

where the remainder term $R(v) = F(a+v) - F(a) - DF(a)v$ satisfies $|R(v)|/|v| \to 0$ as $v \to 0$. Thus the total derivative represents the "best linear approximation" to $F(a+v) - F(a)$ near a.

► **Exercise C.2.** Suppose V, W, X are finite-dimensional vector spaces, $U \subseteq V$ is an open subset, a is a point in U, and $F, G\colon U \to W$ and $f, g\colon U \to \mathbb{R}$ are maps. Prove the following statements.

J.M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics 218, DOI 10.1007/978-1-4419-9982-5,

(a) If F is differentiable at a, then it is continuous at a.
(b) If F is a constant map, then F is differentiable at a and $DF(a) = 0$.
(c) If F and G are differentiable at a, then $F + G$ is also, and

$$D(F + G)(a) = DF(a) + DG(a).$$

(d) If f and g are differentiable at a, then fg is also, and

$$D(fg)(a) = f(a)Dg(a) + g(a)Df(a).$$

(e) If f and g are differentiable at a and $g(a) \neq 0$, then f/g is differentiable at a, and

$$D(f/g)(a) = \frac{g(a)Df(a) - f(a)Dg(a)}{g(a)^2}.$$

(f) If $T: V \to W$ is a linear map, then T is differentiable at every point $v \in V$, with total derivative equal to T itself: $DT(v) = T$.
(g) If $B: V \times W \to X$ is a bilinear map, then B is differentiable at every point $(v, w) \in V \times W$, and

$$DB(v, w)(x, y) = B(v, y) + B(x, w).$$

Proposition C.3 (The Chain Rule for Total Derivatives). *Suppose V, W, X are finite-dimensional vector spaces, $U \subseteq V$ and $\widetilde{U} \subseteq W$ are open subsets, and $F: U \to \widetilde{U}$ and $G: \widetilde{U} \to X$ 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*

$$D(G \circ F)(a) = DG\big(F(a)\big) \circ DF(a).$$

Proof. Let $A = DF(a)$ and $B = DG\big(F(a)\big)$. We need to show that

$$\lim_{v \to 0} \frac{|G(F(a + v)) - G(F(a)) - BAv|}{|v|} = 0. \tag{C.3}$$

Let us write $b = F(a)$ and $w = F(a + v) - F(a)$. With these substitutions, we can rewrite the quotient in (C.3) as

$$\begin{aligned}\frac{|G(b + w) - G(b) - BAv|}{|v|} &= \frac{|G(b + w) - G(b) - Bw + Bw - BAv|}{|v|} \\ &\leq \frac{|G(b + w) - G(b) - Bw|}{|v|} + \frac{|B(w - Av)|}{|v|}. \end{aligned} \tag{C.4}$$

Since A and B are linear, Exercise B.52 shows that there are constants C, C' such that $|Ax| \leq C|x|$ for all $x \in V$, and $|By| \leq C'|y|$ for all $y \in W$. The differentiability of F at a means that for any $\varepsilon > 0$, we can ensure that

$$|w - Av| = |F(a + v) - F(a) - Av| \leq \varepsilon|v|$$

as long as v lies in a small enough neighborhood of 0. Moreover, as $v \to 0$, $|w| = |F(a + v) - F(a)| \to 0$ by continuity of F. Therefore, the differentiability of G at b means that by making $|v|$ even smaller if necessary, we can also achieve

$$|G(b + w) - G(b) - Bw| \leq \varepsilon|w|.$$

Putting all of these estimates together, we see that for $|v|$ sufficiently small, (C.4) is bounded by

$$\begin{aligned}
\varepsilon\frac{|w|}{|v|} + C'\frac{|w-Av|}{|v|} &= \varepsilon\frac{|w-Av+Av|}{|v|} + C'\frac{|w-Av|}{|v|} \\
&\le \varepsilon\frac{|w-Av|}{|v|} + \varepsilon\frac{|Av|}{|v|} + C'\frac{|w-Av|}{|v|} \\
&\le \varepsilon^2 + \varepsilon C + C'\varepsilon,
\end{aligned}$$

which can be made as small as desired. □

Partial Derivatives

Now we specialize to maps between Euclidean spaces. Suppose $U \subseteq \mathbb{R}^n$ is open and $f\colon U \to \mathbb{R}$ is a real-valued function. For any $a = \left(a^1, \dots, a^n\right) \in U$ and any $j \in \{1, \dots, n\}$, the ***j th partial derivative of f at a*** is defined to be the ordinary derivative of f with respect to x^j while holding the other variables fixed:

$$\begin{aligned}
\frac{\partial f}{\partial x^j}(a) &= \lim_{h\to 0} \frac{f(a^1, \dots, a^j + h, \dots, a^n) - f(a^1, \dots, a^j, \dots, a^n)}{h} \\
&= \lim_{h\to 0} \frac{f(a + he_j) - f(a)}{h},
\end{aligned}$$

if the limit exists.

More generally, for a vector-valued function $F\colon U \to \mathbb{R}^m$, we can write the coordinates of $F(x)$ as $F(x) = \left(F^1(x), \dots, F^m(x)\right)$. This defines m functions $F^1, \dots, F^m\colon U \to \mathbb{R}$ called the ***component functions of F***. The partial derivatives of F are defined simply to be the partial derivatives $\partial F^i/\partial x^j$ of its component functions. The matrix $\left(\partial F^i/\partial x^j\right)$ of partial derivatives is called the ***Jacobian matrix of F***, and its determinant is called the ***Jacobian determinant of F***.

If $F\colon U \to \mathbb{R}^m$ is a function for which each partial derivative exists at each point in U and the functions $\partial F^i/\partial x^j\colon U \to \mathbb{R}$ 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 (first) partial derivatives of those of order $k-1$, when they exist.

In general, if $U \subseteq \mathbb{R}^n$ is an open subset and $k \ge 0$, 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.) Because existence and continuity of derivatives are local properties, clearly F is C^k if and only if it has that property in a neighborhood of each point in U.

A function that is of class C^k for every $k \geq 0$ is said to be of ***class*** $\boldsymbol{C^\infty}$, ***smooth***, or ***infinitely differentiable***. If U and V are open subsets of Euclidean spaces, a function $F: U \to V$ is called a ***diffeomorphism*** if it is smooth and bijective and its inverse function is also smooth.

One consequence of the chain rule is worth noting.

Proposition C.4. *Suppose $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open subsets and $F: U \to V$ is a diffeomorphism. Then $m = n$, and for each $a \in U$, the total derivative $DF(a)$ is invertible, with $DF(a)^{-1} = D(F^{-1})(F(a))$.*

Proof. Because $F^{-1} \circ F = \mathrm{Id}_U$, the chain rule implies that for each $a \in U$,

$$\mathrm{Id}_{\mathbb{R}^n} = D(\mathrm{Id}_U)(a) = D(F^{-1} \circ F)(a) = D(F^{-1})(F(a)) \circ DF(a). \qquad \text{(C.5)}$$

Similarly, $F \circ F^{-1} = \mathrm{Id}_V$ implies that $DF(a) \circ D(F^{-1})(F(a))$ is the identity on $\mathbb{R}^m$. This implies that $DF(a)$ is invertible with inverse $D(F^{-1})(F(a))$, and therefore $m = n$. □

We sometimes need to consider smoothness of functions whose domains are subsets of $\mathbb{R}^n$ that are not open. If $A \subseteq \mathbb{R}^n$ is an *arbitrary* subset, a function $F: A \to \mathbb{R}^m$ is said to be ***smooth on A*** if it admits a smooth extension to an open neighborhood of each point, or more precisely, if for every $x \in A$, there exist an open subset $U_x \subseteq \mathbb{R}^n$ containing x and a smooth function $\widetilde{F}: U_x \to \mathbb{R}^m$ that agrees with F on $U_x \cap A$. The notion of diffeomorphism extends to arbitrary subsets in the obvious way: given arbitrary subsets $A, B \subseteq \mathbb{R}^n$, a ***diffeomorphism from A to B*** is a smooth bijective map $f: A \to B$ with smooth inverse.

We are especially concerned with real-valued functions, that is, functions whose codomain is $\mathbb{R}$. If $U \subseteq \mathbb{R}^n$ 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 $C^\infty(U)$. Sums, constant multiples, and products of functions are defined pointwise: for $f, g: U \to \mathbb{R}$ and $c \in \mathbb{R}$,

$$\begin{aligned}(f+g)(x) &= f(x) + g(x),\\ (cf)(x) &= c(f(x)),\\ (fg)(x) &= f(x)g(x).\end{aligned}$$

► **Exercise C.5.** Let $U \subseteq \mathbb{R}^n$ be an open subset, and suppose $f, g \in C^\infty(U)$ and $c \in \mathbb{R}$.

(a) Show that $f + g$, cf, and fg are smooth.
(b) Show that these operations turn $C^\infty(U)$ into a commutative ring and a commutative and associative algebra over $\mathbb{R}$ (see p. 624).
(c) Show that if g never vanishes on U, then f/g is smooth.

The following important result shows that for most interesting functions, the order in which we take partial derivatives is irrelevant. For a proof, see [Apo74, Rud76, Str00].

Proposition C.6 (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 function 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 C.7. *If $F\colon U \to \mathbb{R}^m$ is smooth, then the mixed partial derivatives of F of any order are independent of the order of differentiation.* □

Next we study the relationship between total and partial derivatives. Suppose $U \subseteq \mathbb{R}^n$ 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 proposition identifies that matrix as the Jacobian of F.

Proposition C.8. *Let $U \subseteq \mathbb{R}^n$ 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 $B = DF(a)$, and for $v \in \mathbb{R}^n$ small enough that $a + v \in U$, let $R(v) = F(a+v) - F(a) - Bv$. The fact that F is differentiable at a implies that each component of the vector-valued function $R(v)/|v|$ goes to zero as $v \to 0$. The ith partial derivative of F^j at a, if it exists, is

$$\begin{aligned} \frac{\partial F^j}{\partial x^i}(a) &= \lim_{t \to 0} \frac{F^j(a + te_i) - F^j(a)}{t} = \lim_{t \to 0} \frac{B_i^j t + R^j(te_i)}{t} \\ &= B_i^j + \lim_{t \to 0} \frac{R^j(te_i)}{t}. \end{aligned}$$

The norm of the quotient on the right above is $\left|R^j(te_i)\right| / |te_i|$, which approaches zero as $t \to 0$. It follows that $\partial F^j / \partial x^i(a)$ exists and is equal to B_i^j as claimed. □

► **Exercise C.9.** Suppose $U \subseteq \mathbb{R}^n$ is open. Show that a function $F\colon U \to \mathbb{R}^m$ is differentiable at $a \in U$ if and only if each of its component functions $F^1, \dots, F^m$ is differentiable at a. Show that if this is the case, then

$$DF(a) = \begin{pmatrix} DF^1(a) \\ \vdots \\ DF^m(a) \end{pmatrix}.$$

The preceding exercise implies that for an open interval $J \subseteq \mathbb{R}$, a map $\gamma\colon J \to \mathbb{R}^m$ is differentiable if and only if its component functions are differentiable in the sense of one-variable calculus.

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 [Apo74, Rud76, Str00].

Proposition C.10. *Let $U \subseteq \mathbb{R}^n$ be open. If $F\colon U \to \mathbb{R}^m$ is of class C^1, then it is differentiable at each point of U.*

For functions between Euclidean spaces, the chain rule can be rephrased in terms of partial derivatives.

Corollary C.11 (The Chain Rule for Partial Derivatives). *Let $U \subseteq \mathbb{R}^n$ and $\widetilde{U} \subseteq \mathbb{R}^m$ be open subsets, and let $x = \left(x^1, \dots, x^n\right)$ denote the standard coordinates on U and $y = \left(y^1, \dots, y^m\right)$ those on $\widetilde{U}$.*

(a) *A 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*

$$\frac{\partial (G^i \circ F)}{\partial x^j}(x) = \sum_{k=1}^{m} \frac{\partial G^i}{\partial y^k}\big(F(x)\big)\frac{\partial F^k}{\partial x^j}(x).$$

(b) *If F and G are smooth, then $G \circ F$ is smooth.*

► **Exercise C.12.** Prove Corollary C.11.

From the chain rule and induction one can derive formulas for the higher partial derivatives of a composite function as needed, provided the functions in question are sufficiently differentiable.

► **Exercise C.13.** Suppose $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ are arbitrary subsets, and $F\colon A \to \mathbb{R}^m$ and $G\colon B \to \mathbb{R}^p$ are smooth maps (in the sense that they have smooth extensions in a neighborhood of each point) such that $F(A) \subseteq B$. Show that $G \circ F\colon A \to \mathbb{R}^p$ is smooth.

Now suppose $f\colon U \to \mathbb{R}$ is a smooth real-valued function on an open subset $U \subseteq \mathbb{R}^n$, and $a \in U$. For each vector $v \in \mathbb{R}^n$, we define the ***directional derivative of f in the direction v at a*** to be the number

$$D_v f(a) = \left.\frac{d}{dt}\right|_{t=0} f(a + tv). \tag{C.6}$$

(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 function $t \mapsto a + tv \mapsto f(a + tv)$, by the chain rule it 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 fundamental theorem of calculus expresses one well-known relationship between integrals and derivatives. Another 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 [Apo74, Rud76, Str00].

Theorem C.14 (Differentiation Under an Integral Sign). *Let $U \subseteq \mathbb{R}^n$ be an open subset, let $a, b \in \mathbb{R}$, and let $f: U \times [a,b] \to \mathbb{R}$ be a continuous function such that the partial derivatives $\partial f/\partial x^i: U \times [a,b] \to \mathbb{R}$ exist and are continuous on $U \times [a,b]$ for $i = 1, \dots, n$. Define $F: U \to \mathbb{R}$ 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:

$$\frac{\partial F}{\partial x^i}(x) = \int_a^b \frac{\partial f}{\partial x^i}(x,t)\,dt.$$

You are probably familiar with Taylor's theorem, which shows how a sufficiently smooth function can be approximated near a point by a polynomial. We need a version of Taylor's theorem in several variables that gives an explicit integral form for the remainder term. In order to express it concisely, it helps to introduce some shorthand notation. For any m-tuple $I = (i_1, \dots, i_m)$ of indices with $1 \le i_j \le n$, we let $|I| = m$ denote the number of indices in I, and

$$\partial_I = \frac{\partial^m}{\partial x^{i_1} \cdots \partial x^{i_m}},$$
$$(x-a)^I = \left(x^{i_1} - a^{i_1}\right) \cdots \left(x^{i_m} - a^{i_m}\right).$$

Theorem C.15 (Taylor's Theorem). *Let $U \subseteq \mathbb{R}^n$ be an open subset, and let $a \in U$ be fixed. Suppose $f \in C^{k+1}(U)$ for some $k \ge 0$. If W is any convex subset of U containing a, then for all $x \in W$,*

$$f(x) = P_k(x) + R_k(x), \tag{C.7}$$

*where P_k is the **kth-order Taylor polynomial of f at a**, defined by*

$$P_k(x) = f(a) + \sum_{m=1}^{k} \frac{1}{m!} \sum_{I:|I|=m} \partial_I f(a)(x-a)^I, \tag{C.8}$$

*and R_k is the **kth remainder term**, given by*

$$R_k(x) = \frac{1}{k!} \sum_{I:|I|=k+1} (x-a)^I \int_0^1 (1-t)^k \partial_I f\big(a + t(x-a)\big)\,dt. \tag{C.9}$$

Proof. For $k = 0$ (where we interpret P_0 to mean $f(a)$), this is just the fundamental theorem of calculus applied to the function $u(t) = f\big(a + t(x-a)\big)$, together with the chain rule. Assuming the result holds for some k, integration by parts applied to the integral in the remainder term yields

$$\int_0^1 (1-t)^k \partial_I f\big(a + t(x-a)\big)\,dt$$

$$= \left[-\frac{(1-t)^{k+1}}{k+1} \partial_I f\big(a + t(x-a)\big) \right]_{t=0}^{t=1}$$
$$+ \int_0^1 \frac{(1-t)^{k+1}}{k+1} \frac{\partial}{\partial t}\big(\partial_I f\big(a + t(x-a)\big)\big)\, dt$$
$$= \frac{1}{k+1} \partial_I f(a)$$
$$+ \frac{1}{k+1} \sum_{j=1}^n \big(x^j - a^j\big) \int_0^1 (1-t)^{k+1} \frac{\partial}{\partial x^j} \partial_I f\big(a + t(x-a)\big)\, dt.$$

When we insert this into (C.7), we obtain the analogous formula with k replaced by $k+1$. □

Corollary C.16. *Suppose $U \subseteq \mathbb{R}^n$ is an open subset, $a \in U$, and $f \in C^{k+1}(U)$ for some $k \geq 0$. If W is a convex subset of U containing a on which all of the $(k+1)$st partial derivatives of f are bounded in absolute value by a constant M, then for all $x \in W$,*

$$\big|f(x) - P_k(x)\big| \leq \frac{n^{k+1} M}{(k+1)!} |x-a|^{k+1},$$

where P_k is the kth Taylor polynomial of f at a, defined by (C.8).

Proof. There are n^{k+1} terms on the right-hand side of (C.9), and each term is bounded in absolute value by $(1/(k+1)!)|x-a|^{k+1} M$. □

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 is 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 consider. For more details on the aspects of integration theory described here, you can consult [Apo74, Rud76, Str00].

A ***closed rectangle*** in $\mathbb{R}^n$ is a product set of the form $\left[a^1, b^1\right] \times \cdots \times \left[a^n, b^n\right]$, for real numbers $a^i < b^i$. Analogously, an ***open rectangle*** is a set of the form $\left(a^1, b^1\right) \times \cdots \times \left(a^n, b^n\right)$. If A is a rectangle of either type, the ***volume of A***, denoted by $\operatorname{Vol}(A)$, is defined to be the product of the lengths of its component intervals:

$$\operatorname{Vol}(A) = \big(b^1 - a^1\big) \cdots \big(b^n - a^n\big). \tag{C.10}$$

A rectangle is called a ***cube*** if all of its side lengths $\big(b^i - a^i\big)$ are equal.

Given a closed interval $[a,b] \subseteq \mathbb{R}$, a ***partition of $[a,b]$*** is a finite sequence $P = (a_0, \ldots, a_k)$ 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 P***. Similarly, if $A = \left[a^1, b^1\right] \times \cdots \times \left[a^n, b^n\right]$ is a closed rectangle, a ***partition of A*** is an n-tuple $P =$

$(P_1, \dots, P_n)$, where each P_i is a partition of $\left[a^i, b^i\right]$. 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 intersect only on their boundaries.

Suppose $A \subseteq \mathbb{R}^n$ is a closed rectangle and $f : A \to \mathbb{R}$ is a bounded function. For each partition P of A, we define the ***lower sum of f with respect to P*** by

$$\mathrm{L}(f, P) = \sum_j \left(\inf_{R_j} f \right) \mathrm{Vol}(R_j),$$

where the sum is over all the subrectangles R_j of P. Similarly, the ***upper sum*** is

$$\mathrm{U}(f, P) = \sum_j \left(\sup_{R_j} f \right) \mathrm{Vol}(R_j).$$

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 C.17. *Let $A \subseteq \mathbb{R}^n$ be a closed rectangle, and let $f : A \to \mathbb{R}$ be a bounded function. For any pair of partitions P and P' of A,*

$$\mathrm{L}(f, P) \le \mathrm{U}(f, P').$$

Proof. Write $P = (P_1, \dots, P_n)$ and $P' = (P_1', \dots, P_n')$, and let Q be the partition $Q = (P_1 \cup P_1', \dots, P_n \cup P_n')$. Each subrectangle of P or P' is a union of finitely many subrectangles of Q. An easy computation shows that

$$\mathrm{L}(f, P) \le \mathrm{L}(f, Q) \le \mathrm{U}(f, Q) \le \mathrm{U}(f, P'),$$

from which the result follows. □

The ***lower integral of f over A*** is

$$\underline{\int_A} f \, dV = \sup\left\{\mathrm{L}(f, P) : P \text{ is a partition of } A\right\},$$

and the ***upper integral*** is

$$\overline{\int_A} f \, dV = \inf\left\{\mathrm{U}(f, P) : P \text{ is a partition of } A\right\}.$$

Clearly, both numbers exist, because f is bounded, and Lemma C.17 implies that the lower integral is less than or equal to the upper integral.

If $f : A \to \mathbb{R}$ is a bounded function whose upper and lower integrals are equal, we say that f is ***(Riemann) integrable over A***, 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, has no meaning on its own; it is just a "closing bracket" for the integral sign. Other common notations are

$$\int_A f \quad \text{or} \quad \int_A f\,dx^1 \cdots dx^n \quad \text{or} \quad \int_A f\left(x^1,\dots,x^n\right)dx^1\cdots dx^n.$$

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 $X \subseteq \mathbb{R}^n$ is said to have ***measure zero*** if for every $\delta > 0$, there exists a countable cover of X by open rectangles $\{C_i\}$ such that $\sum_i \operatorname{Vol}(C_i) < \delta$. (Those who are familiar with the theory of Lebesgue measure will notice that this is equivalent to the condition that the Lebesgue measure of X be equal to zero.)

Proposition C.18 (Properties of Sets of Measure Zero).

(a) *If $X \subseteq \mathbb{R}^n$ has measure zero and $x_0 \in \mathbb{R}^n$, then the translated subset $x_0 + X = \{x_0 + a : a \in X\}$ also has measure zero.*
(b) *Every subset of a set of measure zero in $\mathbb{R}^n$ has measure zero.*
(c) *A countable union of sets of measure zero in $\mathbb{R}^n$ has measure zero.*
(d) *If $k < n$, then every subset of $\mathbb{R}^k$ (viewed as the set of points $x \in \mathbb{R}^n$ with $x^{k+1} = \cdots = x^n = 0$) has measure zero in $\mathbb{R}^n$.*

► **Exercise C.19.** Prove Proposition C.18.

Part (d) of this proposition 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 that 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 C.20 (Lebesgue's Integrability Criterion). *Let $A \subseteq \mathbb{R}^n$ be a closed 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. Suppose the set S has measure zero, and let $\varepsilon > 0$ be given. By definition of measure zero sets, S can be covered by a countable collection of open rectangles $\{C_i\}$, the sum of whose volumes is less than ε. For each $q \in A \smallsetminus S$, since f is continuous at q, there is an open rectangle D_q centered at q such that $|f(x) - f(q)| < \varepsilon$ for all $x \in D_q \cap A$. By shrinking D_q a little, we can arrange that the same inequality holds for all $x \in \bar{D}_q \cap A$. This implies $\sup_{\bar{D}_q} f - \inf_{\bar{D}_q} f \le 2\varepsilon$.

The collection of all rectangles of the form C_i or D_q is an open cover of A. By compactness, finitely many of them cover A. Let us relabel these rectangles as $\{C_1, \dots, C_k, D_1, \dots, D_l\}$. Replacing each C_i or D_j by its intersection with $\operatorname{Int} A$, we may assume that each $\overline{C}_i$ and each $\overline{D}_j$ is contained in A.

Since there are only finitely many rectangles $\{C_i, D_j\}$, there is a partition P of A with the property that each $\overline{C}_i$ or $\overline{D}_j$ 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 $\overline{C}_i$ for some i, and every subrectangle in $\mathcal{D}$ is contained in $\overline{D}_j$ for some j. Then

$$\begin{aligned}
&\mathrm{U}(f,P) - \mathrm{L}(f,P)\\
&\quad = \sum_i \Big(\sup_{R_i} f\Big) \operatorname{Vol}(R_i) - \sum_i \Big(\inf_{R_i} f\Big) \operatorname{Vol}(R_i)\\
&\quad = \sum_{R_i \in \mathcal{C}} \Big(\sup_{R_i} f - \inf_{R_i} f\Big) \operatorname{Vol}(R_i) + \sum_{R_i \in \mathcal{D}} \Big(\sup_{R_i} f - \inf_{R_i} f\Big) \operatorname{Vol}(R_i)\\
&\quad \le \Big(\sup_A f - \inf_A f\Big) \sum_{R_i \in \mathcal{C}} \operatorname{Vol}(R_i) + 2\varepsilon \sum_{R_i \in \mathcal{D}} \operatorname{Vol}(R_i)\\
&\quad \le \Big(\sup_A f - \inf_A f\Big)\varepsilon + 2\varepsilon \operatorname{Vol}(A).
\end{aligned}$$

It follows that

$$\overline{\int_A} f\, dV - \underline{\int_A} f\, dV \le \Big(\sup_A f - \inf_A f\Big)\varepsilon + 2\varepsilon \operatorname{Vol}(A),$$

which can be made as small as desired by taking ε sufficiently small. This implies that the upper and lower integrals of f are equal, so f is integrable. □

In fact, Lebesgue's criterion is both necessary and sufficient for Riemann integrability, but we do not need that.

Now suppose $D \subseteq \mathbb{R}^n$ is an arbitrary bounded set, and $f\colon D \to \mathbb{R}$ is a bounded function. Define $f_D\colon \mathbb{R}^n \to \mathbb{R}$ by

$$f_D(x) = \begin{cases} f(x), & x \in D,\\ 0, & x \in \mathbb{R}^n \smallsetminus D. \end{cases} \tag{C.11}$$

If the integral

$$\int_A f_D\, dV \tag{C.12}$$

exists for some closed rectangle A containing D, then f is said to be ***integrable over D***. The integral (C.12) is denoted by $\int_D f\, dV$ and called the ***integral of f***

over $\boldsymbol{D}$. It is easy to check that both the integrability of f and the value of the integral are independent of the rectangle chosen.

In practice, we are interested only in integrals of bounded continuous functions. However, since we 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 (C.11). The main reason for proving Proposition C.20 is that it allows us to give a simple description of domains on which all bounded continuous functions are integrable.

A subset $D \subseteq \mathbb{R}^n$ is called a ***domain of integration*** if D is bounded and ∂D has n-dimensional measure zero. It follows from Proposition C.18 that every open or closed rectangle is a domain of integration, and a finite union of domains of integration is again a domain of integration.

Proposition C.21. *If $D \subseteq \mathbb{R}^n$ is a domain of integration, then every bounded continuous real-valued function on D is integrable over D.*

Proof. Let $f: D \to \mathbb{R}$ be bounded and continuous, let $f_D: \mathbb{R}^n \to \mathbb{R}$ be the function defined by (C.11), and let A be a closed rectangle containing D. To prove the theorem, we need only show that the set of points in A where f_D is discontinuous has measure zero.

If $x \in \operatorname{Int} D$, then $f_D = f$ on a neighborhood of x, so f_D is continuous at x. Similarly, if $x \in \mathbb{R}^n \smallsetminus \overline{D}$, 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. □

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*** $\boldsymbol{D}$ is defined to be

$$\operatorname{Vol}(D) = \int_D 1\, dV. \tag{C.13}$$

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 C.22 (Properties of Volume). *Let $D \subseteq \mathbb{R}^n$ be a domain of integration.*

(a) *If D is an open or closed rectangle, then the two definitions* (C.10) *and* (C.13) *of* $\operatorname{Vol}(D)$ *agree.*
(b) $\operatorname{Vol}(D) \geq 0$, *with equality if and only if D has measure zero.*
(c) *If $D_1, \dots, D_k$ are domains of integration whose union is D, then*

$$\operatorname{Vol}(D) \leq \operatorname{Vol}(D_1) + \cdots + \operatorname{Vol}(D_k),$$

with equality if and only if $D_i \cap D_j$ has measure zero for each $i \neq j$.
(d) *If D_1 is a domain of integration contained in D, then* $\operatorname{Vol}(D_1) \leq \operatorname{Vol}(D)$, *with equality if and only if $D \smallsetminus D_1$ has measure zero.*

Proposition C.23 (Properties of Integrals). *Let $D \subseteq \mathbb{R}^n$ 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*

$$\int_D f\, dV = \int_{D_1} f\, dV + \dots + \int_{D_k} f\, dV.$$

(d) *If $f \geq 0$ on D, then $\int_D f\, dV \geq 0$, with equality if and only if $f \equiv 0$ on* $\operatorname{Int} D$.
(e) $(\inf_D f) \operatorname{Vol}(D) \leq \int_D f\, dV \leq (\sup_D f) \operatorname{Vol}(D)$.
(f) $\left| \int_D f\, dV \right| \leq \int_D |f|\, dV$.

► **Exercise C.24.** Prove Propositions C.22 and C.23.

Corollary C.25. *A set of measure zero in $\mathbb{R}^n$ contains no nonempty open subset.*

Proof. Assume for the sake of contradiction that $D \subseteq \mathbb{R}^n$ has measure zero and contains a nonempty open subset U. Then U contains a nonempty open rectangle, which has positive volume and therefore does not have measure zero by Proposition C.22. But this contradicts the fact that every subset of D has measure zero by Proposition C.18. □

There are two more fundamental properties of multiple integrals that we need. The proofs are too involved to be included in this summary, but you can look them up in [Apo74, Rud76, Str00] if you are interested. Each of these theorems can be stated in various ways, some stronger than others. The versions we give here are quite sufficient for our applications.

Theorem C.26 (Change of Variables). *Suppose D and E are open domains of integration in $\mathbb{R}^n$, and $G\colon \overline{D} \to \overline{E}$ is smooth map that restricts to a diffeomorphism from D to E. For every continuous function $f\colon \overline{E} \to \mathbb{R}$,*

$$\int_E f\, dV = \int_D (f \circ G)\, |\det DG|\, dV.$$

Theorem C.27 (Fubini's Theorem). *Let $A = \left[a^1, b^1\right] \times \dots \times \left[a^n, b^n\right]$ 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.

Integrals of Vector-Valued Functions

If $D \subseteq \mathbb{R}^n$ is a domain of integration and $F: D \to \mathbb{R}^k$ is a bounded continuous vector-valued function, we define the integral of F over D to be the vector in $\mathbb{R}^k$ obtained by integrating F component by component:

$$\int_D F\,dV = \left(\int_D F^1\,dV, \dots, \int_D F^k\,dV\right).$$

The analogues of parts (a)–(c) of Proposition C.23 obviously hold for vector-valued integrals, just by applying them to each component. Part (f) holds as well, but requires a bit more work to prove.

Proposition C.28. *Suppose $D \subseteq \mathbb{R}^n$ is a domain of integration and $F: D \to \mathbb{R}^k$ is a bounded continuous vector-valued function. Then*

$$\left|\int_D F\,dV\right| \le \int_D |F|\,dV. \tag{C.14}$$

Proof. Let G denote the vector $\int_D F\,dV \in \mathbb{R}^k$. If $G = 0$, then (C.14) obviously holds, so we may as well assume that $G \ne 0$. We compute

$$|G|^2 = \sum_{i=1}^k (G^i)^2 = \sum_{i=1}^k G^i \int_D F^i\,dV = \sum_{i=1}^k \int_D G^i F^i\,dV = \int_D (G \cdot F)\,dV.$$

Applying Proposition C.23(f) to the scalar integral $\int_D (G \cdot F)\,dV$, we obtain

$$|G|^2 \le \int_D |G \cdot F|\,dV \le \int_D |G|\,|F|\,dV = |G| \int_D |F|\,dV.$$

Dividing both sides of the inequality above by $|G|$ yields (C.14). □

As an application of (C.14), we prove an important estimate for the local behavior of a C^1 function in terms of its total derivative.

Proposition C.29 (Lipschitz Estimate for C^1 Functions). *Let $U \subseteq \mathbb{R}^n$ be an open subset, and suppose $F: U \to \mathbb{R}^m$ is of class C^1. Then F is Lipschitz continuous on every compact convex subset $K \subseteq U$. The Lipschitz constant can be taken to be* $\sup_{x \in K} |DF(x)|$.

Proof. Since $|DF(x)|$ is a continuous function of x, it is bounded on the compact set K. (The norm here is the Frobenius norm on matrices defined in (B.15).) Let $M = \sup_{x \in K} |DF(x)|$. For arbitrary $a, b \in K$, we have $a + t(b - a) \in K$ for all $t \in I$ because K is convex. By the fundamental theorem of calculus applied to each component of F, together with the chain rule,

$$F(b) - F(a) = \int_0^1 \frac{d}{dt} F\big(a + t(b-a)\big)\,dt$$

$$= \int_0^1 DF\big(a + t(b-a)\big)(b-a)\,dt.$$

Therefore, by (C.14) and Exercise B.48,

$$\begin{aligned}\big|F(b) - F(a)\big| &\le \int_0^1 \big|DF\big(a + t(b-a)\big)\big|\,|b-a|\,dt \\ &\le \int_0^1 M\,|b-a|\,dt = M|b-a|.\end{aligned}$$

□

Corollary C.30. *If $U \subseteq \mathbb{R}^n$ is an open subset and $F : U \to \mathbb{R}^m$ is of class C^1, then f is locally Lipschitz continuous.*

Proof. Each point of U is contained in a ball whose closure is contained in U, and Proposition C.29 shows that the restriction of F to such a ball is Lipschitz continuous. □

Sequences and Series of Functions

Let $S \subseteq \mathbb{R}^n$, and suppose we are given functions $f : S \to \mathbb{R}^m$ and $f_i : S \to \mathbb{R}^m$ for each integer $i \ge 1$. The sequence $(f_i)_{i=1}^{\infty}$ is said to ***converge pointwise to f*** if for each $a \in S$ and each $\varepsilon > 0$, there exists an integer N such that $i \ge N$ implies $|f_i(a) - f(a)| < \varepsilon$. 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 $i \ge N$ implies $|f_i(a) - f(a)| < \varepsilon$ for all $a \in S$. The sequence is ***uniformly Cauchy*** if for every $\varepsilon > 0$ there exists N such that $i, j \ge N$ implies $|f_i(a) - f_j(a)| < \varepsilon$ for all $a \in S$.

Theorem C.31 (Properties of Uniform Convergence). *Let $S \subseteq \mathbb{R}^n$, and suppose $f_i : S \to \mathbb{R}^m$ is continuous for each integer $i \ge 1$.*

(a) *If $f_i \to f$ uniformly, then f is continuous.*
(b) *If the sequence $(f_i)_{i=1}^{\infty}$ is uniformly Cauchy, then it converges uniformly to a continuous function.*
(c) *If $f_i \to f$ uniformly and S is a compact domain of integration, then*

$$\lim_{i\to\infty} \int_S f_i\,dV = \int_S f\,dV.$$

(d) *If S is open, each f_i is of class C^1, $f_i \to f$ pointwise, and $\big(\partial f_i/\partial x^j\big)$ 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 [Apo74, Rud76, Str00].

Given an infinite series of (real-valued or vector-valued) functions $\sum_{i=0}^{\infty} f_i$ on $S \subseteq \mathbb{R}^n$, one says ***the series converges pointwise*** if the corresponding sequence of partial sums converges pointwise to some function f:

$$f(x) = \lim_{N\to\infty} \sum_{i=0}^{N} f_i(x) \quad \text{for all } x \in S.$$

We say ***the series converges uniformly*** if its partial sums do so.

Proposition C.32 (Weierstrass M-test). *Suppose $S \subseteq \mathbb{R}^n$, and $f_i\colon S \to \mathbb{R}^k$ are functions. If there exist positive real numbers M_i such that $\sup_S |f_i| \le M_i$ and $\sum_i M_i$ converges, then $\sum_i f_i$ converges uniformly on S.*

► **Exercise C.33.** Prove Proposition C.32.

The Inverse and Implicit Function Theorems

The last two theorems in this appendix are central results about smooth functions. They say that under certain hypotheses, the local behavior of a smooth function is modeled by the behavior of its total derivative.

Theorem C.34 (Inverse Function Theorem). *Suppose U and V are open subsets of $\mathbb{R}^n$, and $F\colon U \to V$ is a smooth function. If $DF(a)$ is invertible at some point $a \in U$, then there exist connected neighborhoods $U_0 \subseteq U$ of a and $V_0 \subseteq V$ of $F(a)$ such that $F|_{U_0}\colon U_0 \to V_0$ is a diffeomorphism.*

The proof of this theorem is based on an elementary result about metric spaces, which we describe first.

Let X be a metric space. A map $G\colon X \to X$ is said to be a ***contraction*** if there is a constant $\lambda \in (0,1)$ such that $d\big(G(x), G(y)\big) \le \lambda d(x,y)$ for all $x, y \in X$. Clearly, every contraction is continuous. A ***fixed point*** of a map $G\colon X \to X$ is a point $x \in X$ such that $G(x) = x$.

Lemma C.35 (Contraction Lemma). *Let X be a nonempty complete metric space. Every contraction $G\colon X \to X$ has a unique fixed point.*

Proof. Uniqueness is immediate, for if x and x' are both fixed points of G, the contraction property implies $d(x,x') = d\big(G(x), G(x')\big) \le \lambda d(x,x')$, 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 $(x_n)_{n=0}^{\infty}$ inductively by $x_{n+1} = G(x_n)$. For any $i \ge 1$ we have $d(x_i, x_{i+1}) = d\big(G(x_{i-1}), G(x_i)\big) \le \lambda d(x_{i-1}, x_i)$, and therefore by induction

$$d(x_i, x_{i+1}) \le \lambda^i d(x_0, x_1).$$

If N is a positive integer and $j \ge i \ge N$,

$$d(x_i, x_j) \le d(x_i, x_{i+1}) + d(x_{i+1}, x_{i+2}) + \cdots + d(x_{j-1}, x_j)$$

$$
\begin{aligned}
&\le \left(\lambda^i + \cdots + \lambda^{j-1}\right) d(x_0, x_1) \\
&\le \lambda^i \left(\sum_{n=0}^{\infty} \lambda^n\right) d(x_0, x_1) \\
&\le \lambda^N \frac{1}{1-\lambda} d(x_0, x_1).
\end{aligned}
$$

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. We begin by making some simple modifications to the function F to streamline the proof. First, the function F_1 defined by

$$F_1(x) = F(x+a) - F(a)$$

is smooth on a neighborhood of 0 and satisfies $F_1(0) = 0$ and $DF_1(0) = DF(a)$; clearly, F is a diffeomorphism on a connected neighborhood of a if and only if F_1 is a diffeomorphism on a connected neighborhood of 0. Second, the function $F_2 = DF_1(0)^{-1} \circ F_1$ is smooth on the same neighborhood of 0 and satisfies $F_2(0) = 0$ and $DF_2(0) = I_n$; and if F_2 is a diffeomorphism in a neighborhood of 0, then so is F_1 and therefore also F. Henceforth, replacing F by F_2, we assume that F is defined in a neighborhood U of 0, $F(0) = 0$, and $DF(0) = I_n$. Because the determinant of $DF(x)$ is a continuous function of x, by shrinking U if necessary, we may assume that $DF(x)$ is invertible for each $x \in U$.

Let $H(x) = x - F(x)$ for $x \in U$. Then $DH(0) = I_n - I_n = 0$. Because the matrix entries of $DH(x)$ are continuous functions of x, there is a number $\delta > 0$ such that $B_\delta(0) \subseteq U$ and for all $x \in \overline{B}_\delta(0)$, we have $|DH(x)| \le \frac{1}{2}$. If $x, x' \in \overline{B}_\delta(0)$, the Lipschitz estimate for smooth functions (Proposition C.29) implies

$$|H(x') - H(x)| \le \tfrac{1}{2}|x' - x|. \tag{C.15}$$

In particular, taking $x' = 0$, this implies

$$|H(x)| \le \tfrac{1}{2}|x|. \tag{C.16}$$

Since $x' - x = F(x') - F(x) + H(x') - H(x)$, it follows that

$$|x' - x| \le \left|F(x') - F(x)\right| + \left|H(x') - H(x)\right| \le \left|F(x') - F(x)\right| + \tfrac{1}{2}|x' - x|,$$

and rearranging gives

$$|x' - x| \le 2|F(x') - F(x)| \tag{C.17}$$

for all $x, x' \in \overline{B}_\delta(0)$. In particular, this shows that F is injective on $\overline{B}_\delta(0)$.

Now let $y \in B_{\delta/2}(0)$ be arbitrary. We will show that there exists a unique point $x \in B_\delta(0)$ such that $F(x) = y$. Let $G(x) = y + H(x) = y + x - F(x)$, so that $G(x) = x$ if and only if $F(x) = y$. If $|x| \le \delta$, (C.16) implies

$$|G(x)| \le |y| + |H(x)| < \frac{\delta}{2} + \frac{1}{2}|x| \le \delta, \tag{C.18}$$

so G maps $\bar{B}_\delta(0)$ to itself. It then follows from (C.15) that $|G(x) - G(x')| = |H(x) - H(x')| \le \frac{1}{2}|x - x'|$, so G is a contraction. Since $\bar{B}_\delta(0)$ is a complete metric space (see Example A.6), the contraction lemma implies that G has a unique fixed point $x \in \bar{B}_\delta(0)$. From (C.18), $|x| = |G(x)| < \delta$, so in fact $x \in B_\delta(0)$, thus proving the claim.

Let $V_0 = B_{\delta/2}(0)$ and $U_0 = B_\delta(0) \cap F^{-1}(V_0)$. Then U_0 is open in $\mathbb{R}^n$, and the argument above shows that $F\colon U_0 \to V_0$ is bijective, so $F^{-1}\colon V_0 \to U_0$ exists. Substituting $x = F^{-1}(y)$ and $x' = F^{-1}(y')$ into (C.17) shows that F^{-1} is continuous. Thus $F\colon U_0 \to V_0$ is a homeomorphism, and it follows that U_0 is connected because V_0 is.

The only thing that remains to be proved is that F^{-1} is smooth. If we knew it were smooth, Proposition C.4 would imply that $D\left(F^{-1}\right)(y) = DF(x)^{-1}$, where $x = F^{-1}(y)$. We begin by showing that F^{-1} is differentiable at each point of V_0, with total derivative given by this formula.

Let $y \in V_0$ be arbitrary, and set $x = F^{-1}(y)$ and $L = DF(x)$. We need to show that

$$\lim_{y' \to y} \frac{F^{-1}(y') - F^{-1}(y) - L^{-1}(y' - y)}{|y' - y|} = 0.$$

Given $y' \in V_0 \smallsetminus \{y\}$, write $x' = F^{-1}(y') \in U_0 \smallsetminus \{x\}$. Then

$$\begin{aligned}
&\frac{F^{-1}(y') - F^{-1}(y) - L^{-1}(y' - y)}{|y' - y|} \\
&\quad = L^{-1}\left(\frac{L(x' - x) - (y' - y)}{|y' - y|}\right) \\
&\quad = \frac{|x' - x|}{|y' - y|} L^{-1}\left(-\frac{F(x') - F(x) - L(x' - x)}{|x' - x|}\right).
\end{aligned}$$

The factor $|x' - x|/|y' - y|$ above is bounded thanks to (C.17), and because L^{-1} is linear and therefore bounded (Exercise B.52), the norm of the second factor is bounded by a constant multiple of

$$\frac{|F(x') - F(x) - L(x' - x)|}{|x' - x|}. \tag{C.19}$$

As $y' \to y$, it follows that $x' \to x$ by continuity of F^{-1}, and then (C.19) goes to zero because $L = DF(x)$ and F is differentiable. This completes the proof that F^{-1} is differentiable.

By Proposition C.8, the partial derivatives of F^{-1} are defined at each point $y \in V_0$. Observe that the formula $D\big(F^{-1}\big)(y) = DF\left(F^{-1}(y)\right)^{-1}$ implies that the matrix-valued function $y \mapsto D\big(F^{-1}\big)(y)$ can be written as the composition

$$y \overset{F^{-1}}{\longmapsto} F^{-1}(x) \overset{DF}{\longmapsto} DF\left(F^{-1}(y)\right) \overset{i}{\longmapsto} DF\left(F^{-1}(y)\right)^{-1}, \tag{C.20}$$

where i is matrix inversion. In this composition, F^{-1} is continuous; DF is smooth because its component functions are the partial derivatives of F; and i is smooth because Cramer's rule expresses the entries of an inverse matrix as rational functions of the entries of the matrix. Because $D\big(F^{-1}\big)$ is a composition of continuous functions, it is continuous. Thus the partial derivatives of F^{-1} are continuous, so F^{-1} is of class C^1.

Now assume by induction that we have shown that F^{-1} is of class C^k. This means that each of the functions in (C.20) is of class C^k. Because $D\big(F^{-1}\big)$ 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. □

Corollary C.36. *Suppose $U \subseteq \mathbb{R}^n$ is an open subset, and $F\colon U \to \mathbb{R}^n$ is a smooth function whose Jacobian determinant is nonzero at every point in U.*

(a) *F is an open map.*
(b) *If F is injective, then $F\colon U \to F(U)$ is a diffeomorphism.*

Proof. For each $a \in U$, the fact that the Jacobian determinant of F is nonzero implies that $DF(a)$ is invertible, so the inverse function theorem implies that there exist open subsets $U_a \subseteq U$ containing a and $V_a \subseteq F(U)$ containing $F(a)$ such that F restricts to a diffeomorphism $F|_{U_a}\colon U_a \to V_a$. In particular, this means that each point of $F(U)$ has a neighborhood contained in $F(U)$, so $F(U)$ is open. If $U_0 \subseteq U$ is an arbitrary open subset, the same argument with U replaced by U_0 shows that $F(U_0)$ is also open; this proves (a). If in addition F is injective, then the inverse map $F^{-1}\colon F(U) \to U$ exists for set-theoretic reasons; on a neighborhood of each point $F(a) \in F(U)$ it is equal to the inverse of $F|_{U_a}$, so it is smooth. □

The next two examples illustrate the use of the preceding corollary.

Example C.37 (Polar Coordinates). As you know from calculus, ***polar coordinates*** (r,θ) in the plane are defined implicitly by the relations $x = r\cos\theta$, $y = r\sin\theta$. The map $F\colon (0,\infty)\times\mathbb{R} \to \mathbb{R}^2$ defined by $F(r,\theta) = (r\cos\theta, r\sin\theta)$ is smooth and has Jacobian determinant equal to r, which is nonzero everywhere on the domain. Thus, Corollary C.36 shows that the restriction of F to any open subset on which it is injective is a diffeomorphism onto its image. One such subset is $\{(r,\theta) : r > 0,\ -\pi < \theta < \pi\}$, which is mapped bijectively by F onto the complement of the nonpositive part of the x-axis. //

Example C.38 (Spherical Coordinates). Similarly, ***spherical coordinates*** on $\mathbb{R}^3$ are the functions (ρ,φ,θ) defined by the relations

$$x = \rho\sin\varphi\cos\theta,$$

$$y = \rho \sin\varphi \sin\theta,$$
$$z = \rho \cos\varphi.$$

Geometrically, ρ is the distance from the origin, φ is the angle from the positive z-axis, and θ is the angle from the $x > 0$ half of the (x,z)-plane. If we define $G\colon (0,\infty) \times (0,\pi) \times \mathbb{R} \to \mathbb{R}^3$ by $G(\rho,\varphi,\theta) = (\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\varphi)$, a computation shows that the Jacobian determinant of G is $\rho^2 \sin\varphi \neq 0$. Thus, the restriction of G to any open subset on which it is injective is a diffeomorphism onto its image. One such subset is

$$\{(\rho,\varphi,\theta) : \rho > 0,\ 0 < \varphi < \pi, -\pi < \theta < \pi\}.$$

Notice how much easier it is to argue this way than to try to construct an inverse map explicitly out of inverse trigonometric functions. //

► **Exercise C.39.** Verify the claims in the preceding two examples.

The next result is one of the most important consequences of the inverse function theorem. It gives conditions under which a level set of a smooth function is locally the graph of a smooth function.

Theorem C.40 (Implicit Function Theorem). *Let $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ be an open subset, and let $(x,y) = \left(x^1,\dots,x^n,y^1,\dots,y^k\right)$ denote the standard coordinates on U. Suppose $\Phi\colon U \to \mathbb{R}^k$ is a smooth function, $(a,b) \in U$, 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 $V_0 \subseteq \mathbb{R}^n$ of a and $W_0 \subseteq \mathbb{R}^k$ of b and a smooth function $F\colon V_0 \to W_0$ such that $\Phi^{-1}(c) \cap (V_0 \times W_0)$ is the graph of F, that is, $\Phi(x,y) = c$ for $(x,y) \in V_0 \times W_0$ if and only if $y = F(x)$.

Proof. Consider the smooth function $\Psi\colon U \to \mathbb{R}^n \times \mathbb{R}^k$ defined by $\Psi(x,y) = \big(x, \Phi(x,y)\big)$. 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 because it is block lower triangular and the two blocks on the main diagonal are nonsingular. 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 that $U_0 = V \times W$ is a product neighborhood.

Writing $\Psi^{-1}(x,y) = \big(A(x,y), B(x,y)\big)$ for some smooth functions A and B, we compute

$$\begin{aligned}(x,y) &= \Psi\big(\Psi^{-1}(x,y)\big) = \Psi\big(A(x,y), B(x,y)\big)\\ &= \Big(A(x,y), \Phi\big(A(x,y), B(x,y)\big)\Big).\end{aligned} \tag{C.21}$$

Comparing the first components in this equation, we find that $A(x,y) = x$, so Ψ^{-1} has the form $\Psi^{-1}(x,y) = \big(x, B(x,y)\big)$.

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 the second components in (C.21) yields

$$c = \Phi\big(x, B(x,c)\big) = \Phi\big(x, F(x)\big)$$

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 $\Psi(x,y) = \big(x, \Phi(x,y)\big) = (x,c)$, so

$$(x,y) = \Psi^{-1}(x,c) = \big(x, B(x,c)\big) = \big(x, F(x)\big),$$

which implies that $y = F(x)$. This completes the proof. □

Appendix D
Review of Differential Equations

The theory of ordinary differential equations (ODEs) underlies much of the study of smooth manifolds. In this appendix, we review both the theoretical and the practical aspects of the subject. Since we need to work only with first-order equations and systems, we concentrate our attention on those. For more detail, consult any good ODE textbook, such as [BR89] or [BD09].

Existence, Uniqueness, and Smoothness

Here is the general setting in which ODEs appear in this book: we are given n real-valued continuous functions $V^1, \dots, V^n$ defined on some open subset $W \subseteq \mathbb{R}^{n+1}$, and the goal is to find differentiable real-valued functions $y^1, \dots, y^n$ solving the following ***initial value problem***:

$$\dot{y}^i(t) = V^i\big(t, y^1(t), \dots, y^n(t)\big), \quad i = 1, \dots, n, \tag{D.1}$$

$$y^i(t_0) = c^i, \qquad i = 1, \dots, n, \tag{D.2}$$

where $(t_0, c^1, \dots, c^n)$ is an arbitrary point in W. (Here and elsewhere in the book, we use a dot to denote an ordinary derivative with respect to t whenever convenient, primarily when there are superscripts that would make the prime notation cumbersome.)

The fundamental fact about ordinary differential equations is that for smooth equations, there always exists a unique solution to the initial value problem, at least for a short time, and the solution is a smooth function of the initial conditions as well as time. The existence and uniqueness parts of this theorem are proved in most ODE textbooks, but the smoothness part is often omitted. Because this result is so fundamental to smooth manifold theory, we give a complete proof here.

Most of our applications of the theory are confined to the following special case: if the functions V^i on the right-hand side of (D.1) do not depend explicitly on t, the system is said to be ***autonomous***; otherwise, it is ***nonautonomous***. We begin by stating and proving our main theorem in the autonomous case. Afterwards, we show how the general case follows from this one.

J.M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics 218,
DOI 10.1007/978-1-4419-9982-5,

Theorem D.1 (Fundamental Theorem for Autonomous ODEs). *Suppose $U \subseteq \mathbb{R}^n$ is open, and $V: U \to \mathbb{R}^n$ is a smooth vector-valued function. Consider the initial value problem*

$$\dot{y}^i(t) = V^i\big(y^1(t), \dots, y^n(t)\big), \quad i = 1, \dots, n, \tag{D.3}$$

$$y^i(t_0) = c^i, \quad i = 1, \dots, n, \tag{D.4}$$

for arbitrary $t_0 \in \mathbb{R}$ and $c = (c^1, \dots, c^n) \in U$.

(a) EXISTENCE: *For any $t_0 \in \mathbb{R}$ and $x_0 \in U$, there exist an open interval J_0 containing t_0 and an open subset $U_0 \subseteq U$ containing x_0 such that for each $c \in U_0$, there is a C^1 map $y: J_0 \to U$ that solves* (D.3)–(D.4).
(b) UNIQUENESS: *Any two differentiable solutions to* (D.3)–(D.4) *agree on their common domain.*
(c) SMOOTHNESS: *Let J_0 and U_0 be as in* (a), *and let $\theta: J_0 \times U_0 \to U$ be the map defined by $\theta(t, x) = y(t)$, where $y: J_0 \to U$ is the unique solution to* (D.3) *with initial condition $y(t_0) = x$. Then θ is smooth.*

The existence, uniqueness, and smoothness parts of this theorem will be proved separately below. The following comparison theorem is useful in the proofs to follow.

Theorem D.2 (ODE Comparison Theorem). *Let $J \subseteq \mathbb{R}$ be an open interval, and suppose the differentiable function $u: J \to \mathbb{R}^n$ satisfies the following differential inequality for all $t \in J$:*

$$|u'(t)| \le f\big(|u(t)|\big),$$

where $f: [0, \infty) \to [0, \infty)$ is Lipschitz continuous. If for some $t_0 \in J$, $v: [0, \infty) \to [0, \infty)$ is a differentiable real-valued function satisfying the initial-value problem

$$v'(t) = f\big(v(t)\big),$$
$$v(0) = |u(t_0)|,$$

then the following inequality holds for all $t \in J$:

$$|u(t)| \le v\big(|t - t_0|\big). \tag{D.5}$$

Proof. Assume first that $t_0 = 0$, and let $J^+ = \{t \in J : t \ge 0\}$. We begin by proving that (D.5) holds for all $t \in J^+$. On the open subset of J^+ where $|u(t)| > 0$, $|u(t)|$ is a differentiable function of t, and the Cauchy–Schwarz inequality shows that

$$\begin{aligned}\frac{d}{dt}|u(t)| &= \frac{d}{dt}\big(u(t) \cdot u(t)\big)^{1/2} = \frac{1}{2}\big(u(t) \cdot u(t)\big)^{-1/2}\big(2u(t) \cdot u'(t)\big) \\ &\le \frac{1}{2}|u(t)|^{-1}\left(2\,|u(t)|\,\left|u'(t)\right|\right) = \left|u'(t)\right| \le f\big(|u(t)|\big).\end{aligned}$$

Let A be a Lipschitz constant for f, and consider the continuous function $w: J^+ \to \mathbb{R}$ defined by

$$w(t) = e^{-At}\big(|u(t)| - v(t)\big).$$

Then $w(0) = 0$, and (D.5) for $t \in J^+$ is equivalent to $w(t) \le 0$.

At any $t \in J^+$ such that $w(t) > 0$ (and therefore $|u(t)| > v(t) \ge 0$), w is differentiable and satisfies

$$\begin{aligned} w'(t) &= -Ae^{-At}\big(|u(t)| - v(t)\big) + e^{-At}\frac{d}{dt}\big(|u(t)| - v(t)\big) \\ &\le -Ae^{-At}\big(|u(t)| - v(t)\big) + e^{-At}\big(f\big(|u(t)|\big) - f\big(v(t)\big)\big) \\ &\le 0, \end{aligned}$$

where the last inequality follows from the Lipschitz estimate for f.

Now suppose there is some $t_1 \in J^+$ such that $w(t_1) > 0$. Let

$$\tau = \sup\big\{t \in [0, t_1] : w(t) \le 0\big\}.$$

Then $w(\tau) = 0$ by continuity, and $w(t) > 0$ for $t \in (\tau, t_1]$. Since w is continuous on $[\tau, t_1]$ and differentiable on (τ, t_1), the mean value theorem implies that there must exist $t \in (\tau, t_1)$ such that $w(t) > 0$ and $w'(t) > 0$. But this contradicts the calculation above, which showed that $w'(t) \le 0$ whenever $w(t) > 0$, thus proving that $w(t) \le 0$ for all $t \in J^+$.

Now, the result for $t \le 0$ follows easily by substituting $-t$ for t in the argument above. Finally, for the general case in which $t_0 \ne 0$, we simply apply the above argument to the function $\widetilde{u}(t) = u(t + t_0)$ on the interval $\widetilde{J} = \{t : t + t_0 \in J\}$. □

Remark. In the statement of the comparison theorem, we have assumed for simplicity that both f and v are defined for all nonnegative t, but these hypotheses can be weakened: the proof goes through essentially without modification as long as v is defined on an interval $[0, b)$ large enough that $J \subseteq (t_0 - b, t_0 + b)$, and f is defined on some interval that contains $|u(t)|$ and $v(t)$ for all $t \in J$.

Theorem D.3 (Existence of ODE Solutions). *Let $U \subseteq \mathbb{R}^n$ be an open subset, and suppose $V: U \to \mathbb{R}^n$ is locally Lipschitz continuous. Let $(t_0, x_0) \in \mathbb{R} \times U$ be given. There exist an open interval $J_0 \subseteq \mathbb{R}$ containing t_0, an open subset $U_0 \subseteq U$ containing x_0, and for each $c \in U_0$, a C^1 map $y: J_0 \to U$ satisfying the initial value problem* (D.3)–(D.4).

Proof. By shrinking U if necessary, we may assume that V is Lipschitz continuous on U. We begin by showing that the system (D.3)–(D.4) is equivalent to a certain integral equation. Suppose y is any solution to (D.3)–(D.4) on some interval J_0 containing t_0. Because y is differentiable, it is continuous, and then the fact that the right-hand side of (D.3) is a continuous function of t implies that y is of class C^1. Integrating (D.3) with respect to t and applying the fundamental theorem of calculus shows that y satisfies the following integral equation:

$$y^i(t) = c^i + \int_{t_0}^{t} V^i\big(y(s)\big)\,ds. \tag{D.6}$$

Conversely, if $y: J_0 \to U$ is a continuous map satisfying (D.6), then the fundamental theorem of calculus implies that y satisfies (D.3)–(D.4) and therefore is actually of class C^1.

This motivates the following definition. Suppose J_0 is an open interval containing t_0. For any continuous map $y\colon J_0 \to U$, we define a new map $Iy\colon J_0 \to \mathbb{R}^n$ by

$$Iy(t) = c + \int_{t_0}^{t} V\big(y(s)\big)\,ds. \tag{D.7}$$

Then we are led to seek a fixed point for I in a suitable metric space of maps.

Let C be a Lipschitz constant for V, so that

$$\big|V(y) - V\big(\tilde{y}\big)\big| \le C\big|y - \tilde{y}\big|, \quad y, \tilde{y} \in U. \tag{D.8}$$

Given $t_0 \in \mathbb{R}$ and $x_0 \in U$, choose $r > 0$ small enough that $\overline{B}_r(x_0) \subseteq U$. Let M be the supremum of $|V|$ on the compact set $\overline{B}_r(x_0)$. Choose $\delta > 0$ and $\varepsilon > 0$ small enough that

$$\delta < \frac{r}{2}, \qquad \varepsilon < \min\left(\frac{r}{2M}, \frac{1}{C}\right),$$

and set $J_0 = (t_0 - \varepsilon, t_0 + \varepsilon) \subseteq \mathbb{R}$ and $U_0 = B_\delta(x_0) \subseteq U$. For any $c \in U_0$, let $\mathcal{M}_c$ denote the set of all continuous maps $y\colon J_0 \to \overline{B}_r(x_0)$ satisfying $y(t_0) = c$. We define a metric on $\mathcal{M}_c$ by

$$d(y, \tilde{y}) = \sup_{t \in J_0} \big|y(t) - \tilde{y}(t)\big|.$$

Any sequence of maps in $\mathcal{M}_c$ that is Cauchy in this metric is uniformly Cauchy, and thus converges to a continuous limit y. Clearly, the conditions that y take its values in $\overline{B}_r(x_0)$ and $y(t_0) = c$ are preserved in the limit. Therefore, $\mathcal{M}_c$ is a complete metric space.

We wish to define a map $I\colon \mathcal{M}_c \to \mathcal{M}_c$ by formula (D.7). The first thing we need to verify is that I really does map $\mathcal{M}_c$ into itself. It is clear from the definition that $Iy(t_0) = c$ and Iy is continuous (in fact, it is differentiable by the fundamental theorem of calculus). Thus, we need only check that Iy takes its values in $\overline{B}_r(x_0)$. If $y \in \mathcal{M}_c$, then for any $t \in J_0$,

$$\begin{aligned}
|Iy(t) - x_0| &= \left|c + \int_{t_0}^{t} V\big(y(s)\big)\,ds - x_0\right| \\
&\le |c - x_0| + \int_{t_0}^{t} \big|V\big(y(s)\big)\big|\,ds \\
&< \delta + M\varepsilon < r
\end{aligned}$$

by our choice of δ and ε.

Next we check that I is a contraction (see p. 657). If $y, \tilde{y} \in \mathcal{M}_c$, then

$$\begin{aligned}
d\big(Iy, I\tilde{y}\big) &= \sup_{t \in J_0} \left|\int_{t_0}^{t} V\big(y(s)\big)\,ds - \int_{t_0}^{t} V\big(\tilde{y}(s)\big)\,ds\right| \\
&\le \sup_{t \in J_0} \int_{t_0}^{t} \big|V\big(y(s)\big) - V\big(\tilde{y}(s)\big)\big|\,ds
\end{aligned}$$

$$\leq \sup_{t \in J_0} \int_{t_0}^{t} C \left| y(s) - \widetilde{y}(s) \right| ds \leq C\varepsilon d\left(y, \widetilde{y}\right).$$

Because we have chosen ε so that $C\varepsilon < 1$, this shows that I is a contraction. By the contraction lemma (Lemma C.35), I has a fixed point $y \in \mathcal{M}_c$, which is a solution to (D.6) and thus also to (D.3)–(D.4). □

Theorem D.4 (Uniqueness of ODE Solutions). *Let $U \subseteq \mathbb{R}^n$ be an open subset, and suppose $V : U \to \mathbb{R}^n$ is locally Lipschitz continuous. For any $t_0 \in \mathbb{R}$ and $c \in U$, any two differentiable solutions to* (D.3)–(D.4) *are equal on their common domain.*

Proof. Suppose first that $y, \widetilde{y} \colon J_0 \to U$ are two differentiable functions that both satisfy (D.3) on the same open interval $J_0 \subseteq \mathbb{R}$, but not necessarily with the same initial conditions. Let J_1 be a bounded open interval containing t_0 such that $\overline{J}_1 \subseteq J_0$. The union of $y\left(\overline{J}_1\right)$ and $\widetilde{y}\left(\overline{J}_1\right)$ is a compact subset of U, and Proposition A.48(b) shows that there is a Lipschitz constant C for V on that set. Thus

$$\left| \frac{d}{dt}\left(\widetilde{y}(t) - y(t)\right) \right| = \left| V\left(\widetilde{y}(t)\right) - V\left(y(t)\right) \right| \leq C \left| \widetilde{y}(t) - y(t) \right|.$$

Applying the ODE comparison theorem (Theorem D.2) with $u(t) = \widetilde{y}(t) - y(t)$, $f(v) = Cv$, and $v(t) = e^{Ct}\left|\widetilde{y}(t_0) - y(t_0)\right|$, we conclude that

$$\left| \widetilde{y}(t) - y(t) \right| \leq e^{C|t - t_0|} \left| \widetilde{y}(t_0) - y(t_0) \right|, \quad t \in \overline{J}_1. \tag{D.9}$$

Thus, $y(t_0) = \widetilde{y}(t_0)$ implies $y \equiv \widetilde{y}$ on all of $\overline{J}_1$. Since every point of J_0 is contained in some such subinterval J_1, it follows that $y \equiv \widetilde{y}$ on all of J_0. □

Theorem D.5 (Smoothness of ODE Solutions). *Suppose $U \subseteq \mathbb{R}^n$ is an open subset and $V : U \to \mathbb{R}^n$ is locally Lipschitz continuous. Suppose also that $U_0 \subseteq U$ is an open subset, $J_0 \subseteq \mathbb{R}$ is an open interval containing t_0, and $\theta \colon J_0 \times U_0 \to U$ is a map such that for each $x \in U_0$, $y(t) = \theta(t, x)$ solves the initial value problem* (D.3)–(D.4) *with initial condition $c = x$. If V is of class C^k for some $k \geq 0$, then so is θ.*

Proof. Let $(t_1, x_1) \in J_0 \times U_0$ be arbitrary. It suffices to prove that θ is C^k on some neighborhood of (t_1, x_1). We prove this claim by induction on k.

Let J_1 be a bounded open interval containing t_0 and t_1 such that $\overline{J}_1 \subseteq J_0$. Because the restriction of θ to $J_0 \times \{x_1\}$ is an integral curve of V, it is continuous and therefore the set $K = \theta\left(\overline{J}_1 \times \{x_1\}\right)$ is compact. Thus, there exists $c > 0$ such that $\overline{B}_{2c}(y) \subseteq U$ for every $y \in K$. Let $W = \bigcup_{y \in K} B_c(y)$, so that W is a precompact neighborhood of K in U. The restriction of V to $\overline{W}$ is bounded by compactness, and is Lipschitz continuous by Proposition A.48(b). Let C be a Lipschitz constant for V on $\overline{W}$, and define constants M and T by

$$M = \sup_{\overline{W}} |V|, \qquad T = \sup_{t \in \overline{J}_1} |t - t_0|.$$

For any $x, \widetilde{x} \in W$, both $t \mapsto \theta(t,x)$ and $t \mapsto \theta(t,\widetilde{x})$ are integral curves of V for $t \in J_1$. As long as both curves stay in W, (D.9) implies

$$|\theta(t,x) - \theta(t,\widetilde{x})| \leq e^{CT} |\widetilde{x} - x|. \tag{D.10}$$

Choose $r > 0$ such that $2re^{CT} < c$, and let $U_1 = B_r(x_1)$ and $U_2 = B_{2r}(x_1)$. We will prove that θ maps $\bar{J}_1 \times \bar{U}_2$ into W. Assume not, which means there is some $(t_2, x_2) \in \bar{J}_1 \times \bar{U}_2$ such that $\theta(t_2, x_2) \notin W$; for simplicity, assume $t_2 > t_0$. Let τ be the infimum of times $t > t_0$ in J_1 such that $\theta(t, x_2) \notin \overline{W}$. By continuity, this means $\theta(\tau, x_2) \in \partial W$. But because both $\theta(t, x_1)$ and $\theta(t, x_2)$ are in W for $t \in [t_0, \tau]$, (D.10) yields $|\theta(\tau, x_2) - \theta(\tau, x_1)| \leq 2re^{CT} < c$, which means that $\theta(\tau, x_2) \in W$, a contradiction. This proves the claim.

For the $k = 0$ step, we need to show that θ is continuous on $\bar{J}_1 \times \bar{U}_1$. It follows from (D.10) that it is Lipschitz continuous there as a function of x. We need to show that it is jointly continuous in (t, x).

Let $(t, x) \in \bar{J}_1 \times \bar{U}_1$ be arbitrary. Since every solution to the initial value problem satisfies the integral equation (D.6), we find that

$$\theta^i(t,x) = x^i + \int_{t_0}^{t} V^i\big(\theta(s,x)\big)\, ds, \tag{D.11}$$

and therefore (assuming for simplicity that $t_1 \geq t$),

$$\begin{aligned}
|\theta(t_1,x_1) - \theta(t,x)| &\leq |x_1 - x| + \left| \int_{t_0}^{t_1} V\big(\theta(s,x_1)\big)\, ds - \int_{t_0}^{t} V\big(\theta(s,x)\big)\, ds \right| \\
&\leq |x_1 - x| + \int_{t_0}^{t} \big| V\big(\theta(s,x_1)\big) - V\big(\theta(s,x)\big) \big|\, ds \\
&\quad + \int_{t}^{t_1} \big| V\big(\theta(s,x_1)\big) \big|\, ds \\
&\leq |x_1 - x| + C \int_{t_0}^{t} |\theta(s,x_1) - \theta(s,x)|\, ds + \int_{t}^{t_1} M\, ds \\
&\leq |x_1 - x| + CTe^{CT} |x_1 - x| + M\,|t_1 - t|.
\end{aligned}$$

It follows that θ is continuous at (t_1, x_1).

Next we tackle the $k = 1$ step, which is the hardest part of the proof. Suppose that V is of class C^1, and let $\bar{J}_1$, $\bar{U}_1$ be defined as above. Expressed in terms of θ, the initial value problem (D.3)–(D.4) with $c = x$ reads

$$\begin{aligned}
\frac{\partial \theta^i}{\partial t}(t,x) &= V^i\big(\theta(t,x)\big), \\
\theta^i(t_0,x) &= x^i.
\end{aligned} \tag{D.12}$$

Because we know that θ is continuous by the argument above, this shows that $\partial\theta^i/\partial t$ is continuous. We will prove that for each j, $\partial\theta^i/\partial x^j$ exists and is continuous on $\bar{J}_1 \times \bar{U}_1$.

For any real number h such that $0 < |h| < r$ and any indices $i, j \in \{1, \dots, n\}$, we let $(\Delta_h)^i_j\colon \bar{J}_1 \times \bar{U}_1 \to \mathbb{R}$ 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. In fact, we will show that $(\Delta_h)^i_j$ converges uniformly on $\bar{J}_1 \times \bar{U}_1$ as $h \to 0$, from which it follows that $\partial\theta^i/\partial x^j$ exists and is continuous there, because it is a uniform limit of continuous functions. Let $\Delta_h\colon \bar{J}_1 \times \bar{U}_1 \to \mathrm{M}(n,\mathbb{R})$ be the matrix-valued function whose matrix entries are $(\Delta_h)^i_j(t,x)$. Note that (D.10) implies $|(\Delta_h)^i_j(t,x)| \le e^{CT}$ for each i and j, and thus

$$|\Delta_h(t,x)| \le n e^{CT}, \tag{D.13}$$

where the norm on the left-hand side is the Frobenius norm on matrices.

Let us compute the derivative of $(\Delta_h)^i_j$ with respect to t:

$$\begin{aligned}\frac{\partial}{\partial t}(\Delta_h)^i_j(t,x) &= \frac{1}{h}\left(\frac{\partial\theta^i}{\partial t}(t, x + he_j) - \frac{\partial\theta^i}{\partial t}(t,x)\right)\\ &= \frac{1}{h}\Big(V^i\big(\theta(t, x + he_j)\big) - V^i\big(\theta(t,x)\big)\Big).\end{aligned} \tag{D.14}$$

The mean value theorem applied to the C^1 function

$$u(s) = V^i\big((1-s)\theta(t,x) + s\theta(t, x + he_j)\big)$$

implies that there is some $c \in (0,1)$ such that $u(1) - u(0) = u'(c)$. If we substitute $y_0 = (1-c)\theta(t,x) + c\theta(t, x + he_j)$ (a point on the line segment between $\theta(t,x)$ and $\theta(t, x + he_j)$), this becomes

$$\begin{aligned}V^i\big(\theta(t, x + he_j)\big) - V^i\big(\theta(t,x)\big) &= \sum_{k=1}^n \frac{\partial V^i}{\partial y^k}(y_0)\big(\theta^k(t, x + he_j) - \theta^k(t,x)\big)\\ &= h\sum_{k=1}^n \frac{\partial V^i}{\partial y^k}(y_0)(\Delta_h)^k_j(t,x).\end{aligned}$$

Inserting this into (D.14) yields

$$\frac{\partial}{\partial t}(\Delta_h)^i_j(t,x) = \sum_{k=1}^n \frac{\partial V^i}{\partial y^k}(y_0)(\Delta_h)^k_j(t,x).$$

Thus for any sufficiently small nonzero real numbers $h, \widetilde{h}$,

$$\begin{aligned}
&\frac{\partial}{\partial t}\big((\Delta_h)^i_j(t,x) - (\Delta_{\widetilde{h}})^i_j(t,x)\big) \\
&\quad = \sum_{k=1}^n \frac{\partial V^i}{\partial y^k}(y_0)(\Delta_h)^k_j(t,x) - \sum_{k=1}^n \frac{\partial V^i}{\partial y^k}(\widetilde{y}_0)(\Delta_{\widetilde{h}})^k_j(t,x) \\
&\quad = \sum_{k=1}^n \frac{\partial V^i}{\partial y^k}(y_0)\big((\Delta_h)^k_j(t,x) - (\Delta_{\widetilde{h}})^k_j(t,x)\big) \\
&\qquad + \sum_{k=1}^n \left(\frac{\partial V^i}{\partial y^k}(y_0) - \frac{\partial V^i}{\partial y^k}\big(\widetilde{y}_0\big)\right)(\Delta_{\widetilde{h}})^k_j(t,x),
\end{aligned} \tag{D.15}$$

where $\widetilde{y}_0$ is defined similarly to y_0, but with $\widetilde{h}$ in place of h.

Now let $\varepsilon > 0$ be given. Because the continuous functions $\partial V^i/\partial y^k$ are uniformly continuous on $\bar{U}_1$ (Proposition A.48(a)), there exists $\delta > 0$ such that the following inequality holds whenever $|y_1 - y_2| < \delta$:

$$\left|\frac{\partial V^i}{\partial y^k}(y_1) - \frac{\partial V^i}{\partial y^k}(y_2)\right| < \varepsilon.$$

Suppose $|h|$ and $\big|\widetilde{h}\big|$ are both less than $\delta e^{-CT}/n$. Then we have

$$|y_0 - \theta(t,x)| = c|\theta(t,x+he_j) - \theta(t,x)| \le c|h||\Delta_h(t,x)| < \delta, \tag{D.16}$$

and similarly $\big|\widetilde{y}_0 - \theta(t,x)\big| < \delta$, so

$$\begin{aligned}
&\left|\frac{\partial V^i}{\partial y^k}(y_0) - \frac{\partial V^i}{\partial y^k}\big(\widetilde{y}_0\big)\right| \\
&\quad \le \left|\frac{\partial V^i}{\partial y^k}(y_0) - \frac{\partial V^i}{\partial y^k}\big(\theta(t,x)\big)\right| + \left|\frac{\partial V^i}{\partial y^k}\big(\theta(t,x)\big) - \frac{\partial V^i}{\partial y^k}\big(\widetilde{y}_0\big)\right| < 2\varepsilon.
\end{aligned} \tag{D.17}$$

Inserting (D.17) and (D.13) into (D.15), we find that the matrix-valued function $\Delta_h - \Delta_{\widetilde{h}}$ satisfies the following differential inequality:

$$\left|\frac{\partial}{\partial t}\Big(\Delta_h(t,x) - \Delta_{\widetilde{h}}(t,x)\Big)\right| \le E|\Delta_h(t,x) - \Delta_{\widetilde{h}}(t,x)| + 2\varepsilon n e^{CT},$$

where E is the supremum of $|DV|$ on $\bar{U}_1$. Note that $\theta^i(t_0,x) = x^i$ implies that $(\Delta_h)^i_j$ satisfies the following initial condition:

$$(\Delta_h)^i_j(t_0,x) = \frac{\theta^i(t_0,x+he_j) - \theta^i(t_0,x)}{h} = \frac{(x^i + h\delta^i_j) - x^i}{h} = \delta^i_j. \tag{D.18}$$

Thus, $\Delta_h(t_0,x) - \Delta_{\widetilde{h}}(t_0,x) = 0$, and we can apply the ODE comparison theorem with $f(v) = Ev + B$ and $v(t) = (B/E)\big(e^{Et} - 1\big)$ (where $B = 2\varepsilon n e^{CT}$) to conclude that

$$\left|\Delta_h(t,x) - \Delta_{\widetilde{h}}(t,x)\right| \leq \frac{2\varepsilon n e^{CT}}{E}\left(e^{E|t-t_0|} - 1\right) \leq \frac{2\varepsilon n e^{CT}}{E}\left(e^{ET} - 1\right). \quad \text{(D.19)}$$

Since the expression on the right can be made as small as desired by choosing h and $\widetilde{h}$ sufficiently small, this shows that for each i and j, and any sequence $h_k \to 0$, the sequence of functions $\left((\Delta_{h_k})^i_j\right)_{k=1}^\infty$ is uniformly Cauchy and therefore uniformly convergent to a continuous limit function. It follows easily from (D.19) that the limit is independent of the choice of (h_k), so the limit is in fact equal to $\lim_{h\to 0}(\Delta_h)^i_j(t,x)$, which is $\partial\theta^i/\partial x^j(t,x)$ by definition. This shows that θ^i has continuous first partial derivatives, and completes the proof of the $k=1$ case.

Now assume that the theorem is true for some $k \geq 1$, and suppose V is of class C^{k+1}. By the inductive hypothesis, θ is of class C^k, and therefore by (D.12), $\partial\theta^i/\partial t$ is also C^k. We can differentiate under the integral sign in (D.11) to obtain

$$\frac{\partial\theta^i}{\partial x^j}(t,x) = \delta^i_j + \sum_{k=1}^n \int_{t_0}^t \frac{\partial V^i}{\partial y^k}\big(\theta(s,x)\big)\frac{\partial\theta^k}{\partial x^j}(s,x)\,ds.$$

By the fundamental theorem of calculus, this implies that $\partial\theta^i/\partial x^j$ satisfies the differential equation

$$\frac{\partial}{\partial t}\frac{\partial\theta^i}{\partial x^j}(t,x) = \sum_{k=1}^n \frac{\partial V^i}{\partial y^k}\big(\theta(t,x)\big)\frac{\partial\theta^k}{\partial x^j}(t,x).$$

Consider the following initial value problem for the $n+n^2$ unknown functions $\left(\alpha^i, \beta^i_j\right)$:

$$\begin{aligned} \dot\alpha^i(t) &= V^i\big(\alpha(t)\big), & \alpha^i(t_0) &= a^i, \\ \dot\beta^i_j(t) &= \sum_{k=1}^n \frac{\partial V^i}{\partial y^k}\big(\alpha(t)\big)\beta^k_j(t), & \beta^i_j(t_0) &= b^i_j. \end{aligned}$$

The functions on the right-hand side of this system are C^k functions of $\left(\alpha^i, \beta^i_j\right)$, so the inductive hypothesis implies that its solutions are C^k functions of $\left(t, a^i, b^i_j\right)$. The discussion in the preceding paragraph shows that $\alpha^i(t) = \theta^i(t,x)$ and $\beta^i_j(t) = \partial\theta^i/\partial x^j(t,x)$ solve this system with initial conditions $a^i = x^i$, $b^i_j = \delta^i_j$. This shows that $\partial\theta^i/\partial x^j$ is a C^k function of (t,x), so θ itself is of class C^{k+1}, thus completing the induction. □

Proof of the fundamental theorem. Suppose $U \subseteq \mathbb{R}^n$ is open and $V\colon U \to \mathbb{R}^n$ is smooth. Let $t_0 \in \mathbb{R}$ and $x_0 \in U$ be arbitrary. Because V is smooth, it is locally Lipschitz continuous by Corollary C.30, so the theorems of this appendix apply. Theorem D.3 shows that there exist neighborhoods J_0 of t_0 and U_0 of x_0 such that for each $c \in U_0$, there is a C^1 solution $y\colon J_0 \to U$ to (D.3)–(D.4). Uniqueness of solutions is an immediate consequence of Theorem D.4. Finally, Theorem D.5 shows that the solution is C^k for every k as a function of (t,c), so it is smooth. □

Nonautonomous Systems

Many applications of ODEs require the consideration of nonautonomous systems. In this section we show how the main theorem can be extended to cover the nonautonomous case.

Theorem D.6 (Fundamental Theorem for Nonautonomous ODEs). *Let $J \subseteq \mathbb{R}$ be an open interval and $U \subseteq \mathbb{R}^n$ be an open subset, and let $V\colon J \times U \to \mathbb{R}^n$ be a smooth vector-valued function.*

(a) EXISTENCE: *For any $s_0 \in J$ and $x_0 \in U$, there exist an open interval $J_0 \subseteq J$ containing s_0 and an open subset $U_0 \subseteq U$ containing x_0, such that for each $t_0 \in J_0$ and $c = (c^1, \dots, c^n) \in U_0$, there is a C^1 map $y\colon J_0 \to U$ that solves* (D.1)–(D.2).

(b) UNIQUENESS: *Any two differentiable solutions to* (D.1)–(D.2) *agree on their common domain.*

(c) SMOOTHNESS: *Let J_0 and U_0 be as in* (a), *and define a map $\theta\colon J_0 \times J_0 \times U_0 \to U$ by letting $\theta(t, t_0, c) = y(t)$, where $y\colon J_0 \to U$ is the unique solution to* (D.1)–(D.2). *Then θ is smooth.*

Proof. Consider the following autonomous initial value problem for the $n+1$ functions $y^0, \dots, y^n$:

$$\begin{aligned} \dot{y}^0(t) &= 1; \\ \dot{y}^i(t) &= V^i\big(y^0(t), y^1(t), \dots, y^n(t)\big), \quad & i = 1, \dots, n; \\ y^0(t_0) &= t_0; \\ y^i(t_0) &= c^i, & i = 1, \dots, n. \end{aligned} \tag{D.20}$$

Any solution to (D.20) satisfies $y^0(t) = t$ for all t, and therefore $\big(y^1, \dots, y^n\big)$ solves the nonautonomous system (D.1)–(D.2); and conversely, any solution to (D.1)–(D.2) yields a solution to (D.20) by setting $y^0(t) = t$. Theorem D.1 guarantees that there is an open interval $J_0 \subseteq \mathbb{R}$ containing s_0 and an open subset $W_0 \subseteq J \times U$ containing (s_0, x_0), such that for any $(t_0, c) \in W_0$ there exists a unique solution to (D.20) defined for $t \in J_0$, and the solution depends smoothly on (t, t_0, c). Shrinking J_0 and W_0 if necessary, we may assume that $J_0 \subseteq J$ and $W_0 = J_0 \times U_0$ for some open subset $U_0 \subseteq U$. The result follows. □

Simple Solution Techniques

To get the most out of this book, you need to be able to find explicit solutions to a few differential equations and systems of differential equations. You have probably learned a variety of solution techniques for such equations. The following simple types of equations are more than adequate for the needs of this book.

Separable Equations

A first-order differential equation for a single function $y(t)$ that can be written in the form

$$y'(t) = f\big(y(t)\big)g(t),$$

where f and g are continuous functions with f nonvanishing, is said to be ***separable***. Any separable equation can be solved (at least in principle) by dividing through by $f\big(y(t)\big)$, integrating both sides, and using substitution to transform the left-hand integral:

$$\begin{aligned}\frac{y'(t)}{f(y(t))} &= g(t),\\ \int \frac{y'(t)\,dt}{f(y(t))} &= \int g(t)\,dt,\\ \int \frac{dy}{f(y)} &= \int g(t)\,dt.\end{aligned}$$

If the resulting indefinite integrals can be computed explicitly, the result is a relation involving y and t that can (again, in principle) be solved for y. The constant of integration can then be adjusted to achieve the desired initial condition for y. Separable equations include those of the form $y'(t) = g(t)$ as a trivial special case, which can be solved by direct integration.

2×2 *Constant-Coefficient Linear Systems*

A system of the form

$$\begin{aligned}x'(t) &= ax(t) + by(t),\\ y'(t) &= cx(t) + dy(t),\end{aligned} \tag{D.21}$$

where a, b, c, d are constants, can be written in matrix notation as $Z'(t) = AZ(t)$, where

$$Z(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \qquad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The set of solutions to this type of system always forms a 2-dimensional vector space over $\mathbb{R}$. Once two linearly independent solutions have been found, every other solution is a linear combination of these, and the constants can be adjusted to match any initial conditions.

It is always possible to find at least one (perhaps complex-valued) solution of the form $Z(t) = e^{\lambda t} Z_0$, where λ is an eigenvalue of A and Z_0 is a corresponding eigenvector. If A has two distinct eigenvalues, then there are two such solutions, and they span the solution space. (If the initial conditions are real, then the corresponding solution is real.) On the other hand, if A has only one eigenvalue, there are two cases. If $A - \lambda I_2 \neq 0$, then there is a vector Z_1 (called a ***generalized eigenvector***) such that $(A - \lambda I_2)Z_1 = Z_0$, and a second linearly independent solution is given by $Z(t) = e^{\lambda t}(tZ_0 + Z_1)$. Otherwise, $A = \lambda I_2$, and the two equations in (D.21) are uncoupled and can be solved independently.

Partially Uncoupled Systems

If one of the differential equations in (D.1), say the equation for $\dot{y}^i(t)$, involves none of the dependent variables other than $y^i(t)$, then one can attempt to solve that equation first and substitute the solution into the other equations, thus obtaining a system with fewer unknown functions, which might be solvable by one of the methods above.

► **Exercise D.7.** Solve the following initial value problems.

(a) $x'(t) = x(t)^2; \quad x(0) = x_0.$

(b) $x'(t) = \dfrac{1}{x(t)}; \quad x(0) = x_0 > 0.$

(c) $x'(t) = y; \quad x(0) = x_0;$
$y'(t) = 1; \quad y(0) = y_0.$

(d) $x'(t) = x; \quad x(0) = x_0;$
$y'(t) = 2y; \quad y(0) = y_0.$

(e) $x'(t) = -y; \quad x(0) = x_0;$
$y'(t) = x; \quad y(0) = y_0.$

(f) $x'(t) = -x(t) + y(t); \quad x(0) = x_0;$
$y'(t) = -x(t) - y(t); \quad y(0) = y_0.$

(g) $x'(t) = 1; \quad x(0) = x_0;$
$y'(t) = \dfrac{1}{1 + x(t)^2}; \quad y(0) = y_0.$

References

[AM78] Abraham, Ralph, Marsden, Jerrold E.: Foundations of Mechanics, 2nd edn. Benjamin/Cummings, Reading (1978)

[Apo74] Apostol, Tom M.: Mathematical Analysis, 2nd edn. Addison–Wesley, Reading (1974)

[Bae02] Baez, John C.: The octonions. Bull. Am. Math. Soc. (N.S.) **39**(2), 145–205 (2002)

[BR89] Birkhoff, Garrett, Rota, Gian-Carlo: Ordinary Differential Equations, 4th edn. Wiley, New York (1989)

[Boo86] Boothby, William M.: An Introduction to Differentiable Manifolds and Riemannian Geometry, 2nd edn. Academic Press, Orlando (1986)

[MB58] Bott, Raoul, Milnor, John: On the parallelizability of the spheres. Bull. Am. Math. Soc. (N.S.) **64**, 87–89 (1958)

[BD09] Boyce, William E., DiPrima, Richard C.: Elementary Differential Equations and Boundary Value Problems, 9th edn. Wiley, New York (2009)

[Bre93] Bredon, Glen E.: Topology and Geometry. Springer, New York (1993)

[Cha06] Chavel, Isaac: Riemannian Geometry: A Modern Introduction. Cambridge Studies in Advanced Mathematics, vol. 98, 2nd edn. Cambridge University Press, Cambridge (2006)

[dC92] do Carmo, Manfredo Perdigão: Riemannian Geometry. Birkhäuser, Boston (1992). Translated from the second Portuguese edition by Francis Flaherty

[DK90] Donaldson, S.K., Kronheimer, P.B.: The Geometry of Four-Manifolds. Clarendon Press, New York (1990)

[Eva98] Evans, Lawrence C.: Partial Differential Equations. Am. Math. Soc., Providence (1998)

[Fol95] Folland, Gerald B.: Introduction to Partial Differential Equations, 2nd edn. Princeton University Press, Princeton (1995)

[FQ90] Freedman, Michael, Quinn, Frank: Topology of 4-Manifolds. Princeton University Press, Princeton (1990)

[FIS03] Friedberg, Stephen H., Insel, Arnold J., Spence, Lawrence E.: Linear Algebra, 4th edn. Prentice–Hall, Upper Saddle River (2003)

[Gil95] Gilkey, Peter B.: Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, 2nd edn. CRC Press, Boca Raton (1995)

[HL82] Harvey, Reese, Lawson, H. Blaine Jr.: Calibrated geometries. Acta Math.-Djursholm **148**, 47–157 (1982)

[Hat02] Hatcher, Allen: Algebraic Topology. Cambridge University Press, Cambridge (2002)

[HE73] Hawking, S.W., Ellis, G.F.R.: The Large Scale Structure of Space–Time. Cambridge University Press, London (1973)

[Her75] Herstein, Israel N.: Topics in Algebra, 2nd edn. Wiley, New York (1975)

J.M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics 218,
DOI 10.1007/978-1-4419-9982-5,

[Hör90] Hörmander, Lars: The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis, 2nd edn. Springer, Berlin (1990)
[Hun97] Hungerford, Thomas W.: Abstract Algebra: An Introduction, 2nd edn. Saunders College Publishing, Fort Worth (1997)
[Joh91] John, Fritz: Partial Differential Equations. Applied Mathematical Sciences, vol. 1, 4th edn. Springer, New York (1991)
[Ker58] Kervaire, Michel A.: Non-parallelizability of the n sphere for $n > 7$. Proc. Natl. Acad. Sci. USA **44**, 280–283 (1958)
[Ker60] Kervaire, Michel A.: A manifold which does not admit any differentiable structure. Comment. Math. Helv. **34**, 257–270 (1960)
[KM63] Kervaire, Michel A., Milnor, John W.: Groups of homotopy spheres: I. Ann. Math. **77**, 504–537 (1963)
[KN69] Kobayashi, Shoshichi, Nomizu, Katsumi: Foundations of Differential Geometry, vols. I & II. Wiley, New York (1969)
[LeeJeff09] Lee, Jeffrey M.: Manifolds and Differential Geometry. Am. Math. Soc., Providence (2009)
[LeeRM] Lee, John M.: Riemannian Manifolds: An Introduction to Curvature. Springer, New York (1997)
[LeeTM] Lee, John M.: Introduction to Topological Manifolds, 2nd edn. Springer, New York (2011)
[Mar71] Martinet, J.: Formes de contact sur les variétés de dimension 3. In: Proc. Liverpool Singularities Symp. II, pp. 142–163 (1971) (French)
[Mas89] Massey, William S.: Algebraic Topology: An Introduction. Springer, New York (1989)
[Mil56] Milnor, John W.: On manifolds homeomorphic to the 7-sphere. Ann. Math. **64**, 399–405 (1956)
[Mil63] Milnor, John W.: Morse Theory. Princeton University Press, Princeton (1963)
[Moi77] Moise, Edwin E.: Geometric Topology in Dimensions 2 and 3. Springer, New York (1977)
[Mos65] Moser, Jürgen: On the volume elements on a manifold. Trans. Am. Math. Soc. **120**, 286–294 (1965)
[Mos05] Mostow, G.D.: A structure theorem for homogeneous spaces. Geom. Dedic. **114**, 87–102 (2005)
[Mun60] Munkres, James R.: Obstructions to the smoothing of piecewise differentiable homeomorphisms. Ann. Math. **72**, 521–554 (1960)
[Mun66] Munkres, James R.: Elementary Differential Topology. Princeton University Press, Princeton (1966)
[Mun84] Munkres, James R.: Elements of Algebraic Topology. Addison–Wesley, Menlo Park (1984)
[Mun00] Munkres, James R.: Topology, 2nd edn. Prentice–Hall, Upper Saddle River (2000)
[Nes03] Nestruev, Jet: Smooth Manifolds and Observables. Springer, New York (2003)
[Noe71] Noether, Emmy: Invariant variation problems. Transp. Theory Stat. **1**(3), 186–207 (1971). Translated from the German (Nachr. Akad. Wiss. Gött. Math.-Phys. Kl. II 1918, 235–257)
[Osb82] Osborn, Howard: Vector Bundles, vol. 1. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York (1982)
[Pet06] Petersen, Peter: Riemannian Geometry, 2nd edn. Springer, New York (2006)
[Rud76] Rudin, Walter: Principles of Mathematical Analysis, 3rd edn. McGraw–Hill, New York (1976)
[Sha97] Sharpe, R.W.: Differential Geometry: Cartan's Generalization of Klein's Erlangen Program. Springer, New York (1997)
[Sie92] Sieradski, Allan J.: An Introduction to Topology and Homotopy. PWS-Kent, Boston (1992)
[Sma62] Smale, S.: On the structure of manifolds. Am. J. Math. **84**, 387–399 (1962)

[Spi99] Spivak, Michael: A Comprehensive Introduction to Differential Geometry, vols. 1–5, 3rd edn. Publish or Perish, Berkeley (1999)

[Str00] Strichartz, Robert S.: The Way of Analysis, revised edn. Jones & Bartlett, Boston (2000)

[Var84] Varadarajan, V.S.: Lie Groups, Lie Algebras, and Their Representations. Springer, New York (1984)

[Wal65] Wall, C.T.C.: All 3-manifolds imbed in 5-space. Bull. Am. Math. Soc. (N.S.) **71**, 564–567 (1965)

[War83] Warner, Frank W.: Foundations of Differentiable Manifolds and Lie Groups. Springer, New York (1983)

[Wei71] Weinstein, Alan: Symplectic manifolds and their Lagrangian submanifolds. Adv. Math. **6**, 329–346 (1971)

[Whi36] Whitney, Hassler: Differentiable manifolds. Ann. Math. **37**, 645–680 (1936)

[Whi44a] Whitney, Hassler: The self-intersections of a smooth n-manifold in $2n$-space. Ann. Math. **45**, 220–246 (1944)

[Whi44b] Whitney, Hassler: The singularities of a smooth n-manifold in $(2n - 1)$-space. Ann. Math. **45**, 247–293 (1944)

Notation Index

J.M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics 218,
DOI 10.1007/978-1-4419-9982-5, © Springer Science+Business Media New York 2013

Subject Index

J.M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics 218,
DOI 10.1007/978-1-4419-9982-5,

I

O

P