Loring W. Tu

An Introduction to Manifolds

Second Edition

May 19, 2010

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Loring W. Tu
Department of Mathematics
Tufts University
Medford, MA 02155
loring.tu@tufts.edu

Contents

Appendices

A Brief Introduction

Undergraduate calculus progresses from differentiation and integration of functions on the real line to functions on the plane and in 3-space. Then one encounters vector-valued functions and learns about integrals on curves and surfaces. Real analysis extends differential and integral calculus from $\mathbb{R}^3$ to $\mathbb{R}^n$. This book is about the extension of calculus from curves and surfaces to higher dimensions.

The higher-dimensional analogues of smooth curves and surfaces are called *manifolds*. The constructions and theorems of vector calculus become simpler in the more general setting of manifolds; gradient, curl, and divergence are all special cases of the exterior derivative, and the fundamental theorem for line integrals, Green's theorem, Stokes' theorem, and the divergence theorem are different manifestations of a single general Stokes' theorem for manifolds.

Higher-dimensional manifolds arise even if one is interested only in the three-dimensional space which we inhabit. For example, if we call a rotation followed by a translation an affine motion, then the set of all affine motions in $\mathbb{R}^3$ is a six-dimensional manifold. Moreover, this six-dimensional manifold is not $\mathbb{R}^6$.

We consider two manifolds to be topologically the same if there is a homeomorphism between them, that is, a bijection that is continuous in both directions. A topological invariant of a manifold is a property such as compactness that remains unchanged under a homeomorphism. Another example is the number of connected components of a manifold. Interestingly, we can use differential and integral calculus on manifolds to study the topology of manifolds. We obtain a more refined invariant called the de Rham cohomology of the manifold.

Our plan is as follows. First, we recast calculus on $\mathbb{R}^n$ in a way suitable for generalization to manifolds. We do this by giving meaning to the symbols dx, dy, and dz, so that they assume a life of their own, as *differential forms*, instead of being mere notations as in undergraduate calculus.

While it is not logically necessary to develop differential forms on $\mathbb{R}^n$ before the theory of manifolds—after all, the theory of differential forms on a manifold in Chapter V subsumes that on $\mathbb{R}^n$, from a pedagogical point of view it is advantageous to treat $\mathbb{R}^n$ separately first, since it is on $\mathbb{R}^n$ that the essential simplicity of differential forms and exterior differentiation becomes most apparent.

Another reason for not delving into manifolds right away is so that in a course setting the students without a background in point-set topology can read Appendix A on their own while studying the calculus of differential forms on $\mathbb{R}^n$.

Armed with the rudiments of point-set topology, we define a manifold and derive various conditions for a set to be a manifold. A central idea of calculus is the approximation of a nonlinear object by a linear object. With this in mind, we investigate the relation between a manifold and its tangent spaces. Key examples are Lie groups and their Lie algebras.

Finally we do calculus on manifolds, exploiting the interplay of analysis and topology to show on the one hand how the theorems of vector calculus generalize, and on the other hand, how the results on manifolds define new C^∞ invariants of a manifold, the de Rham cohomology groups.

The de Rham cohomology groups are in fact not merely C^∞ invariants, but also topological invariants, a consequence of the celebrated de Rham theorem that establishes an isomorphism between de Rham cohomology and singular cohomology with real coefficients. To prove this theorem would take us too far afield. Interested readers may find a proof in the sequel [4] to this book.

Chapter I

Euclidean Spaces

The Euclidean space $\mathbb{R}^n$ is the prototype of all manifolds. Not only is it the simplest, but locally every manifold looks like $\mathbb{R}^n$. A good understanding of $\mathbb{R}^n$ is essential in generalizing differential and integral calculus to a manifold.

Euclidean space is special in having a set of standard global coordinates. This is both a blessing and a handicap. It is a blessing because all constructions on $\mathbb{R}^n$ can be defined in terms of the standard coordinates and all computations carried out explicitly. It is a handicap because, defined in terms of coordinates, it is often not obvious which concepts are intrinsic, i.e., independent of coordinates. Since a manifold in general does not have standard coordinates, only coordinate-independent concepts will make sense on a manifold. For example, it turns out that on a manifold of dimension n, it is not possible to integrate functions, because the integral of a function depends on a set of coordinates. The objects that can be integrated are differential forms. It is only because the existence of global coordinates permits an identification of functions with differential n-forms on $\mathbb{R}^n$ that integration of functions becomes possible on $\mathbb{R}^n$.

Our goal in this chapter is to recast calculus on $\mathbb{R}^n$ in a coordinate-free way suitable for generalization to manifolds. To this end, we view a tangent vector not as an arrow or as a column of numbers, but as a derivation on functions. This is followed by an exposition of Hermann Grassmann's formalism of alternating multilinear functions on a vector space, which lays the foundation for the theory of differential forms. Finally we introduce differential forms on $\mathbb{R}^n$, together with two of their basic operations, the wedge product and the exterior derivative, and show how they generalize and simplify vector calculus in $\mathbb{R}^3$.

1

Smooth Functions on a Euclidean Space

The calculus of C^∞ functions will be our primary tool for studying higher-dimensional manifolds. For this reason, we begin with a review of C^∞ functions on $\mathbb{R}^n$.

1.1 C^∞ Versus Analytic Functions

Write the coordinates on $\mathbb{R}^n$ as $x^1, \dots, x^n$ and let $p = (p^1, \dots, p^n)$ be a point in an open set U in $\mathbb{R}^n$. In keeping with the conventions of differential geometry, the indices on coordinates are *superscripts*, not subscripts. An explanation of the rules for superscripts and subscripts is given in Subsection 4.7.

Definition 1.1. Let k be a nonnegative integer. A real-valued function $f\colon U \to \mathbb{R}$ is said to be C^k *at* $p \in U$ if its partial derivatives

$$\frac{\partial^j f}{\partial x^{i_1} \cdots \partial x^{i_j}}$$

of all orders $j \le k$ exist and are continuous at p. The function $f\colon U \to \mathbb{R}$ is C^∞ *at* p if it is C^k for all $k \ge 0$; in other words, its partial derivatives $\partial^j f/\partial x^{i_1} \cdots \partial x^{i_j}$ of all orders exist and are continuous at p. A vector-valued function $f\colon U \to \mathbb{R}^m$ is said to be C^k *at* p if all of its component functions $f^1, \dots, f^m$ are C^k at p. We say that $f\colon U \to \mathbb{R}^m$ is C^k *on* U if it is C^k at every point in U. A similar definition holds for a C^∞ function on an open set U. We treat the terms "C^∞" and "smooth" as synonymous.

Example 1.2.
(i) A C^0 function on U is a continuous function on U.
(ii) Let $f\colon \mathbb{R} \to \mathbb{R}$ be $f(x) = x^{1/3}$. Then

$$f'(x) = \begin{cases} \frac{1}{3}x^{-2/3} & \text{for } x \neq 0, \\ \text{undefined} & \text{for } x = 0. \end{cases}$$

Thus the function f is C^0 but not C^1 at $x = 0$.

(iii) Let $g\colon \mathbb{R} \to \mathbb{R}$ be defined by

$$g(x) = \int_0^x f(t)\,dt = \int_0^x t^{1/3}\,dt = \frac{3}{4}x^{4/3}.$$

Then $g'(x) = f(x) = x^{1/3}$, so $g(x)$ is C^1 but not C^2 at $x = 0$. In the same way one can construct a function that is C^k but not C^{k+1} at a given point.

(iv) The polynomial, sine, cosine, and exponential functions on the real line are all C^∞.

A *neighborhood* of a point in $\mathbb{R}^n$ is an open set containing the point. The function f is *real-analytic* at p if in some neighborhood of p it is equal to its Taylor series at p:

$$f(x) = f(p) + \sum_i \frac{\partial f}{\partial x^i}(p)(x^i - p^i) + \frac{1}{2!}\sum_{i,j}\frac{\partial^2 f}{\partial x^i \partial x^j}(p)(x^i - p^i)(x^j - p^j)$$
$$+ \cdots + \frac{1}{k!}\sum_{i_1,\dots,i_k}\frac{\partial^k f}{\partial x^{i_1}\cdots\partial x^{i_k}}(p)(x^{i_1} - p^{i_1})\cdots(x^{i_k} - p^{i_k}) + \cdots,$$

in which the general term is summed over all $1 \le i_1,\dots,i_k \le n$.

A real-analytic function is necessarily C^∞, because as one learns in real analysis, a convergent power series can be differentiated term by term in its region of convergence. For example, if

$$f(x) = \sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots,$$

then term-by-term differentiation gives

$$f'(x) = \cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots.$$

The following example shows that a C^∞ function need not be real-analytic. The idea is to construct a C^∞ function $f(x)$ on $\mathbb{R}$ whose graph, though not horizontal, is "very flat" near 0 in the sense that all of its derivatives vanish at 0.

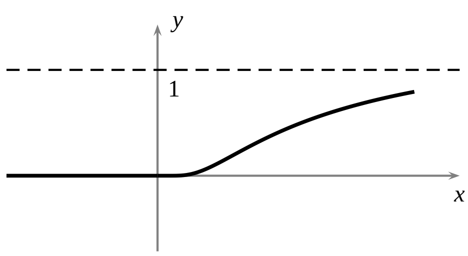


Fig. 1.1. A C^∞ function all of whose derivatives vanish at 0.

Example 1.3 (*A C^∞ function very flat at* 0). Define $f(x)$ on $\mathbb{R}$ by

$$f(x) = \begin{cases} e^{-1/x} & \text{for } x > 0, \\ 0 & \text{for } x \le 0. \end{cases}$$

(See Figure 1.1.) By induction, one can show that f is C^∞ on $\mathbb{R}$ and that the derivatives $f^{(k)}(0) = 0$ for all $k \ge 0$ (Problem 1.2).

The Taylor series of this function at the origin is identically zero in any neighborhood of the origin, since all derivatives $f^{(k)}(0) = 0$. Therefore, $f(x)$ cannot be equal to its Taylor series and $f(x)$ is not real-analytic at 0.

1.2 Taylor's Theorem with Remainder

Although a C^∞ function need not be equal to its Taylor series, there is a Taylor's theorem with remainder for C^∞ functions which is often good enough for our purposes. In the lemma below, we prove the very first case when the Taylor series consists of only the constant term $f(p)$.

We say that a subset S of $\mathbb{R}^n$ is *star-shaped* with respect to a point p in S if for every x in S, the line segment from p to x lies in S (Figure 1.2).

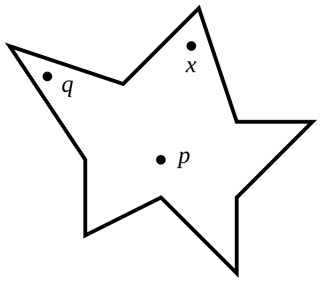


Fig. 1.2. Star-shaped with respect to p, but not with respect to q.

Lemma 1.4 (Taylor's theorem with remainder). *Let f be a C^∞ function on an open subset U of $\mathbb{R}^n$ star-shaped with respect to a point $p = (p^1, \dots, p^n)$ in U. Then there are functions $g_1(x), \dots, g_n(x) \in C^\infty(U)$ such that*

$$f(x) = f(p) + \sum_{i=1}^{n} (x^i - p^i) g_i(x), \quad g_i(p) = \frac{\partial f}{\partial x^i}(p).$$

Proof. Since U is star-shaped with respect to p, for any x in U the line segment $p + t(x - p)$, $0 \le t \le 1$ lies in U (Figure 1.3). So $f(p + t(x - p))$ is defined for $0 \le t \le 1$.

By the chain rule,

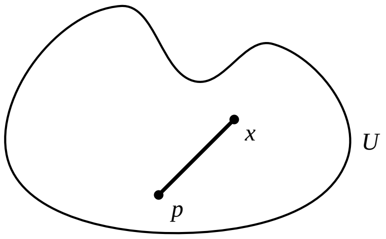


Fig. 1.3. The line segment from p to x.

$$\frac{d}{dt} f(p+t(x-p)) = \sum (x^i - p^i) \frac{\partial f}{\partial x^i}(p+t(x-p)).$$

If we integrate both sides with respect to t from 0 to 1, we get

$$f(p+t(x-p))\Big]_0^1 = \sum (x^i - p^i) \int_0^1 \frac{\partial f}{\partial x^i}(p+t(x-p))\, dt. \tag{1.1}$$

Let

$$g_i(x) = \int_0^1 \frac{\partial f}{\partial x^i}(p+t(x-p))\, dt.$$

Then $g_i(x)$ is C^∞ and (1.1) becomes

$$f(x) - f(p) = \sum (x^i - p^i) g_i(x).$$

Moreover,

$$g_i(p) = \int_0^1 \frac{\partial f}{\partial x^i}(p) dt = \frac{\partial f}{\partial x^i}(p). \qquad \square$$

In case $n = 1$ and $p = 0$, this lemma says that

$$f(x) = f(0) + x g_1(x)$$

for some C^∞ function $g_1(x)$. Applying the lemma repeatedly gives

$$g_i(x) = g_i(0) + x g_{i+1}(x),$$

where g_i, g_{i+1} are C^∞ functions. Hence,

$$\begin{aligned} f(x) &= f(0) + x(g_1(0) + x g_2(x)) \\ &= f(0) + x g_1(0) + x^2 (g_2(0) + x g_3(x)) \\ &\ \ \vdots \\ &= f(0) + g_1(0)x + g_2(0)x^2 + \cdots + g_i(0)x^i + g_{i+1}(x) x^{i+1}. \end{aligned} \tag{1.2}$$

Differentiating (1.2) repeatedly and evaluating at 0, we get

$$g_k(0) = \frac{1}{k!} f^{(k)}(0), \quad k = 1, 2, \ldots, i.$$

So (1.2) is a polynomial expansion of $f(x)$ whose terms up to the last term agree with the Taylor series of $f(x)$ at 0.

Remark. Being star-shaped is not such a restrictive condition, since any open ball

$$B(p,\varepsilon) = \{x \in \mathbb{R}^n \mid \|x-p\| < \varepsilon\}$$

is star-shaped with respect to p. If f is a C^∞ function defined on an open set U containing p, then there is an $\varepsilon > 0$ such that

$$p \in B(p,\varepsilon) \subset U.$$

When its domain is restricted to $B(p,\varepsilon)$, the function f is defined on a star-shaped neighborhood of p and Taylor's theorem with remainder applies.

NOTATION. It is customary to write the standard coordinates on $\mathbb{R}^2$ as x, y, and the standard coordinates on $\mathbb{R}^3$ as x, y, z.

Problems

1.1. A function that is C^2 but not C^3
Let $g\colon \mathbb{R} \to \mathbb{R}$ be the function in Example 1.2(iii). Show that the function $h(x) = \int_0^x g(t)\,dt$ is C^2 but not C^3 at $x = 0$.

1.2.* A C^∞ function very flat at 0
Let $f(x)$ be the function on $\mathbb{R}$ defined in Example 1.3.

(a) Show by induction that for $x > 0$ and $k \geq 0$, the kth derivative $f^{(k)}(x)$ is of the form $p_{2k}(1/x)\,e^{-1/x}$ for some polynomial $p_{2k}(y)$ of degree $2k$ in y.
(b) Prove that f is C^∞ on $\mathbb{R}$ and that $f^{(k)}(0) = 0$ for all $k \geq 0$.

1.3. A diffeomorphism of an open interval with $\mathbb{R}$
Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^n$ be open subsets. A C^∞ map $F\colon U \to V$ is called a *diffeomorphism* if it is bijective and has a C^∞ inverse $F^{-1}\colon V \to U$.

(a) Show that the function $f\colon\,]-\pi/2, \pi/2[\, \to \mathbb{R}$, $f(x) = \tan x$, is a diffeomorphism.
(b) Let a, b be real numbers with $a < b$. Find a linear function $h\colon\,]a,b[\, \to\,]-1,1[$, thus proving that any two finite open intervals are diffeomorphic.

The composite $f \circ h\colon\,]a,b[\, \to \mathbb{R}$ is then a diffeomorphism of an open interval with $\mathbb{R}$.

(c) The exponential function $\exp\colon \mathbb{R} \to\,]0,\infty[$ is a diffeomorphism. Use it to show that for any real numbers a and b, the intervals $\mathbb{R}$, $]a,\infty[$, and $]-\infty,b[$ are diffeomorphic.

1.4. A diffeomorphism of an open cube with $\mathbb{R}^n$
Show that the map

$$f\colon \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[^n \to \mathbb{R}^n, \quad f(x_1,\dots,x_n) = (\tan x_1, \dots, \tan x_n)$$

is a diffeomorphism.

1.5. A diffeomorphism of an open ball with $\mathbb{R}^n$
Let $\mathbf{0} = (0,0)$ be the origin and $B(\mathbf{0},1)$ the open unit disk in $\mathbb{R}^2$. To find a diffeomorphism between $B(\mathbf{0},1)$ and $\mathbb{R}^2$, we identify $\mathbb{R}^2$ with the *xy*-plane in $\mathbb{R}^3$ and introduce the lower open hemisphere

$$S\colon x^2 + y^2 + (z-1)^2 = 1, \quad z < 1$$

in $\mathbb{R}^3$ as an intermediate space (Figure 1.4). First note that the map

$$f : B(\mathbf{0},1) \to S, \qquad (a,b) \mapsto (a,b,1-\sqrt{1-a^2-b^2})$$

is a bijection.

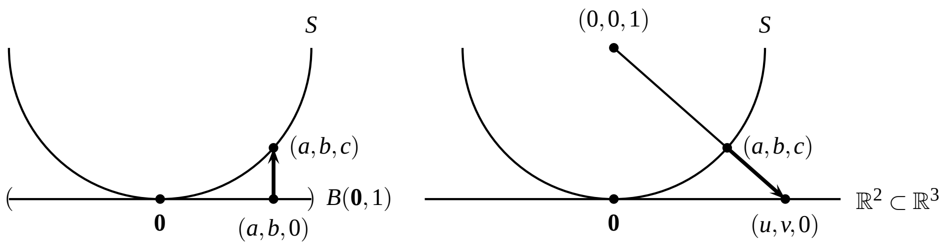


Fig. 1.4. A diffeomorphism of an open disk with $\mathbb{R}^2$.

(a) The *stereographic projection* $g\colon S \to \mathbb{R}^2$ from $(0,0,1)$ is the map that sends a point $(a,b,c) \in S$ to the intersection of the line through $(0,0,1)$ and (a,b,c) with the *xy*-plane. Show that it is given by

$$(a,b,c) \mapsto (u,v) = \left(\frac{a}{1-c}, \frac{b}{1-c}\right), \quad c = 1-\sqrt{1-a^2-b^2}$$

with inverse

$$(u,v) \mapsto \left(\frac{u}{\sqrt{1+u^2+v^2}}, \frac{v}{\sqrt{1+u^2+v^2}}, 1-\frac{1}{\sqrt{1+u^2+v^2}}\right).$$

(b) Composing the two maps f and g gives the map

$$h = g \circ f\colon B(\mathbf{0},1) \to \mathbb{R}^2, \qquad h(a,b) = \left(\frac{a}{\sqrt{1-a^2-b^2}}, \frac{b}{\sqrt{1-a^2-b^2}}\right).$$

Find a formula for $h^{-1}(u,v) = (f^{-1} \circ g^{-1})(u,v)$ and conclude that h is a diffeomorphism of the open disk $B(\mathbf{0},1)$ with $\mathbb{R}^2$.

(c) Generalize part (b) to $\mathbb{R}^n$.

1.6.* Taylor's theorem with remainder to order 2
Prove that if $f\colon \mathbb{R}^2 \to \mathbb{R}$ is C^∞, then there exist C^∞ functions g_{11}, g_{12}, g_{22} on $\mathbb{R}^2$ such that

$$f(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)x + \frac{\partial f}{\partial y}(0,0)y$$
$$+ x^2 g_{11}(x,y) + xy g_{12}(x,y) + y^2 g_{22}(x,y).$$

1.7.* A function with a removable singularity
Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a C^∞ function with $f(0,0) = \partial f/\partial x(0,0) = \partial f/\partial y(0,0) = 0$. Define

$$g(t,u) = \begin{cases} \frac{f(t,tu)}{t} & \text{for } t \neq 0, \\ 0 & \text{for } t = 0. \end{cases}$$

Prove that $g(t,u)$ is C^∞ for $(t,u) \in \mathbb{R}^2$. (*Hint*: Apply Problem 1.6.)

1.8. Bijective C^∞ maps
Define $f\colon \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3$. Show that f is a bijective C^∞ map, but that f^{-1} is not C^∞. (This example shows that a bijective C^∞ map need not have a C^∞ inverse. In complex analysis, the situation is quite different: a bijective holomorphic map $f\colon \mathbb{C} \to \mathbb{C}$ necessarily has a holomorphic inverse.)

2

Tangent Vectors in $\mathbb{R}^n$ as Derivations

In elementary calculus we normally represent a vector at a point p in $\mathbb{R}^3$ algebraically as a column of numbers

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

or geometrically as an arrow emanating from p (Figure 2.1).

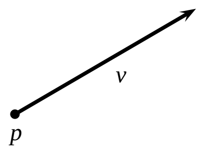


Fig. 2.1. A vector v at p.

Recall that a secant plane to a surface in $\mathbb{R}^3$ is a plane determined by three points of the surface. As the three points approach a point p on the surface, if the corresponding secant planes approach a limiting position, then the plane that is the limiting position of the secant planes is called the tangent plane to the surface at p. Intuitively, the tangent plane to a surface at p is the plane in $\mathbb{R}^3$ that just "touches" the surface at p. A vector at p is tangent to a surface in $\mathbb{R}^3$ if it lies in the tangent plane at p (Figure 2.2).

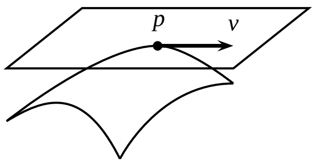


Fig. 2.2. A tangent vector v to a surface at p.

Such a definition of a tangent vector to a surface presupposes that the surface is embedded in a Euclidean space, and so would not apply to the projective plane, for example, which does not sit inside an $\mathbb{R}^n$ in any natural way.

Our goal in this section is to find a characterization of tangent vectors in $\mathbb{R}^n$ that will generalize to manifolds.

2.1 The Directional Derivative

In calculus we visualize the tangent space $T_p(\mathbb{R}^n)$ at p in $\mathbb{R}^n$ as the vector space of all arrows emanating from p. By the correspondence between arrows and column vectors, the vector space $\mathbb{R}^n$ can be identified with this column space. To distinguish between points and vectors, we write a point in $\mathbb{R}^n$ as $p = (p^1, \dots, p^n)$ and a vector in the tangent space $T_p(\mathbb{R}^n)$ as

$$v = \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix} \quad \text{or} \quad \langle v^1, \dots, v^n \rangle.$$

We usually denote the standard basis for $\mathbb{R}^n$ or $T_p(\mathbb{R}^n)$ by $e_1, \dots, e_n$. Then $v = \sum v^i e_i$ for some $v^i \in \mathbb{R}$. Elements of $T_p(\mathbb{R}^n)$ are called *tangent vectors* (or simply *vectors*) at p in $\mathbb{R}^n$. We sometimes drop the parentheses and write $T_p\mathbb{R}^n$ for $T_p(\mathbb{R}^n)$.

The line through a point $p = (p^1, \dots, p^n)$ with direction $v = \langle v^1, \dots, v^n \rangle$ in $\mathbb{R}^n$ has parametrization

$$c(t) = (p^1 + tv^1, \dots, p^n + tv^n).$$

Its ith component $c^i(t)$ is $p^i + tv^i$. If f is C^∞ in a neighborhood of p in $\mathbb{R}^n$ and v is a tangent vector at p, the *directional derivative* of f in the direction v at p is defined to be

$$D_v f = \lim_{t \to 0} \frac{f(c(t)) - f(p)}{t} = \left. \frac{d}{dt} \right|_{t=0} f(c(t)).$$

By the chain rule,

$$D_v f = \sum_{i=1}^{n} \frac{dc^i}{dt}(0) \frac{\partial f}{\partial x^i}(p) = \sum_{i=1}^{n} v^i \frac{\partial f}{\partial x^i}(p). \tag{2.1}$$

In the notation $D_v f$, it is understood that the partial derivatives are to be evaluated at p, since v is a vector at p. So $D_v f$ is a number, not a function. We write

$$D_v = \sum v^i \left. \frac{\partial}{\partial x^i} \right|_p$$

for the map that sends a function f to the number $D_v f$. To simplify the notation we often omit the subscript p if it is clear from the context.

The association $v \mapsto D_v$ of the directional derivative D_v to a tangent vector v offers a way to characterize tangent vectors as certain operators on functions. To make this precise, in the next two subsections we study in greater detail the directional derivative D_v as an operator on functions.

2.2 Germs of Functions

A *relation* on a set S is a subset R of $S \times S$. Given x, y in S, we write $x \sim y$ if and only if $(x, y) \in R$. The relation R is an *equivalence relation* if it satisfies the following three properties for all $x, y, z \in S$:

(i) (reflexivity) $x \sim x$,
(ii) (symmetry) if $x \sim y$, then $y \sim x$,
(iii) (transitivity) if $x \sim y$ and $y \sim z$, then $x \sim z$.

As long as two functions agree on some neighborhood of a point p, they will have the same directional derivatives at p. This suggests that we introduce an equivalence relation on the C^∞ functions defined in some neighborhood of p. Consider the set of all pairs (f, U), where U is a neighborhood of p and $f\colon U \to \mathbb{R}$ is a C^∞ function. We say that (f, U) is *equivalent* to (g, V) if there is an open set $W \subset U \cap V$ containing p such that $f = g$ when restricted to W. This is clearly an equivalence relation because it is reflexive, symmetric, and transitive. The equivalence class of (f, U) is called the *germ* of f at p. We write $C_p^\infty(\mathbb{R}^n)$ or simply C_p^∞ if there is no possibility of confusion, for the set of all germs of C^∞ functions on $\mathbb{R}^n$ at p.

Example. The functions

$$f(x) = \frac{1}{1-x}$$

with domain $\mathbb{R} - \{1\}$ and

$$g(x) = 1 + x + x^2 + x^3 + \cdots$$

with domain the open interval $]-1, 1[$ have the same germ at any point p in the open interval $]-1, 1[$.

An *algebra* over a field K is a vector space A over K with a multiplication map

$$\mu\colon A \times A \to A,$$

usually written $\mu(a, b) = a \cdot b$, such that for all $a, b, c \in A$ and $r \in K$,

(i) (associativity) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$,
(ii) (distributivity) $(a + b) \cdot c = a \cdot c + b \cdot c$ and $a \cdot (b + c) = a \cdot b + a \cdot c$,
(iii) (homogeneity) $r(a \cdot b) = (ra) \cdot b = a \cdot (rb)$.

Equivalently, an algebra over a field K is a ring A (with or without multiplicative identity) which is also a vector space over K such that the ring multiplication satisfies the homogeneity condition (iii). Thus, an algebra has three operations: the addition and multiplication of a ring and the scalar multiplication of a vector space. Usually we omit the multiplication sign and write ab instead of $a \cdot b$.

A map $L\colon V \to W$ between vector spaces over a field K is called a *linear map* or a *linear operator* if for any $r \in K$ and $u, v \in V$,

(i) $L(u + v) = L(u) + L(v)$;

(ii) $L(rv) = rL(v)$.

To emphasize the fact that the scalars are in the field K, such a map is also said to be *K-linear*.

If A and A' are algebras over a field K, then an *algebra homomorphism* is a linear map $L\colon A \to A'$ that preserves the algebra multiplication: $L(ab) = L(a)L(b)$ for all $a, b \in A$.

The addition and multiplication of functions induce corresponding operations on C_p^∞, making it into an algebra over $\mathbb{R}$ (Problem 2.2).

2.3 Derivations at a Point

For each tangent vector v at a point p in $\mathbb{R}^n$, the directional derivative at p gives a map of real vector spaces

$$D_v\colon C_p^\infty \to \mathbb{R}.$$

By (2.1), D_v is $\mathbb{R}$-linear and satisfies the Leibniz rule

$$D_v(fg) = (D_v f)g(p) + f(p)D_v g, \tag{2.2}$$

precisely because the partial derivatives $\partial/\partial x^i|_p$ have these properties.

In general, any linear map $D\colon C_p^\infty \to \mathbb{R}$ satisfying the Leibniz rule (2.2) is called a *derivation at p* or a *point-derivation* of C_p^∞. Denote the set of all derivations at p by $\mathcal{D}_p(\mathbb{R}^n)$. This set is in fact a real vector space, since the sum of two derivations at p and a scalar multiple of a derivation at p are again derivations at p (Problem 2.3).

Thus far, we know that directional derivatives at p are all derivations at p, so there is a map

$$\begin{aligned} \phi\colon T_p(\mathbb{R}^n) &\to \mathcal{D}_p(\mathbb{R}^n), \\ v &\mapsto D_v = \sum v^i \left.\frac{\partial}{\partial x^i}\right|_p. \end{aligned} \tag{2.3}$$

Since D_v is clearly linear in v, the map ϕ is a linear map of vector spaces.

Lemma 2.1. *If D is a point-derivation of C_p^∞, then $D(c) = 0$ for any constant function c.*

Proof. As we do not know if every derivation at p is a directional derivative, we need to prove this lemma using only the defining properties of a derivation at p.

By $\mathbb{R}$-linearity, $D(c) = cD(1)$. So it suffices to prove that $D(1) = 0$. By the Leibniz rule (2.2)

$$D(1) = D(1\cdot 1) = D(1)\cdot 1 + 1\cdot D(1) = 2D(1).$$

Subtracting $D(1)$ from both sides gives $0 = D(1)$. □

The *Kronecker delta* δ is a useful notation that we frequently call upon:

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

Theorem 2.2. *The linear map* $\phi\colon T_p(\mathbb{R}^n) \to \mathcal{D}_p(\mathbb{R}^n)$ *defined in* (2.3) *is an isomorphism of vector spaces.*

Proof. To prove injectivity, suppose $D_v = 0$ for $v \in T_p(\mathbb{R}^n)$. Applying D_v to the coordinate function x^j gives

$$0 = D_v(x^j) = \sum_i v^i \left.\frac{\partial}{\partial x^i}\right|_p x^j = \sum_i v^i \delta^j_i = v^j.$$

Hence, $v = 0$ and ϕ is injective.

To prove surjectivity, let D be a derivation at p and let (f, V) be a representative of a germ in C^∞_p. Making V smaller if necessary, we may assume that V is an open ball, hence star-shaped. By Taylor's theorem with remainder (Lemma 1.4) there are C^∞ functions $g_i(x)$ in a neighborhood of p such that

$$f(x) = f(p) + \sum (x^i - p^i) g_i(x), \quad g_i(p) = \frac{\partial f}{\partial x^i}(p).$$

Applying D to both sides and noting that $D(f(p)) = 0$ and $D(p^i) = 0$ by Lemma 2.1, we get by the Leibniz rule (2.2)

$$\begin{aligned} Df(x) &= \sum (Dx^i) g_i(p) + \sum (p^i - p^i) Dg_i(x) \\ &= \sum (Dx^i) \frac{\partial f}{\partial x^i}(p). \end{aligned}$$

This proves that $D = D_v$ for $v = \langle Dx^1, \ldots, Dx^n \rangle$. □

This theorem shows that one may identify the tangent vectors at p with the derivations at p. Under the vector space isomorphism $T_p(\mathbb{R}^n) \simeq \mathcal{D}_p(\mathbb{R}^n)$, the standard basis $e_1, \ldots, e_n$ for $T_p(\mathbb{R}^n)$ corresponds to the set $\partial/\partial x^1|_p, \ldots, \partial/\partial x^n|_p$ of partial derivatives. From now on, we will make this identification and write a tangent vector $v = \langle v^1, \ldots, v^n \rangle = \sum v^i e_i$ as

$$v = \sum v^i \left.\frac{\partial}{\partial x^i}\right|_p. \tag{2.4}$$

The vector space $\mathcal{D}_p(\mathbb{R}^n)$ of derivations at p, although not as geometric as arrows, turns out to be more suitable for generalization to manifolds.

2.4 Vector Fields

A *vector field* X on an open subset U of $\mathbb{R}^n$ is a function that assigns to each point p in U a tangent vector X_p in $T_p(\mathbb{R}^n)$. Since $T_p(\mathbb{R}^n)$ has basis $\{\partial/\partial x^i|_p\}$, the vector X_p is a linear combination

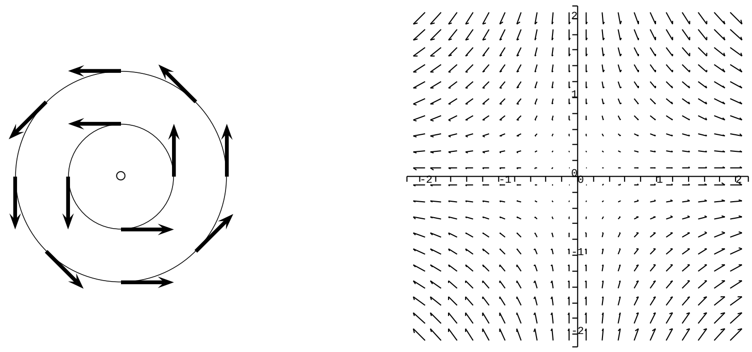
$$X_p = \sum a^i(p) \left.\frac{\partial}{\partial x^i}\right|_p, \quad p \in U, \quad a^i(p) \in \mathbb{R}.$$

Omitting p, we may write $X = \sum a^i \, \partial/\partial x^i$, where the a^i are now functions on U. We say that the vector field X is *C^∞ on U* if the coefficient functions a^i are all C^∞ on U.

Example 2.3. On $\mathbb{R}^2 - \{\mathbf{0}\}$, let $p = (x,y)$. Then

$$X = \frac{-y}{\sqrt{x^2+y^2}} \frac{\partial}{\partial x} + \frac{x}{\sqrt{x^2+y^2}} \frac{\partial}{\partial y} = \left\langle \frac{-y}{\sqrt{x^2+y^2}}, \frac{x}{\sqrt{x^2+y^2}} \right\rangle$$

is the vector field in Figure 2.3(a). As is customary, we draw a vector at p as an arrow emanating from p. The vector field $Y = x\,\partial/\partial x - y\,\partial/\partial y = \langle x, -y \rangle$, suitably rescaled, is sketched in Figure 2.3(b).



(a) The vector field X on $\mathbb{R}^2 - \{\mathbf{0}\}$ (b) The vector field $\langle x, -y \rangle$ on $\mathbb{R}^2$

Fig. 2.3. Vector fields on open subsets of $\mathbb{R}^2$.

One can identify vector fields on U with column vectors of C^∞ functions on U:

$$X = \sum a^i \frac{\partial}{\partial x^i} \quad \longleftrightarrow \quad \begin{bmatrix} a^1 \\ \vdots \\ a^n \end{bmatrix}.$$

This is the same identification as (2.4), but now we are allowing the point p to move in U.

The ring of C^∞ functions on an open set U is commonly denoted $C^\infty(U)$ or $\mathcal{F}(U)$. Multiplication of vector fields by functions on U is defined pointwise:

$$(fX)_p = f(p)X_p, \quad p \in U.$$

Clearly, if $X = \sum a^i \partial/\partial x^i$ is a C^∞ vector field and f is a C^∞ function on U, then $fX = \sum (fa^i)\partial/\partial x^i$ is a C^∞ vector field on U. Thus, the set of all C^∞ vector fields on U, denoted $\mathfrak{X}(U)$, is not only a vector space over $\mathbb{R}$, but also a *module* over the ring $C^\infty(U)$. We recall the definition of a module.

Definition 2.4. If R is a commutative ring with identity, then a (left) *R-module* is an abelian group A with a scalar multiplication map

$$\mu\colon R \times A \to A,$$

usually written $\mu(r,a) = ra$, such that for all $r,s \in R$ and $a,b \in A$,

(i) (associativity) $(rs)a = r(sa)$,
(ii) (identity) if 1 is the multiplicative identity in R, then $1a = a$,
(iii) (distributivity) $(r+s)a = ra + sa$, $r(a+b) = ra + rb$.

If R is a field, then an R-module is precisely a vector space over R. In this sense, a module generalizes a vector space by allowing scalars in a ring rather than a field.

Definition 2.5. Let A and A' be R-modules. An *R-module homomorphism* from A to A' is a map $f\colon A \to A'$ that preserves both addition and scalar multiplication: for all a, $b \in A$ and $r \in R$,

(i) $f(a+b) = f(a) + f(b)$,
(ii) $f(ra) = rf(a)$.

2.5 Vector Fields as Derivations

If X is a C^∞ vector field on an open subset U of $\mathbb{R}^n$ and f is a C^∞ function on U, we define a new function Xf on U by

$$(Xf)(p) = X_p f \quad \text{for any } p \in U.$$

Writing $X = \sum a^i \partial/\partial x^i$, we get

$$(Xf)(p) = \sum a^i(p) \frac{\partial f}{\partial x^i}(p),$$

or

$$Xf = \sum a^i \frac{\partial f}{\partial x^i},$$

which shows that Xf is a C^∞ function on U. Thus, a C^∞ vector field X gives rise to an $\mathbb{R}$-linear map

$$\begin{aligned} C^\infty(U) &\to C^\infty(U), \\ f &\mapsto Xf. \end{aligned}$$

Proposition 2.6 (Leibniz rule for a vector field). *If X is a C^∞ vector field and f and g are C^∞ functions on an open subset U of $\mathbb{R}^n$, then $X(fg)$ satisfies the product rule (Leibniz rule):*

$$X(fg) = (Xf)g + fXg.$$

Proof. At each point $p \in U$, the vector X_p satisfies the Leibniz rule:

$$X_p(fg) = (X_p f)g(p) + f(p)X_p g.$$

As p varies over U, this becomes an equality of functions:

$$X(fg) = (Xf)g + fXg. \qquad \square$$

If A is an algebra over a field K, a *derivation* of A is a K-linear map $D\colon A \to A$ such that

$$D(ab) = (Da)b + aDb \quad \text{for all } a, b \in A.$$

The set of all derivations of A is closed under addition and scalar multiplication and forms a vector space, denoted $\mathrm{Der}(A)$. As noted above, a C^∞ vector field on an open set U gives rise to a derivation of the algebra $C^\infty(U)$. We therefore have a map

$$\begin{aligned} \varphi\colon \mathfrak{X}(U) &\to \mathrm{Der}(C^\infty(U)), \\ X &\mapsto (f \mapsto Xf). \end{aligned}$$

Just as the tangent vectors at a point p can be identified with the point-derivations of C_p^∞, so the vector fields on an open set U can be identified with the derivations of the algebra $C^\infty(U)$; i.e., the map φ is an isomorphism of vector spaces. The injectivity of φ is easy to establish, but the surjectivity of φ takes some work (see Problem 19.12).

Note that a derivation at p is not a derivation of the algebra C_p^∞. A derivation at p is a map from C_p^∞ to $\mathbb{R}$, while a derivation of the algebra C_p^∞ is a map from C_p^∞ to C_p^∞.

Problems

2.1. Vector fields
Let X be the vector field $x\,\partial/\partial x + y\,\partial/\partial y$ and $f(x,y,z)$ the function $x^2+y^2+z^2$ on $\mathbb{R}^3$. Compute Xf.

2.2. Algebra structure on C_p^∞
Define carefully addition, multiplication, and scalar multiplication in C_p^∞. Prove that addition in C_p^∞ is commutative.

2.3. Vector space structure on derivations at a point
Let D and D' be derivations at p in $\mathbb{R}^n$, and $c \in \mathbb{R}$. Prove that

(a) the sum $D + D'$ is a derivation at p.
(b) the scalar multiple cD is a derivation at p.

2.4. Product of derivations
Let A be an algebra over a field K. If D_1 and D_2 are derivations of A, show that $D_1 \circ D_2$ is not necessarily a derivation (it is if D_1 or $D_2 = 0$), but $D_1 \circ D_2 - D_2 \circ D_1$ is always a derivation of A.

3

The Exterior Algebra of Multicovectors

As noted in the introduction, manifolds are higher-dimensional analogues of curves and surfaces. As such, they are usually not linear spaces. Nonetheless, a basic principle in manifold theory is the linearization principle, according to which every manifold can be locally approximated by its tangent space at a point, a linear object. In this way linear algebra enters into manifold theory.

Instead of working with tangent vectors, it turns out to be more fruitful to adopt the dual point of view and work with linear functions on a tangent space. After all, there is only so much that one can do with tangent vectors, which are essentially arrows, but functions, far more flexible, can be added, multiplied, scalar-multiplied, and composed with other maps. Once one admits linear functions on a tangent space, it is but a small step to consider functions of several arguments linear in each argument. These are the multilinear functions on a vector space. The determinant of a matrix, viewed as a function of the column vectors of the matrix, is an example of a multilinear function. Among the multilinear functions, certain ones such as the determinant and the cross product have an *antisymmetric* or *alternating* property: they change sign if two arguments are switched. The alternating multilinear functions with k arguments on a vector space are called *multicovectors of degree k*, or *k-covectors* for short.

It took the genius of Hermann Grassmann (1809–1877), a German mathematician, linguist, and high-school teacher, to recognize the importance of multicovectors. He constructed a vast edifice based on multicovectors, now called the *exterior algebra*, that generalizes parts of vector calculus from $\mathbb{R}^3$ to $\mathbb{R}^n$. For example, the wedge product of two multicovectors on an n-dimensional vector space is a generalization of the cross product in $\mathbb{R}^3$ (see Problem 4.6). Grassmann's work was little appreciated in his lifetime. In fact, he was turned down for a university position and his Ph. D. thesis rejected, because the leading mathematicians of his day such as Möbius and Kummer failed to understand his work. It was only at the turn of twentieth century, in the hands of the great differential geometer Élie Cartan (1869–1951), that Grassmann's exterior algebra found its just recognition as the algebraic basis of the theory of differential forms. This section is an exposition, using modern terminology, of some of Grassmann's ideas.

3.1 Dual Space

If V and W are real vector spaces, we denote by $\mathrm{Hom}(V,W)$ the vector space of all linear maps $f\colon V \to W$. Define the *dual space* $V^\vee$ of V to be the vector space of all real-valued linear functions on V:

$$V^\vee = \mathrm{Hom}(V,\mathbb{R}).$$

The elements of $V^\vee$ are called *covectors* or 1-*covectors* on V.

In the rest of this section, assume V to be a *finite-dimensional* vector space. Let $e_1,\dots,e_n$ be a basis for V. Then every v in V is uniquely a linear combination $v = \sum v^i e_i$ with $v^i \in \mathbb{R}$. Let $\alpha^i\colon V \to \mathbb{R}$ be the linear function that picks out the ith coordinate, $\alpha^i(v) = v^i$. Note that α^i is characterized by

$$\alpha^i(e_j) = \delta^i_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

Proposition 3.1. *The functions $\alpha^1,\dots,\alpha^n$ form a basis for $V^\vee$.*

Proof. We first prove that $\alpha^1,\dots,\alpha^n$ span $V^\vee$. If $f \in V^\vee$ and $v = \sum v^i e_i \in V$, then

$$f(v) = \sum v^i f(e_i) = \sum f(e_i)\alpha^i(v).$$

Hence,

$$f = \sum f(e_i)\alpha^i,$$

which shows that $\alpha^1,\dots,\alpha^n$ span $V^\vee$.

To show linear independence, suppose $\sum c_i\alpha^i = 0$ for some $c_i \in \mathbb{R}$. Applying both sides to the vector e_j gives

$$0 = \sum_i c_i\alpha^i(e_j) = \sum_i c_i\delta^i_j = c_j, \quad j = 1,\dots,n.$$

Hence, $\alpha^1,\dots,\alpha^n$ are linearly independent. □

This basis $\alpha^1,\dots,\alpha^n$ for $V^\vee$ is said to be *dual* to the basis $e_1,\dots,e_n$ for V.

Corollary 3.2. *The dual space $V^\vee$ of a finite-dimensional vector space V has the same dimension as V.*

Example 3.3 (*Coordinate functions*). With respect to a basis $e_1,\dots,e_n$ for a vector space V, every $v \in V$ can be written uniquely as a linear combination $v = \sum b^i(v)e_i$, where $b^i(v) \in \mathbb{R}$. Let $\alpha^1,\dots,\alpha^n$ be the basis of $V^\vee$ dual to $e_1,\dots,e_n$. Then

$$\alpha^i(v) = \alpha^i\left(\sum_j b^j(v)e_j\right) = \sum_j b^j(v)\alpha^i(e_j) = \sum_j b^j(v)\delta^i_j = b^i(v).$$

Thus, the dual basis to $e_1,\dots,e_n$ is precisely the set of coordinate functions $b^1,\dots,b^n$ with respect to the basis $e_1,\dots,e_n$.

3.2 Permutations

Fix a positive integer k. A *permutation* of the set $A = \{1,\dots,k\}$ is a bijection $\sigma\colon A \to A$. More concretely, σ may be thought of as a reordering of the list $1,2,\dots,k$ from its natural increasing order to a new order $\sigma(1),\sigma(2),\dots,\sigma(k)$. The *cyclic permutation*, $(a_1\ a_2\ \cdots\ a_r)$ where the a_i are distinct, is the permutation σ such that $\sigma(a_1) = a_2$, $\sigma(a_2) = a_3$, $\dots$, $\sigma(a_{r-1}) = (a_r)$, $\sigma(a_r) = a_1$, and σ fixes all the other elements of A. A cyclic permutation $(a_1\ a_2\ \cdots\ a_r)$ is also called a *cycle of length r* or an *r-cycle*. A *transposition* is a 2-cycle, that is, a cycle of the form $(a\ b)$ that interchanges a and b, leaving all other elements of A fixed. Two cycles $(a_1 \cdots a_r)$ and $(b_1 \cdots b_s)$ are said to be *disjoint* if the sets $\{a_1,\dots,a_r\}$ and $\{b_1,\dots,b_s\}$ have no elements in common. The *product* $\tau\sigma$ of two permutations τ and σ of A is the composition $\tau\circ\sigma\colon A \to A$, in that order; first apply σ, then τ.

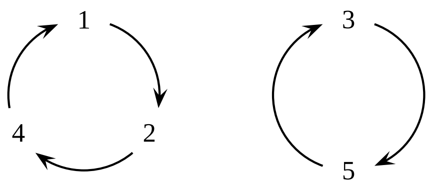
A simple way to describe a permutation $\sigma\colon A \to A$ is by its matrix

$$\begin{bmatrix} 1 & 2 & \cdots & k \\ \sigma(1) & \sigma(2) & \cdots & \sigma(k) \end{bmatrix}.$$

Example 3.4. Suppose the permutation $\sigma\colon \{1,2,3,4,5\} \to \{1,2,3,4,5\}$ maps $1,2,3,4,5$ to $2,4,5,1,3$ in that order. As a matrix,

$$\sigma = \begin{bmatrix} 1\ 2\ 3\ 4\ 5 \\ 2\ 4\ 5\ 1\ 3 \end{bmatrix}. \tag{3.1}$$

To write σ as a product of disjoint cycles, start with any element in $\{1,2,3,4,5\}$, say 1, and apply σ to it repeatedly until we return to the initial element; this gives a cycle: $1 \mapsto 2 \mapsto 4 \to 1$. Next, repeat the procedure to any of the remaining elements, say 3, to get a second cycle: $3 \mapsto 5 \mapsto 3$. Then $\sigma = (1\ 2\ 4)(3\ 5)$.



From this example, it is easy to see that any permutation can be written as a product of disjoint cycles $(a_1\ \cdots\ a_r)(b_1\ \cdots\ b_s)\cdots$.

Let S_k be the group of all permutations of the set $\{1,\dots,k\}$. A permutation is *even* or *odd* depending on whether it is the product of an even or an odd number of transpositions. From the theory of permutations we know that this is a well-defined concept: an even permutation can never be written as the product of an odd number of transpositions and vice versa. The *sign* of a permutation σ, denoted $\operatorname{sgn}(\sigma)$ or $\operatorname{sgn}\sigma$, is defined to be $+1$ or -1 depending on whether the permutation is even or odd. Clearly, the sign of a permutation satisfies

$$\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \tag{3.2}$$

for $\sigma,\tau \in S_k$.

Example 3.5. The decomposition

$$(1\ 2\ 3\ 4\ 5) = (1\ 5)(1\ 4)(1\ 3)(1\ 2)$$

shows that the 5-cycle $(1\ 2\ 3\ 4\ 5)$ is an even permutation.

More generally, the decomposition

$$(a_1\ a_2\ \cdots\ a_r) = (a_1\ a_r)(a_1\ a_{r-1})\cdots(a_1\ a_3)(a_1\ a_2)$$

shows that an r-cycle is an even permutation if and only if r is odd, and an odd permutation if and only if r is even. Thus one way to compute the sign of a permutation is to decompose it into a product of cycles and to count the number of cycles of even length. For example, the permutation $\sigma = (1\ 2\ 4)(3\ 5)$ in Example 3.4 is odd because $(1\ 2\ 4)$ is even and $(3\ 5)$ is odd.

An *inversion* in a permutation σ is an ordered pair $(\sigma(i), \sigma(j))$ such that $i < j$ but $\sigma(i) > \sigma(j)$. To find all the inversions in a permutation σ, it suffices to scan the second row of the matrix of σ from left to right; the inversions are the pairs (a,b) with $a > b$ and a to the left of b. For the permutation σ in Example 3.4, from its matrix (3.1) we can read off its five inversions: $(2,1)$, $(4,1)$, $(5,1)$, $(4,3)$, and $(5,3)$.

Exercise 3.6 (Inversions).* Find the inversions in the permutation $\tau = (1\ 2\ 3\ 4\ 5)$ of Example 3.5.

A second way to compute the sign of a permutation is to count the number of inversions, as we illustrate in the following example.

Example 3.7. Let σ be the permutation of Example 3.4. Our goal is to turn σ into the identity permutation $\mathbb{1}$ by multiplying it on the left by transpositions.

(i) To move 1 to its natural position at the beginning of the second row of the matrix of σ, we need to move it across the three elements $2, 4, 5$. This can be accomplished by multiplying σ on the left by three transpositions: first $(5\ 1)$, then $(4\ 1)$, and finally $(2\ 1)$:

$$\sigma = \begin{bmatrix} 1\ 2\ 3\ 4\ 5 \\ 2\ 4\ 5\ 1\ 3 \end{bmatrix} \xrightarrow{(5\ 1)} \begin{bmatrix} \\ 2\ 4\ 1\ 5\ 3 \end{bmatrix} \xrightarrow{(4\ 1)} \begin{bmatrix} \\ 2\ 1\ 4\ 5\ 3 \end{bmatrix} \xrightarrow{(2\ 1)} \begin{bmatrix} \\ 1\ 2\ 4\ 5\ 3 \end{bmatrix}.$$

The three transpositions $(5\ 1)$, $(4\ 1)$, and $(2\ 1)$ correspond precisely to the three inversions of σ ending in 1.

(ii) The element 2 is already in its natural position in the second row of the matrix.

(iii) To move 3 to its natural position in the second row, we need to move it across two elements 4, 5. This can be accomplished by

$$\begin{bmatrix} 1\ 2\ 3\ 4\ 5 \\ 1\ 2\ 4\ 5\ 3 \end{bmatrix} \xrightarrow{(5\ 3)} \begin{bmatrix} \\ 1\ 2\ 4\ 3\ 5 \end{bmatrix} \xrightarrow{(4\ 3)} \begin{bmatrix} \\ 1\ 2\ 3\ 4\ 5 \end{bmatrix} = \mathbb{1}.$$

Thus,

$$(4\ 3)(5\ 3)(2\ 1)(4\ 1)(5\ 1)\sigma = \mathbb{1}. \tag{3.3}$$

Note that the two transpositions $(5\ 3)$ and $(4\ 3)$ correspond to the two inversions ending in 3. Multiplying both sides of (3.3) on the left by the transpositions $(4\ 3)$, then $(5\ 3)$, then $(2\ 1)$, and so on eventually yields

$$\sigma = (5\ 1)(4\ 1)(2\ 1)(5\ 3)(4\ 3).$$

This shows that σ can be written as a product of as many transpositions as the number of inversions in it.

With this example in mind, we prove the following proposition.

Proposition 3.8. *A permutation is even if and only if it has an even number of inversions.*

Proof. We will obtain the identity permutation $\mathbb{1}$ by multiplying σ on the left by a number of transpositions. This can be achieved in k steps.

(i) First, look for the number 1 among $\sigma(1), \sigma(2), \dots, \sigma(k)$. Every number preceding 1 in this list gives rise to an inversion, for if $1 = \sigma(i)$, then $(\sigma(1), 1), \dots, (\sigma(i-1), 1)$ are inversions of σ. Now move 1 to the beginning of the list across the $i-1$ elements $\sigma(1), \dots, \sigma(i-1)$. This requires multiplying σ on the left by $i-1$ transpositions:

$$\sigma_1 = (\sigma(1)\ 1) \cdots (\sigma(i-1)\ 1)\sigma = \begin{bmatrix} \\ 1\ \sigma(1)\ \cdots\ \sigma(i-1)\ \sigma(i+1)\ \cdots\ \sigma(k) \end{bmatrix}.$$

Note that the number of transpositions is the number of inversions ending in 1.

(ii) Next look for the number 2 in the list: $1, \sigma(1), \dots, \sigma(i-1), \sigma(i+1), \dots, \sigma(k)$. Every number other than 1 preceding 2 in this list gives rise to an inversion $(\sigma(m), 2)$. Suppose there are i_2 such numbers. Then there are i_2 inversions ending in 2. In moving 2 to its natural position $1, 2, \sigma(1), \sigma(2), \dots,$ we need to move it across i_2 numbers. This can be accomplished by multiplying σ_1 on the left by i_2 transpositions.

Repeating this procedure, we see that for each $j = 1, \dots, k$, the number of transpositions required to move j to its natural position is the same as the number of inversions ending in j. In the end we achieve the identity permutation, i.e, the ordered list $1, 2, \dots, k$, from $\sigma(1), \sigma(2), \dots, \sigma(k)$ by multiplying σ by as many transpositions as the total number of inversions in σ. Therefore, $\operatorname{sgn}(\sigma) = (-1)^{\#\text{ inversions in }\sigma}$. □

3.3 Multilinear Functions

Denote by $V^k = V \times \cdots \times V$ the Cartesian product of k copies of a real vector space V. A function $f \colon V^k \to \mathbb{R}$ is *k-linear* if it is linear in each of its k arguments:

$$f(\dots, av + bw, \dots) = af(\dots, v, \dots) + bf(\dots, w, \dots)$$

for all $a, b \in \mathbb{R}$ and $v, w \in V$. Instead of 2-linear and 3-linear, it is customary to say "bilinear" and "trilinear." A k-linear function on V is also called a *k-tensor* on V. We will denote the vector space of all k-tensors on V by $L_k(V)$. If f is a k-tensor on V, we also call k the *degree* of f.

Example 3.9 (*Dot product on* $\mathbb{R}^n$). With respect to the standard basis $e_1, \ldots, e_n$ for $\mathbb{R}^n$, the *dot product,* defined by

$$f(v,w) = v \bullet w = \sum_i v^i w^i, \quad \text{where } v = \sum v^i e_i,\ w = \sum w^i e_i,$$

is bilinear.

Example. The determinant $f(v_1, \ldots, v_n) = \det[v_1 \ \cdots \ v_n]$, viewed as a function of the n column vectors $v_1, \ldots, v_n$ in $\mathbb{R}^n$, is n-linear.

Definition 3.10. A k-linear function $f \colon V^k \to \mathbb{R}$ is *symmetric* if

$$f\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) = f(v_1, \ldots, v_k)$$

for all permutations $\sigma \in S_k$; it is *alternating* if

$$f\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) = (\operatorname{sgn} \sigma) f(v_1, \ldots, v_k)$$

for all $\sigma \in S_k$.

Examples.

(i) The dot product $f(v,w) = v \bullet w$ on $\mathbb{R}^n$ is symmetric.
(ii) The determinant $f(v_1, \ldots, v_n) = \det[v_1 \ \cdots \ v_n]$ on $\mathbb{R}^n$ is alternating.
(iii) The cross product $v \times w$ on $\mathbb{R}^3$ is alternating.
(iv) For any two linear functions $f, g \colon V \to \mathbb{R}$ on a vector space V, the function $f \wedge g \colon V \times V \to \mathbb{R}$ defined by

$$(f \wedge g)(u,v) = f(u)g(v) - f(v)g(u)$$

is alternating. This is a special case of the *wedge product*, which we will soon define.

We are especially interested in the space $A_k(V)$ of all alternating k-linear functions on a vector space V for $k > 0$. These are also called *alternating k-tensors*, *k-covectors*, or *multicovectors of degree k* on V. For $k = 0$, we define a 0-*covector* to be a constant so that $A_0(V)$ is the vector space $\mathbb{R}$. A 1-covector is simply a covector.

3.4 The Permutation Action on Multilinear Functions

If f is a k-linear function on a vector space V and σ is a permutation in S_k, we define a new k-linear function σf by

$$(\sigma f)(v_1,\dots,v_k) = f\left(v_{\sigma(1)},\dots,v_{\sigma(k)}\right).$$

Thus, f is symmetric if and only if $\sigma f = f$ for all $\sigma \in S_k$ and f is alternating if and only if $\sigma f = (\operatorname{sgn}\sigma) f$ for all $\sigma \in S_k$.

When there is only one argument, the permutation group S_1 is the identity group and a 1-linear function is both symmetric and alternating. In particular,

$$A_1(V) = L_1(V) = V^\vee.$$

Lemma 3.11. *If $\sigma, \tau \in S_k$ and f is a k-linear function on V, then $\tau(\sigma f) = (\tau\sigma) f$.*

Proof. For $v_1,\dots,v_k \in V$,

$$\begin{aligned}
\tau(\sigma f)(v_1,\dots,v_k) &= (\sigma f)\left(v_{\tau(1)},\dots,v_{\tau(k)}\right) \\
&= (\sigma f)(w_1,\dots,w_k) \qquad\qquad (\text{letting } w_i = v_{\tau(i)}) \\
&= f\left(w_{\sigma(1)},\dots,w_{\sigma(k)}\right) \\
&= f\left(v_{\tau(\sigma(1))},\dots,v_{\tau(\sigma(k))}\right) = f\left(v_{(\tau\sigma)(1)},\dots,v_{(\tau\sigma)(k)}\right) \\
&= (\tau\sigma) f(v_1,\dots,v_k).
\end{aligned}$$

□

In general, if G is a group and X is a set, a map

$$\begin{aligned} G \times X &\to X, \\ (\sigma, x) &\mapsto \sigma \cdot x \end{aligned}$$

is called a *left action* of G on X if

(i) $e \cdot x = x$ where e is the identity element in G and x is any element in X, and
(ii) $\tau \cdot (\sigma \cdot x) = (\tau\sigma) \cdot x$ for all $\tau, \sigma \in G, x \in X$.

The *orbit* of an element $x \in X$ is defined to be the set $Gx := \{\sigma \cdot x \in X \mid \sigma \in G\}$. In this terminology, we have defined a left action of the permutation group S_k on the space $L_k(V)$ of k-linear functions on V. Note that each permutation acts as a linear function on the vector space $L_k(V)$ since σf is $\mathbb{R}$-linear in f.

A *right action* of G on X is defined similarly; it is a map $X \times G \to X$ such that

(i) $x \cdot e = x$, and
(ii) $(x \cdot \sigma) \cdot \tau = x \cdot (\sigma\tau)$

for all $\sigma, \tau \in G$ and $x \in X$.

Remark. In some books the notation for σf is f^σ. In that notation, $(f^\sigma)^\tau = f^{\tau\sigma}$, not $f^{\sigma\tau}$.

3.5 The Symmetrizing and Alternating Operators

Given any k-linear function f on a vector space V, there is a way to make a symmetric k-linear function Sf from it:

$$(Sf)(v_1,\ldots,v_k) = \sum_{\sigma\in S_k} f\left(v_{\sigma(1)},\ldots,v_{\sigma(k)}\right)$$

or, in our new shorthand,

$$Sf = \sum_{\sigma\in S_k} \sigma f.$$

Similarly, there is a way to make an alternating k-linear function from f. Define

$$Af = \sum_{\sigma\in S_k} (\operatorname{sgn}\sigma)\sigma f.$$

Proposition 3.12. *If f is a k-linear function on a vector space V, then*

(i) *the k-linear function Sf is symmetric, and*
(ii) *the k-linear function Af is alternating.*

Proof. We prove (ii) only, leaving (i) as an exercise. For $\tau \in S_k$,

$$\begin{aligned}
\tau(Af) &= \sum_{\sigma\in S_k} (\operatorname{sgn}\sigma)\tau(\sigma f) \\
&= \sum_{\sigma\in S_k} (\operatorname{sgn}\sigma)(\tau\sigma) f \qquad \text{(by Lemma 3.11)} \\
&= (\operatorname{sgn}\tau)\sum_{\sigma\in S_k} (\operatorname{sgn}\tau\sigma)(\tau\sigma) f \qquad \text{(by (3.2))} \\
&= (\operatorname{sgn}\tau)Af,
\end{aligned}$$

since as σ runs through all permutations in S_k, so does $\tau\sigma$. □

Exercise 3.13 (Symmetrizing operator).* Show that the k-linear function Sf is symmetric.

Lemma 3.14. *If f is an alternating k-linear function on a vector space V, then $Af = (k!)f$.*

Proof. Since for alternating f we have $\sigma f = (\operatorname{sgn}\sigma)f$, and $\operatorname{sgn}\sigma = \pm 1$,

$$Af = \sum_{\sigma\in S_k} (\operatorname{sgn}\sigma)\sigma f = \sum_{\sigma\in S_k} (\operatorname{sgn}\sigma)(\operatorname{sgn}\sigma) f = (k!)f.$$ □

Exercise 3.15 (Alternating operator).* If f is a 3-linear function on a vector space V and $v_1, v_2, v_3 \in V$, what is $(Af)(v_1,v_2,v_3)$?

3.6 The Tensor Product

Let f be a k-linear function and g an ℓ-linear function on a vector space V. Their *tensor product* is the $(k+\ell)$-linear function $f \otimes g$ defined by

$$(f \otimes g)(v_1, \ldots, v_{k+\ell}) = f(v_1, \ldots, v_k) g(v_{k+1}, \ldots, v_{k+\ell}).$$

Example 3.16 (*Bilinear maps*). Let $e_1, \ldots, e_n$ be a basis for a vector space V, $\alpha^1, \ldots, \alpha^n$ the dual basis in $V^\vee$, and $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$ a bilinear map on V. Set $g_{ij} = \langle e_i, e_j \rangle \in \mathbb{R}$. If $v = \sum v^i e_i$ and $w = \sum w^i e_i$, then as we observed in Example 3.3, $v^i = \alpha^i(v)$ and $w^j = \alpha^j(w)$. By bilinearity, we can express $\langle\ ,\ \rangle$ in terms of the tensor product:

$$\begin{aligned} \langle v, w \rangle = \sum v^i w^j \langle e_i, e_j \rangle &= \sum \alpha^i(v) \alpha^j(w) g_{ij} \\ &= \sum g_{ij} (\alpha^i \otimes \alpha^j)(v, w). \end{aligned}$$

Hence, $\langle\ ,\ \rangle = \sum g_{ij} \alpha^i \otimes \alpha^j$. This notation is often used in differential geometry to describe an inner product on a vector space.

Exercise 3.17 (Associativity of the tensor product). Check that the tensor product of multilinear functions is associative: if f, g, and h are multilinear functions on V, then

$$(f \otimes g) \otimes h = f \otimes (g \otimes h).$$

3.7 The Wedge Product

If two multilinear functions f and g on a vector space V are alternating, then we would like to have a product that is alternating as well. This motivates the definition of the *wedge product*, also called the *exterior product*: for $f \in A_k(V)$ and $g \in A_\ell(V)$,

$$f \wedge g = \frac{1}{k!\ell!} A(f \otimes g); \tag{3.4}$$

or explicitly,

$$\begin{aligned} &(f \wedge g)(v_1, \ldots, v_{k+\ell}) \\ &\quad = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} (\operatorname{sgn} \sigma) f\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) g\left(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\right). \end{aligned} \tag{3.5}$$

By Proposition 3.12, $f \wedge g$ is alternating.

When $k = 0$, the element $f \in A_0(V)$ is simply a constant c. In this case, the wedge product $c \wedge g$ is scalar multiplication, since the right-hand side of (3.5) is

$$\frac{1}{\ell!} \sum_{\sigma \in S_\ell} (\operatorname{sgn} \sigma) c g\left(v_{\sigma(1)}, \ldots, v_{\sigma(\ell)}\right) = c g\left(v_1, \ldots, v_\ell\right).$$

Thus $c \wedge g = cg$ for $c \in \mathbb{R}$ and $g \in A_\ell(V)$.

The coefficient $1/k!\ell!$ in the definition of the wedge product compensates for repetitions in the sum: for every permutation $\sigma \in S_{k+\ell}$, there are $k!$ permutations τ in S_k that permute the first k arguments $v_{\sigma(1)},\dots,v_{\sigma(k)}$ and leave the arguments of g alone; for all τ in S_k, the resulting permutations $\sigma\tau$ in $S_{k+\ell}$ contribute the same term to the sum since

$$\begin{aligned}(\operatorname{sgn}\sigma\tau) f\left(v_{\sigma\tau(1)},\dots,v_{\sigma\tau(k)}\right) &= (\operatorname{sgn}\sigma\tau)(\operatorname{sgn}\tau) f\left(v_{\sigma(1)},\dots,v_{\sigma(k)}\right)\\ &= (\operatorname{sgn}\sigma) f\left(v_{\sigma(1)},\dots,v_{\sigma(k)}\right),\end{aligned}$$

where the first equality follows from the fact that $(\tau(1),\dots,\tau(k))$ is a permutation of $(1,\dots,k)$. So we divide by $k!$ to get rid of the $k!$ repeating terms in the sum coming from permutations of the k arguments of f; similarly, we divide by $\ell!$ on account of the ℓ arguments of g.

Example 3.18. For $f \in A_2(V)$ and $g \in A_1(V)$,

$$\begin{aligned}A(f\otimes g)(v_1,v_2,v_3) = {}& f(v_1,v_2)g(v_3) - f(v_1,v_3)g(v_2) + f(v_2,v_3)g(v_1)\\ &- f(v_2,v_1)g(v_3) + f(v_3,v_1)g(v_2) - f(v_3,v_2)g(v_1).\end{aligned}$$

Among these six terms, there are three pairs of equal terms, which we have lined up vertically in the display above:

$$f(v_1,v_2)g(v_3) = -f(v_2,v_1)g(v_3), \quad \text{and so on.}$$

Therefore, after dividing by 2,

$$(f\wedge g)(v_1,v_2,v_3) = f(v_1,v_2)g(v_3) - f(v_1,v_3)g(v_2) + f(v_2,v_3)g(v_1).$$

One way to avoid redundancies in the definition of $f\wedge g$ is to stipulate that in the sum (3.5), $\sigma(1),\dots,\sigma(k)$ be in ascending order and $\sigma(k+1),\dots,\sigma(k+\ell)$ also be in ascending order. We call a permutation $\sigma \in S_{k+\ell}$ a (k,ℓ)-*shuffle* if

$$\sigma(1) < \cdots < \sigma(k) \quad \text{and} \quad \sigma(k+1) < \cdots < \sigma(k+\ell).$$

By the paragraph before Example 3.18, one may rewrite (3.5) as

$$\begin{aligned}&(f\wedge g)(v_1,\dots,v_{k+\ell})\\ &\quad = \sum_{\substack{(k,\ell)\text{-shuffles}\\ \sigma}} (\operatorname{sgn}\sigma) f\left(v_{\sigma(1)},\dots,v_{\sigma(k)}\right) g\left(v_{\sigma(k+1)},\dots,v_{\sigma(k+\ell)}\right).\end{aligned} \tag{3.6}$$

Written this way, the definition of $(f\wedge g)(v_1,\dots,v_{k+\ell})$ is a sum of $\binom{k+\ell}{k}$ terms, instead of $(k+\ell)!$ terms.

Example 3.19 (*Wedge product of two covectors*).* If f and g are covectors on a vector space V and $v_1, v_2 \in V$, then by (3.6)

$$(f\wedge g)(v_1,v_2) = f(v_1)g(v_2) - f(v_2)g(v_1).$$

Exercise 3.20 (Wedge product of two 2-covectors). For $f,g \in A_2(V)$, write out the definition of $f\wedge g$ using $(2,2)$-shuffles.

3.8 Anticommutativity of the Wedge Product

It follows directly from the definition of the wedge product (3.5) that $f \wedge g$ is bilinear in f and in g.

Proposition 3.21. *The wedge product is* anticommutative*: if $f \in A_k(V)$ and $g \in A_\ell(V)$, then*

$$f \wedge g = (-1)^{k\ell} g \wedge f.$$

Proof. Define $\tau \in S_{k+\ell}$ to be the permutation

$$\tau = \begin{bmatrix} 1 & \cdots & \ell & \ell+1 & \cdots & \ell+k \\ k+1 & \cdots & k+\ell & 1 & \cdots & k \end{bmatrix}.$$

This means that

$$\tau(1) = k+1, \ldots, \tau(\ell) = k+\ell, \tau(\ell+1) = 1, \ldots, \tau(\ell+k) = k.$$

Then

$$\begin{aligned} \sigma(1) &= \sigma\tau(\ell+1), \ldots, \sigma(k) = \sigma\tau(\ell+k), \\ \sigma(k+1) &= \sigma\tau(1), \ldots, \sigma(k+\ell) = \sigma\tau(\ell). \end{aligned}$$

For any $v_1, \ldots, v_{k+\ell} \in V$,

$$\begin{aligned} &A(f \otimes g)(v_1, \ldots, v_{k+\ell}) \\ &\quad = \sum_{\sigma \in S_{k+\ell}} (\operatorname{sgn} \sigma) f\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) g\left(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\right) \\ &\quad = \sum_{\sigma \in S_{k+\ell}} (\operatorname{sgn} \sigma) f\left(v_{\sigma\tau(\ell+1)}, \ldots, v_{\sigma\tau(\ell+k)}\right) g\left(v_{\sigma\tau(1)}, \ldots, v_{\sigma\tau(\ell)}\right) \\ &\quad = (\operatorname{sgn} \tau) \sum_{\sigma \in S_{k+\ell}} (\operatorname{sgn} \sigma\tau) g\left(v_{\sigma\tau(1)}, \ldots, v_{\sigma\tau(\ell)}\right) f\left(v_{\sigma\tau(\ell+1)}, \ldots, v_{\sigma\tau(\ell+k)}\right) \\ &\quad = (\operatorname{sgn} \tau) A(g \otimes f)(v_1, \ldots, v_{k+\ell}). \end{aligned}$$

The last equality follows from the fact that as σ runs through all permutations in $S_{k+\ell}$, so does $\sigma\tau$.

We have proven

$$A(f \otimes g) = (\operatorname{sgn} \tau) A(g \otimes f).$$

Dividing by $k!\ell!$ gives

$$f \wedge g = (\operatorname{sgn} \tau) g \wedge f.$$

Exercise 3.22 (Sign of a permutation).* Show that $\operatorname{sgn} \tau = (-1)^{k\ell}$. □

Corollary 3.23. *If f is a multicovector of odd degree on V, then $f \wedge f = 0$.*

Proof. Let k be the degree of f. By anticommutativity,

$$\begin{aligned} f \wedge f &= (-1)^{k^2} f \wedge f \\ &= -f \wedge f, \end{aligned}$$

since k is odd. Hence, $2f \wedge f = 0$. Dividing by 2 gives $f \wedge f = 0$. □

3.9 Associativity of the Wedge Product

The wedge product of a k-covector f and an ℓ-covector g on a vector space V is by definition the $(k+\ell)$-covector

$$f \wedge g = \frac{1}{k!\ell!} A(f \otimes g).$$

To prove the associativity of the wedge product, we will follow Godbillon [14] by first proving a lemma on the alternating operator A.

Lemma 3.24. *Suppose f is a k-linear function and g an ℓ-linear function on a vector space V. Then*

(i) $A(A(f)\otimes g) = k!A(f\otimes g)$, *and*
(ii) $A(f \otimes A(g)) = \ell! A(f \otimes g)$.

Proof. (i) By definition,

$$A(A(f)\otimes g) = \sum_{\sigma\in S_{k+\ell}} (\operatorname{sgn}\sigma)\sigma\left(\sum_{\tau\in S_k}(\operatorname{sgn}\tau)(\tau f)\otimes g\right).$$

We can view $\tau \in S_k$ also as a permutation in $S_{k+\ell}$ fixing $k+1,\dots,k+\ell$. Viewed this way, τ satisfies

$$(\tau f)\otimes g = \tau(f\otimes g).$$

Hence,

$$A(A(f)\otimes g) = \sum_{\sigma\in S_{k+\ell}}\sum_{\tau\in S_k}(\operatorname{sgn}\sigma)(\operatorname{sgn}\tau)(\sigma\tau)(f\otimes g). \tag{3.7}$$

For each $\mu \in S_{k+\ell}$ and each $\tau \in S_k$, there is a unique element $\sigma = \mu\tau^{-1} \in S_{k+\ell}$ such that $\mu = \sigma\tau$, so each $\mu \in S_{k+\ell}$ appears once in the double sum (3.7) for each $\tau \in S_k$, and hence $k!$ times in total. So the double sum (3.7) can be rewritten as

$$\begin{aligned} A(A(f)\otimes g) &= k! \sum_{\mu\in S_{k+\ell}} (\operatorname{sgn}\mu)\mu(f\otimes g) \\ &= k! A(f\otimes g). \end{aligned}$$

The equality in (ii) is proved in the same way. □

Proposition 3.25 (Associativity of the wedge product). *Let V be a real vector space and f,g,h alternating multilinear functions on V of degrees k,ℓ,m, respectively. Then*

$$(f\wedge g)\wedge h = f\wedge(g\wedge h).$$

Proof. By the definition of the wedge product,

$$\begin{aligned}(f\wedge g)\wedge h &= \frac{1}{(k+\ell)!m!}A((f\wedge g)\otimes h)\\ &= \frac{1}{(k+\ell)!m!}\frac{1}{k!\ell!}A(A(f\otimes g)\otimes h)\\ &= \frac{(k+\ell)!}{(k+\ell)!m!k!\ell!}A((f\otimes g)\otimes h) \quad \text{(by Lemma 3.24(i))}\\ &= \frac{1}{k!\ell!m!}A((f\otimes g)\otimes h).\end{aligned}$$

Similarly,

$$\begin{aligned}f\wedge(g\wedge h) &= \frac{1}{k!(\ell+m)!}A\left(f\otimes \frac{1}{\ell!m!}A(g\otimes h)\right)\\ &= \frac{1}{k!\ell!m!}A(f\otimes(g\otimes h)).\end{aligned}$$

Since the tensor product is associative, we conclude that

$$(f\wedge g)\wedge h = f\wedge(g\wedge h).$$ □

By associativity, we can omit the parentheses in a multiple wedge product such as $(f\wedge g)\wedge h$ and write simply $f\wedge g\wedge h$.

Corollary 3.26. *Under the hypotheses of the proposition,*

$$f\wedge g\wedge h = \frac{1}{k!\ell!m!}A(f\otimes g\otimes h).$$

This corollary easily generalizes to an arbitrary number of factors: if $f_i \in A_{d_i}(V)$, then

$$f_1\wedge\cdots\wedge f_r = \frac{1}{(d_1)!\cdots(d_r)!}A(f_1\otimes\cdots\otimes f_r). \tag{3.8}$$

In particular, we have the following proposition. We use the notation $[b^i_j]$ to denote the matrix whose (i,j)-entry is b^i_j.

Proposition 3.27 (Wedge product of 1-covectors). *If $\alpha^1,\dots,\alpha^k$ are linear functions on a vector space V and $v_1,\dots,v_k\in V$, then*

$$(\alpha^1\wedge\cdots\wedge\alpha^k)(v_1,\dots,v_k) = \det[\alpha^i(v_j)].$$

Proof. By (3.8),

$$\begin{aligned}(\alpha^1\wedge\cdots\wedge\alpha^k)(v_1,\dots,v_k) &= A(\alpha^1\otimes\cdots\otimes\alpha^k)(v_1,\dots,v_k)\\ &= \sum_{\sigma\in S_k}(\operatorname{sgn}\sigma)\alpha^1\left(v_{\sigma(1)}\right)\cdots\alpha^k\left(v_{\sigma(k)}\right)\\ &= \det[\alpha^i(v_j)].\end{aligned}$$ □

An algebra A over a field K is said to be *graded* if it can be written as a direct sum $A = \bigoplus_{k=0}^{\infty} A^k$ of vector spaces over K so that the multiplication map sends $A^k \times A^\ell$ to $A^{k+\ell}$. The notation $A = \bigoplus_{k=0}^{\infty} A^k$ means that each nonzero element of A is uniquely a *finite* sum

$$a = a_{i_1} + \cdots + a_{i_m},$$

where $a_{i_j} \neq 0 \in A^{i_j}$. A graded algebra $A = \oplus_{k=0}^{\infty} A^k$ is said to be *anticommutative* or *graded commutative* if for all $a \in A^k$ and $b \in A^\ell$,

$$ab = (-1)^{k\ell} ba.$$

A *homomorphism of graded algebras* is an algebra homomorphism that preserves the degree.

Example. The polynomial algebra $A = \mathbb{R}[x,y]$ is graded by degree; A^k consists of all homogeneous polynomials of total degree k in the variables x and y.

For a finite-dimensional vector space V, say of dimension n, define

$$A_*(V) = \bigoplus_{k=0}^{\infty} A_k(V) = \bigoplus_{k=0}^{n} A_k(V).$$

With the wedge product of multicovectors as multiplication, $A_*(V)$ becomes an anticommutative graded algebra, called the *exterior algebra* or the *Grassmann algebra* of multicovectors on the vector space V.

3.10 A Basis for k-Covectors

Let $e_1, \dots, e_n$ be a basis for a real vector space V, and let $\alpha^1, \dots, \alpha^n$ be the dual basis for $V^\vee$. Introduce the multi-index notation

$$I = (i_1, \dots, i_k)$$

and write e_I for $(e_{i_1}, \dots, e_{i_k})$ and α^I for $\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_k}$.

A k-linear function f on V is completely determined by its values on all k-tuples $(e_{i_1}, \dots, e_{i_k})$. If f is alternating, then it is completely determined by its values on $(e_{i_1}, \dots, e_{i_k})$ with $1 \le i_1 < \cdots < i_k \le n$; that is, it suffices to consider e_I with I in *strictly ascending* order.

Lemma 3.28. *Let $e_1, \dots, e_n$ be a basis for a vector space V and let $\alpha^1, \dots, \alpha^n$ be its dual basis in $V^\vee$. If $I = (1 \le i_1 < \cdots < i_k \le n)$ and $J = (1 \le j_1 < \cdots < j_k \le n)$ are strictly ascending multi-indices of length k, then*

$$\alpha^I(e_J) = \delta^I_J = \begin{cases} 1 & \text{for } I = J, \\ 0 & \text{for } I \neq J. \end{cases}$$

Proof. By Proposition 3.27,

$$\alpha^I(e_J) = \det[\alpha^i(e_j)]_{i\in I, j\in J}.$$

If $I = J$, then $[\alpha^i(e_j)]$ is the identity matrix and its determinant is 1.

If $I \neq J$, we compare them term by term until the terms differ:

$$i_1 = j_1, \ \ldots, \ i_{\ell-1} = j_{\ell-1}, \ i_\ell \neq j_\ell, \ \ldots \ .$$

Without loss of generality, we may assume $i_\ell < j_\ell$. Then i_ℓ will be different from $j_1, \ldots, j_{\ell-1}$ (because these are the same as $i_1, \ldots, i_\ell$, and I is strictly ascending), and i_ℓ will also be different from $j_\ell, j_{\ell+1}, \ldots, j_k$ (because J is strictly ascending). Thus, i_ℓ will be different from $j_1, \ldots, j_k$, and the ℓth row of the matrix $[a^i(e_j)]$ will be all zero. Hence, $\det[a^i(e_j)] = 0$. □

Proposition 3.29. *The alternating k-linear functions α^I, $I = (i_1 < \cdots < i_k)$, form a basis for the space $A_k(V)$ of alternating k-linear functions on V.*

Proof. First, we show linear independence. Suppose $\sum c_I \alpha^I = 0$, $c_I \in \mathbb{R}$, and I runs over all strictly ascending multi-indices of length k. Applying both sides to e_J, $J = (j_1 < \cdots < j_k)$, we get by Lemma 3.28,

$$0 = \sum_I c_I \alpha^I(e_J) = \sum_I c_I \delta^I_J = c_J,$$

since among all strictly ascending multi-indices I of length k, there is only one equal to J. This proves that the α^I are linearly independent.

To show that the α^I span $A_k(V)$, let $f \in A_k(V)$. We claim that

$$f = \sum f(e_I)\alpha^I,$$

where I runs over all strictly ascending multi-indices of length k. Let $g = \sum f(e_I)\alpha^I$. By k-linearity and the alternating property, if two k-covectors agree on all e_J, where $J = (j_1 < \cdots < j_k)$, then they are equal. But

$$g(e_J) = \sum f(e_I)\alpha^I(e_J) = \sum f(e_I)\delta^I_J = f(e_J).$$

Therefore, $f = g = \sum f(e_I)\alpha^I$. □

Corollary 3.30. *If the vector space V has dimension n, then the vector space $A_k(V)$ of k-covectors on V has dimension $\binom{n}{k}$.*

Proof. A strictly ascending multi-index $I = (i_1 < \cdots < i_k)$ is obtained by choosing a subset of k numbers from $1, \ldots, n$. This can be done in $\binom{n}{k}$ ways. □

Corollary 3.31. *If $k > \dim V$, then $A_k(V) = 0$.*

Proof. In $\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_k}$, at least two of the factors must be the same, say $\alpha^j = \alpha^\ell = \alpha$. Because α is a 1-covector, $\alpha \wedge \alpha = 0$ by Corollary 3.23, so $\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_k} = 0$. □

Problems

3.1. Tensor product of covectors
Let $e_1, \dots, e_n$ be a basis for a vector space V and let $\alpha^1, \dots, \alpha^n$ be its dual basis in $V^\vee$. Suppose $[g_{ij}] \in \mathbb{R}^{n\times n}$ is an $n \times n$ matrix. Define a bilinear function $f\colon V \times V \to \mathbb{R}$ by

$$f(v,w) = \sum_{1\le i,j\le n} g_{ij} v^i w^j$$

for $v = \sum v^i e_i$ and $w = \sum w^j e_j$ in V. Describe f in terms of the tensor products of α^i and α^j, $1 \le i, j \le n$.

3.2. Hyperplanes

(a) Let V be a vector space of dimension n and $f\colon V \to \mathbb{R}$ a nonzero linear functional. Show that $\dim \ker f = n-1$. A linear subspace of V of dimension $n-1$ is called a *hyperplane* in V.
(b) Show that a nonzero linear functional on a vector space V is determined up to a multiplicative constant by its kernel, a hyperplane in V. In other words, if f and $g\colon V \to \mathbb{R}$ are nonzero linear functionals and $\ker f = \ker g$, then $g = cf$ for some constant $c \in \mathbb{R}$.

3.3. A basis for k-tensors
Let V be a vector space of dimension n with basis $e_1, \dots, e_n$. Let $\alpha^1, \dots, \alpha^n$ be the dual basis for $V^\vee$. Show that a basis for the space $L_k(V)$ of k-linear functions on V is $\{\alpha^{i_1} \otimes \cdots \otimes \alpha^{i_k}\}$ for all multi-indices $(i_1, \dots, i_k)$ (not just the strictly ascending multi-indices as for $A_k(L)$). In particular, this shows that $\dim L_k(V) = n^k$. (This problem generalizes Problem 3.1.)

3.4. A characterization of alternating k-tensors
Let f be a k-tensor on a vector space V. Prove that f is alternating if and only if f changes sign whenever two successive arguments are interchanged:

$$f(\dots, v_{i+1}, v_i, \dots) = -f(\dots, v_i, v_{i+1}, \dots)$$

for $i = 1, \dots, k-1$.

3.5. Another characterization of alternating k-tensors
Let f be a k-tensor on a vector space V. Prove that f is alternating if and only if $f(v_1, \dots, v_k) = 0$ whenever two of the vectors $v_1, \dots, v_k$ are equal.

3.6. Wedge product and scalars
Let V be a vector space. For $a, b \in \mathbb{R}$, $f \in A_k(V)$ and $g \in A_\ell(V)$, show that $af \wedge bg = (ab) f \wedge g$.

3.7. Transformation rule for a wedge product of covectors
Suppose two sets of covectors on a vector space V, $\beta^1, \dots, \beta^k$ and $\gamma^1, \dots, \gamma^k$, are related by

$$\beta^i = \sum_{j=1}^{k} a^i_j \gamma^j, \quad i = 1, \ldots, k,$$

for a $k \times k$ matrix $A = [a^i_j]$. Show that

$$\beta^1 \wedge \cdots \wedge \beta^k = (\det A)\, \gamma^1 \wedge \cdots \wedge \gamma^k.$$

3.8. Transformation rule for k-covectors
Let f be a k-covector on a vector space V. Suppose two sets of vectors $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$ in V are related by

$$u_j = \sum_{i=1}^{k} a^i_j v_i, \quad j = 1, \ldots, k,$$

for a $k \times k$ matrix $A = [a^i_j]$. Show that

$$f(u_1, \ldots, u_k) = (\det A)\, f(v_1, \ldots, v_k).$$

3.9. Vanishing of a covector of top degree
Let V be a vector space of dimension n. Prove that if an n-covector ω vanishes on a basis $e_1, \ldots, e_n$ for V, then ω is the zero covector on V.

3.10.* Linear independence of covectors
Let $\alpha^1, \ldots, \alpha^k$ be 1-covectors on a vector space V. Show that $\alpha^1 \wedge \cdots \wedge \alpha^k \neq 0$ if and only if $\alpha^1, \ldots, \alpha^k$ are linearly independent in the dual space $V^\vee$.

3.11.* Exterior multiplication
Let α be a nonzero 1-covector and γ a k-covector on a finite-dimensional vector space V. Show that $\alpha \wedge \gamma = 0$ if and only if $\gamma = \alpha \wedge \beta$ for some $(k-1)$-covector β on V.

4

Differential Forms on $\mathbb{R}^n$

Just as a vector field assigns a tangent vector to each point of an open subset U of $\mathbb{R}^n$, so dually a differential k-form assigns a k-covector on the tangent space to each point of U. The wedge product of differential forms is defined pointwise as the wedge product of multicovectors. Since differential forms exist on an open set, not just at a single point, there is a notion of differentiation for differential forms. In fact, there is a unique one, called the *exterior derivative*, characterized by three natural properties. Although we define it using the standard coordinates of $\mathbb{R}^n$, the exterior derivative turns out to be independent of coordinates, as we shall see later, and is therefore intrinsic to a manifold. It is the ultimate abstract extension to a manifold of the gradient, curl, and divergence of vector calculus in $\mathbb{R}^3$. Differential forms extend Grassmann's exterior algebra from the tangent space at a point globally to an entire manifold. Since its creation around the turn of the twentieth century, generally credited to É. Cartan [5] and H. Poincaré [34], the calculus of differential forms has had far-reaching consequences in geometry, topology, and physics. In fact, certain physical concepts such as electricity and magnetism are best formulated in terms of differential forms.

In this section we will study the simplest case, that of differential forms on an open subset of $\mathbb{R}^n$. Even in this setting, differential forms already provide a way to unify the main theorems of vector calculus in $\mathbb{R}^3$.

4.1 Differential 1-Forms and the Differential of a Function

The *cotangent space* to $\mathbb{R}^n$ at p, denoted by $T_p^*(\mathbb{R}^n)$ or $T_p^*\mathbb{R}^n$, is defined to be the dual space $(T_p\mathbb{R}^n)^\vee$ of the tangent space $T_p(\mathbb{R}^n)$. Thus, an element of the cotangent space $T_p^*(\mathbb{R}^n)$ is a covector or a linear functional on the tangent space $T_p(\mathbb{R}^n)$. In parallel with the definition of a vector field, a *covector field* or a *differential* 1*-form* on an open subset U of $\mathbb{R}^n$ is a function ω that assigns to each point p in U a covector $\omega_p \in T_p^*(\mathbb{R}^n)$,

$$\omega\colon U \to \bigsqcup_{p\in U} T_p^*(\mathbb{R}^n),$$
$$p \mapsto \omega_p \in T_p^*(\mathbb{R}^n).$$

Here the notation $\bigsqcup$ stands for "disjoint union," meaning that the sets $T_p^*(\mathbb{R}^n)$ are all disjoint. We call a differential 1-form a 1-*form* for short.

From any C^∞ function $f\colon U \to \mathbb{R}$, we can construct a 1-form df, called the *differential* of f, as follows. For $p \in U$ and $X_p \in T_pU$, define

$$(df)_p(X_p) = X_p f.$$

A few words may be in order about the definition of the differential. The directional derivative of a function in the direction of a tangent vector at a point p sets up a bilinear pairing

$$T_p(\mathbb{R}^n) \times C_p^\infty(\mathbb{R}^n) \to \mathbb{R},$$
$$(X_p, f) \mapsto \langle X_p, f\rangle = X_p f.$$

One may think of a tangent vector as a function on the second argument of this pairing: $\langle X_p, \cdot\,\rangle$. The differential $(df)_p$ at p is a function on the first argument of the pairing:

$$(df)_p = \langle\,\cdot\,, f\rangle.$$

The value of the differential df at p is also written $df|_p$.

Let $x^1,\dots,x^n$ be the standard coordinates on $\mathbb{R}^n$. We saw in Subsection 2.3 that the set $\{\partial/\partial x^1|_p,\dots,\partial/\partial x^n|_p\}$ is a basis for the tangent space $T_p(\mathbb{R}^n)$.

Proposition 4.1. *If $x^1,\dots,x^n$ are the standard coordinates on $\mathbb{R}^n$, then at each point $p\in\mathbb{R}^n$, $\{(dx^1)_p,\dots,(dx^n)_p\}$ is the basis for the cotangent space $T_p^*(\mathbb{R}^n)$ dual to the basis $\{\partial/\partial x^1|_p,\dots,\partial/\partial x^n|_p\}$ for the tangent space $T_p(\mathbb{R}^n)$.*

Proof. By definition,

$$(dx^i)_p\left(\left.\frac{\partial}{\partial x^j}\right|_p\right) = \left.\frac{\partial}{\partial x^j}\right|_p x^i = \delta^i_j. \qquad \square$$

If ω is a 1-form on an open subset U of $\mathbb{R}^n$, then by Proposition 4.1, at each point p in U, ω can be written as a linear combination

$$\omega_p = \sum a_i(p)\,(dx^i)_p,$$

for some $a_i(p)\in\mathbb{R}$. As p varies over U, the coefficients a_i become functions on U, and we may write $\omega = \sum a_i\,dx^i$. The covector field ω is said to be C^∞ *on* U if the coefficient functions a_i are all C^∞ on U.

If x, y, and z are the coordinates on $\mathbb{R}^3$, then dx, dy, and dz are 1-forms on $\mathbb{R}^3$. In this way, we give meaning to what was merely a notation in elementary calculus.

Proposition 4.2 (The differential in terms of coordinates). *If $f\colon U \to \mathbb{R}$ is a C^∞ function on an open set U in $\mathbb{R}^n$, then*

$$df = \sum \frac{\partial f}{\partial x^i}\, dx^i. \tag{4.1}$$

Proof. By Proposition 4.1, at each point p in U,

$$(df)_p = \sum a_i(p)\,(dx^i)_p \tag{4.2}$$

for some real numbers $a_i(p)$ depending on p. Thus, $df = \sum a_i\, dx^i$ for some real functions a_i on U. To find a_j, apply both sides of (4.2) to the coordinate vector field $\partial/\partial x^j$:

$$df\left(\frac{\partial}{\partial x^j}\right) = \sum_i a_i\, dx^i\left(\frac{\partial}{\partial x^j}\right) = \sum_i a_i \delta^i_j = a_j.$$

On the other hand, by the definition of the differential,

$$df\left(\frac{\partial}{\partial x^j}\right) = \frac{\partial f}{\partial x^j}. \qquad \square$$

Equation (4.1) shows that if f is a C^∞ function, then the 1-form df is also C^∞.

Example. Differential 1-forms arise naturally even if one is interested only in tangent vectors. Every tangent vector $X_p \in T_p(\mathbb{R}^n)$ is a linear combination of the standard basis vectors:

$$X_p = \sum_i b^i(X_p)\, \frac{\partial}{\partial x^i}\bigg|_p .$$

In Example 3.3 we saw that at each point $p \in \mathbb{R}^n$, we have $b^i(X_p) = (dx^i)_p(X_p)$. Hence, the coefficient b^i of a vector at p with respect to the standard basis $\partial/\partial x^1|_p$, $\ldots$, $\partial/\partial x^n|_p$ is none other than the dual covector $dx^i|_p$ on $\mathbb{R}^n$. As p varies, $b^i = dx^i$.

4.2 Differential k-Forms

More generally, a *differential form ω of degree k* or a *k-form* on an open subset U of $\mathbb{R}^n$ is a function that assigns to each point p in U an alternating k-linear function on the tangent space $T_p(\mathbb{R}^n)$, i.e., $\omega_p \in A_k(T_p\mathbb{R}^n)$. Since $A_1(T_p\mathbb{R}^n) = T_p^*(\mathbb{R}^n)$, the definition of a k-form generalizes that of a 1-form in Subsection 4.1.

By Proposition 3.29, a basis for $A_k(T_p\mathbb{R}^n)$ is

$$dx^I_p = dx^{i_1}_p \wedge \cdots \wedge dx^{i_k}_p, \quad 1 \le i_1 < \cdots < i_k \le n.$$

Therefore, at each point p in U, ω_p is a linear combination

$$\omega_p = \sum a_I(p)\, dx^I_p, \quad 1 \le i_1 < \cdots < i_k \le n,$$

and a k-form ω on U is a linear combination

$$\omega = \sum a_I \, dx^I,$$

with function coefficients $a_I : U \to \mathbb{R}$. We say that a k-form ω is C^∞ on U if all the coefficients a_I are C^∞ functions on U.

Denote by $\Omega^k(U)$ the vector space of C^∞ k-forms on U. A 0-form on U assigns to each point p in U an element of $A_0(T_p\mathbb{R}^n) = \mathbb{R}$. Thus, a 0-form on U is simply a function on U, and $\Omega^0(U) = C^\infty(U)$.

There are no nonzero differential forms of degree $> n$ on an open subset of $\mathbb{R}^n$. This is because if $\deg dx^I > n$, then in the expression dx^I at least two of the 1-forms dx^{i_α} must be the same, forcing $dx^I = 0$.

The *wedge product* of a k-form ω and an ℓ-form τ on an open set U is defined pointwise:

$$(\omega \wedge \tau)_p = \omega_p \wedge \tau_p, \quad p \in U.$$

In terms of coordinates, if $\omega = \sum_I a_I \, dx^I$ and $\tau = \sum_J b_J \, dx^J$, then

$$\omega \wedge \tau = \sum_{I,J} (a_I b_J) \, dx^I \wedge dx^J.$$

In this sum, if I and J are not disjoint on the right-hand side, then $dx^I \wedge dx^J = 0$. Hence, the sum is actually over disjoint multi-indices:

$$\omega \wedge \tau = \sum_{I,J \text{ disjoint}} (a_I b_J) \, dx^I \wedge dx^J,$$

which shows that the wedge product of two C^∞ forms is C^∞. So the wedge product is a bilinear map

$$\wedge : \Omega^k(U) \times \Omega^\ell(U) \to \Omega^{k+\ell}(U).$$

By Propositions 3.21 and 3.25, the wedge product of differential forms is anticommutative and associative.

In case one of the factors has degree 0, say $k = 0$, the wedge product

$$\wedge : \Omega^0(U) \times \Omega^\ell(U) \to \Omega^\ell(U)$$

is the pointwise multiplication of a C^∞ ℓ-form by a C^∞ function:

$$(f \wedge \omega)_p = f(p) \wedge \omega_p = f(p)\omega_p,$$

since as we noted in Subsection 3.7, the wedge product with a 0-covector is scalar multiplication. Thus, if $f \in C^\infty(U)$ and $\omega \in \Omega^\ell(U)$, then $f \wedge \omega = f\omega$.

Example. Let x, y, z be the coordinates on $\mathbb{R}^3$. The C^∞ 1-forms on $\mathbb{R}^3$ are

$$f \, dx + g \, dy + h \, dz,$$

where f, g, h range over all C^∞ functions on $\mathbb{R}^3$. The C^∞ 2-forms are

$$f \, dy \wedge dz + g \, dx \wedge dz + h \, dx \wedge dy$$

and the C^∞ 3-forms are

$$f \, dx \wedge dy \wedge dz.$$

Exercise 4.3 (A basis for 3-covectors).* Let x^1, x^2, x^3, x^4 be the coordinates on $\mathbb{R}^4$ and p a point in $\mathbb{R}^4$. Write down a basis for the vector space $A_3(T_p(\mathbb{R}^4))$.

With the wedge product as multiplication and the degree of a form as the grading, the direct sum $\Omega^*(U) = \bigoplus_{k=0}^n \Omega^k(U)$ becomes an anticommutative graded algebra over $\mathbb{R}$. Since one can multiply C^∞ k-forms by C^∞ functions, the set $\Omega^k(U)$ of C^∞ k-forms on U is both a vector space over $\mathbb{R}$ and a module over $C^\infty(U)$, and so the direct sum $\Omega^*(U) = \bigoplus_{k=0}^n \Omega^k(U)$ is also a module over the ring $C^\infty(U)$ of C^∞ functions.

4.3 Differential Forms as Multilinear Functions on Vector Fields

If ω is a C^∞ 1-form and X is a C^∞ vector field on an open set U in $\mathbb{R}^n$, we define a function $\omega(X)$ on U by the formula

$$\omega(X)_p = \omega_p(X_p), \quad p \in U.$$

Written out in coordinates,

$$\omega = \sum a_i\, dx^i, \qquad X = \sum b^j \frac{\partial}{\partial x^j} \qquad \text{for some } a_i, b^j \in C^\infty(U),$$

so

$$\omega(X) = \left(\sum a_i\, dx^i\right)\left(\sum b^j \frac{\partial}{\partial x^j}\right) = \sum a_i b^i,$$

which shows that $\omega(X)$ is C^∞ on U. Thus, a C^∞ 1-form on U gives rise to a map from $\mathfrak{X}(U)$ to $C^\infty(U)$.

This function is actually linear over the ring $C^\infty(U)$; i.e., if $f \in C^\infty(U)$, then $\omega(fX) = f\omega(X)$. To show this, it suffices to evaluate $\omega(fX)$ at an arbitrary point $p \in U$:

$$\begin{aligned}(\omega(fX))_p &= \omega_p(f(p)X_p) && \text{(definition of } \omega(fX))\\ &= f(p)\omega_p(X_p) && (\omega_p \text{ is } \mathbb{R}\text{-linear})\\ &= (f\omega(X))_p && \text{(definitition of } f\omega(X)).\end{aligned}$$

Let $\mathcal{F}(U) = C^\infty(U)$. In this notation, a 1-form ω on U gives rise to an $\mathcal{F}(U)$-linear map $\mathfrak{X}(U) \to \mathcal{F}(U)$, $X \mapsto \omega(X)$. Similarly, a k-form ω on U gives rise to a k-linear map over $\mathcal{F}(U)$

$$\underbrace{\mathfrak{X}(U) \times \cdots \times \mathfrak{X}(U)}_{k \text{ times}} \to \mathcal{F}(U),$$
$$(X_1, \ldots, X_k) \mapsto \omega(X_1, \ldots, X_k).$$

Exercise 4.4 (Wedge product of a 2-form with a 1-form).* Let ω be a 2-form and τ a 1-form on $\mathbb{R}^3$. If X, Y, Z are vector fields on M, find an explicit formula for $(\omega \wedge \tau)(X, Y, Z)$ in terms of the values of ω and τ on the vector fields X, Y, Z.

4.4 The Exterior Derivative

To define the *exterior derivative* of a C^∞ k-form on an open subset U of $\mathbb{R}^n$, we first define it on 0-forms: the exterior derivative of a C^∞ function $f \in C^\infty(U)$ is defined to be its differential $df \in \Omega^1(U)$; in terms of coordinates, Proposition 4.2 gives

$$df = \sum \frac{\partial f}{\partial x^i}\, dx^i.$$

Definition 4.5. For $k \geq 1$, if $\omega = \sum_I a_I\, dx^I \in \Omega^k(U)$, then

$$d\omega = \sum_I da_I \wedge dx^I = \sum_I \left(\sum_j \frac{\partial a_I}{\partial x^j}\, dx^j \right) \wedge dx^I \in \Omega^{k+1}(U).$$

Example. Let ω be the 1-form $f\,dx + g\,dy$ on $\mathbb{R}^2$, where f and g are C^∞ functions on $\mathbb{R}^2$. To simplify the notation, write $f_x = \partial f/\partial x,\ f_y = \partial f/\partial y$. Then

$$\begin{aligned} d\omega &= df \wedge dx + dg \wedge dy \\ &= (f_x\, dx + f_y\, dy) \wedge dx + (g_x\, dx + g_y\, dy) \wedge dy \\ &= (g_x - f_y)\, dx \wedge dy. \end{aligned}$$

In this computation $dy \wedge dx = -dx \wedge dy$ and $dx \wedge dx = dy \wedge dy = 0$ by the anticommutative property of the wedge product (Proposition 3.21 and Corollary 3.23).

Definition 4.6. Let $A = \oplus_{k=0}^{\infty} A^k$ be a graded algebra over a field K. An *antiderivation* of the graded algebra A is a K-linear map $D\colon A \to A$ such that for $a \in A^k$ and $b \in A^\ell$,

$$D(ab) = (Da)b + (-1)^k a Db. \tag{4.3}$$

If there is an integer m so that the antiderivation D sends A^k to A^{k+m} for all k, then we say that it is an antiderivation of *degree m*. By defining $A_k = 0$ for $k < 0$, we can extend the grading of a graded algebra A to negative integers. With this extension, the degree m of an antiderivation can be negative. (An example of an antiderivation of degree -1 is interior multiplication, to be discussed in Subsection 20.4.)

Proposition 4.7.
(i) *The exterior differentiation* $d\colon \Omega^*(U) \to \Omega^*(U)$ *is an antiderivation of degree* 1:

$$d(\omega \wedge \tau) = (d\omega) \wedge \tau + (-1)^{\deg \omega}\, \omega \wedge d\tau.$$

(ii) $d^2 = 0$.
(iii) *If* $f \in C^\infty(U)$ *and* $X \in \mathfrak{X}(U)$, *then* $(df)(X) = Xf$.

Proof.
(i) Since both sides of (4.3) are linear in ω and in τ, it suffices to check the equality for $\omega = f\,dx^I$ and $\tau = g\,dx^J$. Then

$$\begin{aligned} d(\omega \wedge \tau) &= d(fg\,dx^I \wedge dx^J) \\ &= \sum \frac{\partial(fg)}{\partial x^i}\,dx^i \wedge dx^I \wedge dx^J \\ &= \sum \frac{\partial f}{\partial x^i} g\,dx^i \wedge dx^I \wedge dx^J + \sum f \frac{\partial g}{\partial x^i}\,dx^i \wedge dx^I \wedge dx^J. \end{aligned}$$

In the second sum, moving the 1-form $(\partial g/\partial x^i)\,dx^i$ across the k-form dx^I results in the sign $(-1)^k$ by anticommutativity. Hence,

$$\begin{aligned} d(\omega \wedge \tau) &= \sum \frac{\partial f}{\partial x^i}\,dx^i \wedge dx^I \wedge g\,dx^J + (-1)^k \sum f\,dx^I \wedge \frac{\partial g}{\partial x^i}\,dx^i \wedge dx^J \\ &= d\omega \wedge \tau + (-1)^k \omega \wedge d\tau. \end{aligned}$$

(ii) Again by the $\mathbb{R}$-linearity of d, it suffices to show that $d^2\omega = 0$ for $\omega = f\,dx^I$. We compute:

$$\begin{aligned} d^2(f\,dx^I) &= d\left(\sum \frac{\partial f}{\partial x^i}\,dx^i \wedge dx^I\right) \\ &= \sum \frac{\partial^2 f}{\partial x^j \partial x^i}\,dx^j \wedge dx^i \wedge dx^I. \end{aligned}$$

In this sum if $i = j$, then $dx^j \wedge dx^i = 0$; if $i \neq j$, then $\partial^2 f/\partial x^i \partial x^j$ is symmetric in i and j, but $dx^j \wedge dx^i$ is alternating in i and j, so the terms with $i \neq j$ pair up and cancel each other. For example,

$$\begin{aligned} &\frac{\partial^2 f}{\partial x^1 \partial x^2}\,dx^1 \wedge dx^2 + \frac{\partial^2 f}{\partial x^2 \partial x^1}\,dx^2 \wedge dx^1 \\ &= \frac{\partial^2 f}{\partial x^1 \partial x^2}\,dx^1 \wedge dx^2 + \frac{\partial^2 f}{\partial x^1 \partial x^2}(-dx^1 \wedge dx^2) = 0. \end{aligned}$$

Therefore, $d^2(f\,dx^I) = 0$.
(iii) This is simply the definition of the exterior derivative of a function as the differential of the function. □

Proposition 4.8 (Characterization of the exterior derivative). *The three properties of Proposition 4.7 uniquely characterize exterior differentiation on an open set U in $\mathbb{R}^n$; that is, if $D\colon \Omega^*(U) \to \Omega^*(U)$ is* (i) *an antiderivation of degree* 1 *such that* (ii) $D^2 = 0$ *and* (iii) $(Df)(X) = Xf$ *for* $f \in C^\infty(U)$ *and* $X \in \mathfrak{X}(U)$, *then* $D = d$.

Proof. Since every k-form on U is a sum of terms such as $f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, by linearity it suffices to show that $D = d$ on a k-form of this type. By (iii), $Df = df$ on C^∞ functions. It follows that $D\,dx^i = DDx^i = 0$ by (ii). A simple induction on k, using the antiderivation property of D, proves that for all k and all multi-indices I of length k,

$$D(dx^I) = D(dx^{i_1} \wedge \cdots \wedge dx^{i_k}) = 0. \tag{4.4}$$

Finally, for every k-form $f\,dx^I$,

$$\begin{aligned} D(f\,dx^I) &= (Df) \wedge dx^I + f\,D(dx^I) && \text{(by (i))} \\ &= (df) \wedge dx^I && \text{(by (ii) and (4.4))} \\ &= d(f\,dx^I) && \text{(definition of } d\text{)}. \end{aligned}$$

Hence, $D = d$ on $\Omega^*(U)$. □

4.5 Closed Forms and Exact Forms

A k-form ω on U is *closed* if $d\omega = 0$; it is *exact* if there is a $(k-1)$-form τ such that $\omega = d\tau$ on U. Since $d(d\tau) = 0$, every exact form is closed. In the next section we will discuss the meaning of closed and exact forms in the context of vector calculus on $\mathbb{R}^3$.

Exercise 4.9 (A closed 1-form on the punctured plane). Define a 1-form ω on $\mathbb{R}^2 - \{\mathbf{0}\}$ by

$$\omega = \frac{1}{x^2+y^2}(-y\,dx + x\,dy).$$

Show that ω is closed.

A collection of vector spaces $\{V^k\}_{k=0}^{\infty}$ with linear maps $d_k \colon V^k \to V^{k+1}$ such that $d_{k+1} \circ d_k = 0$ is called a *differential complex* or a *cochain complex*. For any open subset U of $\mathbb{R}^n$, the exterior derivative d makes the vector space $\Omega^*(U)$ of C^∞ forms on U into a cochain complex, called the *de Rham complex* of U:

$$0 \to \Omega^0(U) \xrightarrow{d} \Omega^1(U) \xrightarrow{d} \Omega^2(U) \to \cdots.$$

The closed forms are precisely the elements of the kernel of d and the exact forms are the elements of the image of d.

4.6 Applications to Vector Calculus

The theory of differential forms unifies many theorems in vector calculus on $\mathbb{R}^3$. We summarize here some results from vector calculus and then show how they fit into the framework of differential forms.

By a *vector-valued function* on an open subset U of $\mathbb{R}^3$, we mean a function $\mathbf{F} = \langle P, Q, R \rangle \colon U \to \mathbb{R}^3$. Such a function assigns to each point $p \in U$ a vector $\mathbf{F}_p \in \mathbb{R}^3 \simeq T_p(\mathbb{R}^3)$. Hence, a vector-valued function on U is precisely a vector field on U.

Recall the three operators gradient, curl, and divergence on scalar- and vector-valued functions on U:

$$\{\text{scalar func.}\} \overset{\text{grad}}{\to} \{\text{vector func.}\} \overset{\text{curl}}{\to} \{\text{vector func.}\} \overset{\text{div}}{\to} \{\text{scalar func.}\}$$

$$\begin{aligned}
\operatorname{grad} f &= \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} f = \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}, \\
\operatorname{curl} \begin{bmatrix} P \\ Q \\ R \end{bmatrix} &= \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} \times \begin{bmatrix} P \\ Q \\ R \end{bmatrix} = \begin{bmatrix} R_y - Q_z \\ -(R_x - P_z) \\ Q_x - P_y \end{bmatrix}, \\
\operatorname{div} \begin{bmatrix} P \\ Q \\ R \end{bmatrix} &= \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} \cdot \begin{bmatrix} P \\ Q \\ R \end{bmatrix} = P_x + Q_y + R_z.
\end{aligned}$$

Since every 1-form on U is a linear combination with function coefficients of dx, dy, and dz, we can identify 1-forms with vector fields on U via

$$P\,dx + Q\,dy + R\,dz \longleftrightarrow \begin{bmatrix} P \\ Q \\ R \end{bmatrix}.$$

Similarly, 2-forms on U can also be identified with vector fields on U:

$$P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy \longleftrightarrow \begin{bmatrix} P \\ Q \\ R \end{bmatrix},$$

and 3-forms on U can be identified with functions on U:

$$f\,dx \wedge dy \wedge dz \longleftrightarrow f.$$

In terms of these identifications, the exterior derivative of a 0-form f is

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \longleftrightarrow \begin{bmatrix} \partial f/\partial x \\ \partial f/\partial y \\ \partial f/\partial x \end{bmatrix} = \operatorname{grad} f;$$

the exterior derivative of a 1-form is

$$\begin{aligned}
&d(P\,dx + Q\,dy + R\,dz) \\
&\quad = (R_y - Q_z)\,dy \wedge dz - (R_x - P_z)\,dz \wedge dx + (Q_x - P_y)\,dx \wedge dy, \qquad (4.5)
\end{aligned}$$

which corresponds to

$$\operatorname{curl} \begin{bmatrix} P \\ Q \\ R \end{bmatrix} = \begin{bmatrix} R_y - Q_z \\ -(R_x - P_z) \\ Q_x - P_y \end{bmatrix};$$

the exterior derivative of a 2-form is

$$\begin{aligned} &d(P\,dy\wedge dz + Q\,dz\wedge dx + R\,dx\wedge dy) \\ &\quad = (P_x + Q_y + R_z)\,dx\wedge dy\wedge dz, \end{aligned} \tag{4.6}$$

which corresponds to

$$\operatorname{div}\begin{bmatrix} P \\ Q \\ R \end{bmatrix} = P_x + Q_y + R_z.$$

Thus, after appropriate identifications, the exterior derivatives d on 0-forms, 1-forms, and 2-forms are simply the three operators grad, curl, and div. In summary, on an open subset U of $\mathbb{R}^3$, there are identifications

$$\begin{array}{ccccccc} \Omega^0(U) & \xrightarrow{d} & \Omega^1(U) & \xrightarrow{d} & \Omega^2(U) & \xrightarrow{d} & \Omega^3(U) \\ \simeq\big\downarrow & & \simeq\big\downarrow & & \simeq\big\downarrow & & \simeq\big\downarrow \\ C^\infty(U) & \xrightarrow[\text{grad}]{} & \mathfrak{X}(U) & \xrightarrow[\text{curl}]{} & \mathfrak{X}(U) & \xrightarrow[\text{div}]{} & C^\infty(U). \end{array}$$

Under these identifications, a vector field $\langle P, Q, R\rangle$ on $\mathbb{R}^3$ is the gradient of a C^∞ function f if and only if the corresponding 1-form $P\,dx + Q\,dy + R\,dz$ is df.

Next we recall three basic facts from calculus concerning grad, curl, and div.

Proposition A. $\operatorname{curl}(\operatorname{grad} f) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Proposition B. $\operatorname{div}\left(\operatorname{curl}\begin{bmatrix} P \\ Q \\ R \end{bmatrix}\right) = 0$.

Proposition C. *On $\mathbb{R}^3$, a vector field $\mathbf{F}$ is the gradient of some scalar function f if and only if* $\operatorname{curl} \mathbf{F} = 0$.

Propositions A and B express the property $d^2 = 0$ of the exterior derivative on open subsets of $\mathbb{R}^3$; these are easy computations. Proposition C expresses the fact that a 1-form on $\mathbb{R}^3$ is exact if and only if it is closed. Proposition C need not be true on a region other than $\mathbb{R}^3$, as the following well-known example from calculus shows.

Example. If $U = \mathbb{R}^3 - \{z\text{-axis}\}$, and $\mathbf{F}$ is the vector field

$$\mathbf{F} = \left\langle \frac{-y}{x^2+y^2},\ \frac{x}{x^2+y^2},\ 0 \right\rangle$$

on $\mathbb{R}^3$, then $\operatorname{curl}\mathbf{F} = \mathbf{0}$, but $\mathbf{F}$ is not the gradient of any C^∞ function on U. The reason is that if $\mathbf{F}$ were the gradient of a C^∞ function f on U, then by the fundamental theorem for line integrals, the line integral

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

over any closed curve C would be zero. However, on the unit circle C in the (x,y)-plane, with $x = \cos t$ and $y = \sin t$ for $0 \le t \le 2\pi$, this integral is

$$\int_C -y\,dx + x\,dy = \int_0^{2\pi} -(\sin t)\,d\cos t + (\cos t)\,d\sin t = 2\pi.$$

In terms of differential forms, the 1-form

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

is closed but not exact on U. (This 1-form is defined by the same formula as the 1-form ω in Exercise 4.9, but is defined on a different space.)

It turns out that whether Proposition C is true for a region U depends only on the topology of U. One measure of the failure of a closed k-form to be exact is the quotient vector space

$$H^k(U) := \frac{\{\text{closed } k\text{-forms on } U\}}{\{\text{exact } k\text{-forms on } U\}},$$

called the kth *de Rham cohomology* of U.

The generalization of Proposition C to any differential form on $\mathbb{R}^n$ is called the *Poincaré lemma*: for $k \ge 1$, every closed k-form on $\mathbb{R}^n$ is exact. This is of course equivalent to the vanishing of the kth de Rham cohomology $H^k(\mathbb{R}^n)$ for $k \ge 1$. We will prove it in Section 27.

The theory of differential forms allows us to generalize vector calculus from $\mathbb{R}^3$ to $\mathbb{R}^n$ and indeed to a manifold of any dimension. The general Stokes' theorem for a manifold that we will prove in Subsection 23.5 subsumes and unifies the fundamental theorem for line integrals, Green's theorem in the plane, the classical Stokes' theorem for a surface in $\mathbb{R}^3$, and the divergence theorem. As a first step in this program, we begin the next chapter with the definition of a manifold.

4.7 Convention on Subscripts and Superscripts

In differential geometry it is customary to index vector fields with subscripts $e_1, \dots, e_n$, and differential forms with superscripts $\omega^1, \dots, \omega^n$. Being 0-forms, coordinate functions take superscripts: $x^1, \dots, x^n$. Their differentials, being 1-forms, should also have superscripts, and indeed they do: $dx^1, \dots, dx^n$. Coordinate vector fields $\partial/\partial x^1, \dots, \partial/\partial x^n$ are considered to have subscripts because the i in $\partial/\partial x^i$, although a superscript for x^i, is in the lower half of the fraction.

Coefficient functions can have superscripts or subscripts depending on whether they are the coefficient functions of a vector field or of a differential form. For a vector field $X = \sum a^i e_i$, the coefficient functions a^i have superscripts; the idea is that the superscript in a^i "cancels out" the subscript in e_i. For the same reason, the coefficient functions b_j in a differential form $\omega = \sum b_j \, dx^j$ have subscripts.

The beauty of this convention is that there is a "conservation of indices" on the two sides of an equality sign. For example, if $X = \sum a^i \, \partial/\partial x^i$, then

$$a^i = (dx^i)(X).$$

Here both sides have a net superscript i. As another example, if $\omega = \sum b_j \, dx^j$, then

$$\omega(X) = \left(\sum b_j \, dx^j\right) \left(\sum a^i \frac{\partial}{\partial x^i}\right) = \sum b_i a^i;$$

after cancellation of superscripts and subscripts, both sides of the equality sign have zero net index. This convention is a useful mnemonic aid in some of the transformation formulas of differential geometry.

Problems

4.1. A 1-form on $\mathbb{R}^3$
Let ω be the 1-form $z\,dx - dz$ and let X be the vector field $y\,\partial/\partial x + x\,\partial/\partial y$ on $\mathbb{R}^3$. Compute $\omega(X)$ and $d\omega$.

4.2. A 2-form on $\mathbb{R}^3$
At each point $p \in \mathbb{R}^3$, define a bilinear function ω_p on $T_p(\mathbb{R}^3)$ by

$$\omega_p(\mathbf{a}, \mathbf{b}) = \omega_p\left(\begin{bmatrix} a^1 \\ a^2 \\ a^3 \end{bmatrix}, \begin{bmatrix} b^1 \\ b^2 \\ b^3 \end{bmatrix}\right) = p^3 \det \begin{bmatrix} a^1 & b^1 \\ a^2 & b^2 \end{bmatrix},$$

for tangent vectors $\mathbf{a}, \mathbf{b} \in T_p(\mathbb{R}^3)$, where p^3 is the third component of $p = (p^1, p^2, p^3)$. Since ω_p is an alternating bilinear function on $T_p(\mathbb{R}^3)$, ω is a 2-form on $\mathbb{R}^3$. Write ω in terms of the standard basis $dx^i \wedge dx^j$ at each point.

4.3. Exterior calculus
Suppose the standard coordinates on $\mathbb{R}^2$ are called r and θ (this $\mathbb{R}^2$ is the (r, θ)-plane, not the (x, y)-plane). If $x = r\cos\theta$ and $y = r\sin\theta$, calculate dx, dy, and $dx \wedge dy$ in terms of dr and $d\theta$.

4.4. Exterior calculus
Suppose the standard coordinates on $\mathbb{R}^3$ are called ρ, ϕ, and θ. If $x = \rho\sin\phi\cos\theta$, $y = \rho\sin\phi\sin\theta$, and $z = \rho\cos\phi$, calculate dx, dy, dz, and $dx \wedge dy \wedge dz$ in terms of $d\rho$, $d\phi$, and $d\theta$.

4.5. Wedge product
Let α be a 1-form and β a 2-form on $\mathbb{R}^3$. Then

$$\alpha = a_1 \, dx^1 + a_2 \, dx^2 + a_3 \, dx^3,$$
$$\beta = b_1 \, dx^2 \wedge dx^3 + b_2 \, dx^3 \wedge dx^1 + b_3 \, dx^1 \wedge dx^2.$$

Simplify the expression $\alpha \wedge \beta$ as much as possible.

4.6. Wedge product and cross product
The correspondence between differential forms and vector fields on an open subset of $\mathbb{R}^3$ in Subsection 4.6 also makes sense pointwise. Let V be a vector space of dimension 3 with basis e_1, e_2, e_3, and dual basis $\alpha^1, \alpha^2, \alpha^3$. To a 1-covector $\alpha = a_1\,\alpha^1 + a_2\,\alpha^2 + a_3\,\alpha^3$ on V, we associate the vector $\mathbf{v}_\alpha = \langle a_1, a_2, a_3\rangle \in \mathbb{R}^3$. To the 2-covector

$$\gamma = c_1\,\alpha^2 \wedge \alpha^3 + c_2\,\alpha^3 \wedge \alpha^1 + c_3\,\alpha^1 \wedge \alpha^2$$

on V, we associate the vector $\mathbf{v}_\gamma = \langle c_1, c_2, c_3\rangle \in \mathbb{R}^3$. Show that under this correspondence, the wedge product of 1-covectors corresponds to the cross product of vectors in $\mathbb{R}^3$: if $\alpha = a_1\,\alpha^1 + a_2\,\alpha^2 + a_3\,\alpha^3$ and $\beta = b_1\,\alpha^1 + b_2\,\alpha^2 + b_3\,\alpha^3$, then $\mathbf{v}_{\alpha\wedge\beta} = \mathbf{v}_\alpha \times \mathbf{v}_\beta$.

4.7. Commutator of derivations and antiderivations
Let $A = \oplus_{k=-\infty}^{\infty} A^k$ be a graded algebra over a field K with $A^k = 0$ for $k < 0$. Let m be an integer. A *superderivation of A of degree m* is a K-linear map $D\colon A \to A$ such that for all k, $D(A^k) \subset A^{k+m}$ and for all $a \in A^k$ and $b \in A^\ell$,

$$D(ab) = (Da)b + (-1)^{km} a(Db).$$

If D_1 and D_2 are two superderivations of A of respective degrees m_1 and m_2, define their *commutator* to be

$$[D_1, D_2] = D_1 \circ D_2 - (-1)^{m_1 m_2} D_2 \circ D_1.$$

Show that $[D_1, D_2]$ is a superderivation of degree $m_1 + m_2$. (A superderivation is said to be *even* or *odd* depending on the parity of its degree. An even superderivation is a derivation; an odd superderivation is an antiderivation.)

Solutions to Selected Exercises Within the Text

3.6 Inversions
As a matrix, $\tau = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{bmatrix}$. Scanning the second row, we see that τ has four inversions: $(2,1)$, $(3,1)$, $(4,1)$, $(5,1)$. ◊

3.13 Symmetrizing operator
A k-linear function $h\colon V \to \mathbb{R}$ is symmetric if and only if $\tau h = h$ for all $\tau \in S_k$. Now

$$\tau(Sf) = \tau \sum_{\sigma \in S_k} \sigma f = \sum_{\sigma \in S_k} (\tau\sigma) f.$$

As σ runs over all elements of the permutation groups S_k, so does $\tau\sigma$. Hence,

$$\sum_{\sigma \in S_k} (\tau\sigma) f = \sum_{\tau\sigma \in S_k} (\tau\sigma) f = Sf.$$

This proves that $\tau(Sf) = Sf$. ◊

3.15 Alternating operator
$f(v_1,v_2,v_3) - f(v_1,v_3,v_2) + f(v_2,v_3,v_1) - f(v_2,v_1,v_3) + f(v_3,v_1,v_2) - f(v_3,v_2,v_1)$. ◊

3.20 Wedge product of two 2-covectors

$$\begin{aligned}
&(f \wedge g)(v_1,v_2,v_3,v_4) \\
&\quad = f(v_1,v_2)g(v_3,v_4) - f(v_1,v_3)g(v_2,v_4) + f(v_1,v_4)g(v_2,v_3) \\
&\qquad + f(v_2,v_3)g(v_1,v_4) - f(v_2,v_4)g(v_1,v_3) + f(v_3,v_4)g(v_1,v_2).
\end{aligned}$$ ◊

3.22 Sign of a permutation
We can achieve the permutation τ from the initial configuration $1,2,\ldots,k+\ell$ in k steps.

(1) First, move the element k to the very end across the ℓ elements $k+1,\ldots,k+\ell$. This requires ℓ transpositions.
(2) Next, move the element $k-1$ across the ℓ elements $k+1,\ldots,k+\ell$.
(3) Then move the element $k-2$ across the same ℓ elements, and so on.

Each of the k steps requires ℓ transpositions. In the end we achieve τ from the identity using ℓk transpositions.

Alternatively, one can count the number of inversions in the permutation τ. There are k inversions starting with $k+1$, namely, $(k+1,1),\dots,(k+1,k)$. Indeed, for each $i=1,\dots,\ell$, there are k inversions starting with $k+i$. Hence, the total number of inversions in τ is $k\ell$. By Proposition 3.8, $\operatorname{sgn}(\tau) = (-1)^{k\ell}$. $\diamondsuit$

4.3 A basis for 3-covectors

By Proposition 3.29, a basis for $A_3(T_p(\mathbb{R}^4))$ is $\left(dx^1\wedge dx^2\wedge dx^3\right)_p$, $\left(dx^1\wedge dx^2\wedge dx^4\right)_p$, $\left(dx^1\wedge dx^3\wedge dx^4\right)_p$, $\left(dx^2\wedge dx^3\wedge dx^4\right)_p$. $\diamondsuit$

4.4 Wedge product of a 2-form with a 1-form

The $(2,1)$-shuffles are $(1<2,3)$, $(1<3,2)$, $(2<3,1)$, with respective signs $+$, $-$, $+$. By Equation (3.6),

$$(\omega\wedge\tau)(X,Y,Z) = \omega(X,Y)\tau(Z) - \omega(X,Z)\tau(Y) + \omega(Y,Z)\tau(X). \qquad \diamondsuit$$

6.14 Smoothness of a map to a circle

Without further justification, the fact that both $\cos t$ and $\sin t$ are C^∞ proves only the smoothness of $(\cos t, \sin t)$ as a map from $\mathbb{R}$ to $\mathbb{R}^2$. To show that $F\colon \mathbb{R}\to S^1$ is C^∞, we need to cover S^1 with charts (U_i,ϕ_i) and examine in turn each $\phi_i\circ F\colon F^{-1}(U_i)\to\mathbb{R}$. Let $\{(U_i,\phi_i)\mid i=1,\dots,4\}$ be the atlas of Example 5.16. On $F^{-1}(U_1)$, $\phi_1\circ F(t) = (x\circ F)(t) = \cos t$ is C^∞. On $F^{-1}(U_3)$, $\phi_3\circ F(t) = \sin t$ is C^∞. Similar computations on $F^{-1}(U_2)$ and $F^{-1}(U_4)$ prove the smoothness of F. $\diamondsuit$

6.18 Smoothness of a map to a Cartesian product

Fix $p\in N$, let (U,ϕ) be a chart about p, and let $(V_1\times V_2, \psi_1\times\psi_2)$ be a chart about $(f_1(p), f_2(p))$. We will be assuming either (f_1,f_2) smooth or both f_i smooth. In either case, (f_1,f_2) is continuous. Hence, by choosing U sufficiently small, we may assume $(f_1,f_2)(U)\subset V_1\times V_2$. Then

$$(\psi_1\times\psi_2)\circ(f_1,f_2)\circ\phi^{-1} = (\psi_1\circ f_1\circ\phi^{-1}, \psi_2\circ f_2\circ\phi^{-1})$$

maps an open subset of $\mathbb{R}^n$ to an open subset of $\mathbb{R}^{m_1+m_2}$. It follows that (f_1,f_2) is C^∞ at p if and only if both f_1 and f_2 are C^∞ at p. $\diamondsuit$

7.11 Real projective space as a quotient of a sphere

Define $\bar{f}\colon \mathbb{R}P^n\to S^n/{\sim}$ by $\bar{f}([x]) = [\frac{x}{\|x\|}]\in S^n/{\sim}$. This map is well defined because $\bar{f}([tx]) = [\frac{tx}{|tx|}] = [\pm\frac{x}{\|x\|}] = [\frac{x}{\|x\|}]$. Note that if $\pi_1\colon \mathbb{R}^{n+1}-\{0\}\to\mathbb{R}P^n$ and $\pi_2\colon S^n\to S^n/{\sim}$ are the projection maps, then there is a commutative diagram

$$\begin{array}{ccc}
\mathbb{R}^n-\{0\} & \xrightarrow{\;f\;} & S^n \\
{\scriptstyle\pi_1}\downarrow & & \downarrow{\scriptstyle\pi_2} \\
\mathbb{R}P^n & \xrightarrow[\;\bar{f}\;]{} & S^n/{\sim}\,.
\end{array}$$

By Proposition 7.1, $\bar{f}$ is continuous because $\pi_2\circ f$ is continuous.

Next define $g\colon S^n\to\mathbb{R}^{n+1}-\{0\}$ by $g(x)=x$. This map induces a map $\bar{g}\colon S^n/{\sim}\to\mathbb{R}P^n$, $\bar{g}([x]) = [x]$. By the same argument as above, $\bar{g}$ is well defined and continuous. Moreover,

$$\bar{g}\circ\bar{f}([x]) = \left[\frac{x}{\|x\|}\right] = [x],$$
$$\bar{f}\circ\bar{g}([x]) = [x],$$

Hints and Solutions to Selected End-of-Section Problems

Problems with complete solutions are starred (*). Equations are numbered consecutively within each problem.

1.2* A C^∞ function very flat at 0
(a) Assume $x > 0$. For $k = 1$, $f'(x) = (1/x^2)e^{-1/x}$. With $p_2(y) = y^2$, this verifies the claim. Now suppose $f^{(k)}(x) = p_{2k}(1/x)e^{-1/x}$. By the product rule and the chain rule,

$$\begin{aligned} f^{(k+1)}(x) &= p_{2k-1}\left(\frac{1}{x}\right)\cdot\left(-\frac{1}{x^2}\right)e^{-\frac{1}{x}} + p_{2k}\left(\frac{1}{x}\right)\cdot\frac{1}{x^2}e^{-\frac{1}{x}} \\ &= \left(q_{2k+1}\left(\frac{1}{x}\right) + q_{2k+2}\left(\frac{1}{x}\right)\right)e^{-\frac{1}{x}} \\ &= p_{2k+2}\left(\frac{1}{x}\right)e^{-\frac{1}{x}}, \end{aligned}$$

where $q_n(y)$ and $p_n(y)$ are polynomials of degree n in y. By induction, the claim is true for all $k \geq 1$. It is trivially true for $k = 0$ also.

(b) For $x > 0$, the formula in (a) shows that $f(x)$ is C^∞. For $x < 0$, $f(x) \equiv 0$, which is trivially C^∞. It remains to show that $f^{(k)}(x)$ is defined and continuous at $x = 0$ for all k.

Suppose $f^{(k)}(0) = 0$. By the definition of the derivative,

$$f^{(k+1)}(0) = \lim_{x\to 0}\frac{f^{(k)}(x) - f^{(k)}(0)}{x} = \lim_{x\to 0}\frac{f^{(k)}(x)}{x}.$$

The limit from the left is clearly 0. So it suffices to compute the limit from the right:

$$\begin{aligned} \lim_{x\to 0^+}\frac{f^{(k)}(x)}{x} &= \lim_{x\to 0^+}\frac{p_{2k}\left(\frac{1}{x}\right)e^{-\frac{1}{x}}}{x} = \lim_{x\to 0^+} p_{2k+1}\left(\frac{1}{x}\right)e^{-\frac{1}{x}} \qquad (1.2.1) \\ &= \lim_{y\to\infty}\frac{p_{2k+1}(y)}{e^y} \quad \left(\text{replacing } \frac{1}{x} \text{ by } y\right). \end{aligned}$$

Applying l'Hôpital's rule $2k+1$ times, we reduce this limit to 0. Hence, $f^{(k+1)}(0) = 0$. By induction, $f^{(k)}(0) = 0$ for all $k \geq 0$.

A similar computation as (1.2.1) for $\lim_{x\to 0} f^{(k)}(x) = 0$ proves that $f^{(k)}(x)$ is continuous at $x = 0$. $\Diamond$

1.3 (b) $h(t) = (\pi/(b-a))(t-a) - (\pi/2)$.

1.5

(a) The line passing through $(0,0,1)$ and (a,b,c) has a parametrization

$$x = at, \quad y = bt, \quad z = (c-1)t + 1.$$

This line intersects the xy-plane when

$$z = 0 \Leftrightarrow t = \frac{1}{1-c} \Leftrightarrow (x,y) = \left(\frac{a}{1-c}, \frac{b}{1-c}\right).$$

To find the inverse of g, write down a parametrization of the line through $(u,v,0)$ and $(0,0,1)$ and solve for the intersection of this line with S.

1.6* Taylor's theorem with remainder to order 2

To simplify the notation, we write $\mathbf{0}$ for $(0,0)$. By Taylor's theorem with remainder, there exist C^∞ functions g_1, g_2 such that

$$f(x,y) = f(\mathbf{0}) + xg_1(x,y) + yg_2(x,y). \tag{1.6.1}$$

Applying the theorem again, but to g_1 and g_2, we obtain

$$g_1(x,y) = g_1(\mathbf{0}) + xg_{11}(x,y) + yg_{12}(x,y), \tag{1.6.2}$$
$$g_2(x,y) = g_2(\mathbf{0}) + xg_{21}(x,y) + yg_{22}(x,y). \tag{1.6.3}$$

Since $g_1(\mathbf{0}) = \partial f/\partial x(\mathbf{0})$ and $g_2(\mathbf{0}) = \partial f/\partial y(\mathbf{0})$, substituting (1.6.2) and (1.6.3) into (1.6.1) gives the result. ◊

1.7* A function with a removable singularity

In Problem 1.6, set $x = t$ and $y = tu$. We obtain

$$f(t,tu) = f(\mathbf{0}) + t\frac{\partial f}{\partial x}(\mathbf{0}) + tu\frac{\partial f}{\partial y}(\mathbf{0}) + t^2(\cdots),$$

where

$$(\cdots) = g_{11}(t,tu) + ug_{12}(t,tu) + u^2 g_{22}(t,tu)$$

is a C^∞ function of t and u. Since $f(\mathbf{0}) = \partial f/\partial x(\mathbf{0}) = \partial f/\partial y(\mathbf{0}) = 0$,

$$\frac{f(t,tu)}{t} = t(\cdots),$$

which is clearly C^∞ in t,u and agrees with g when $t = 0$. ◊

1.8 See Example 1.2(ii).

3.1 $f = \sum g_{ij}\,\alpha^i \otimes \alpha^j$.

3.2

(a) Use the formula $\dim\ker f + \dim\operatorname{im} f = \dim V$.

(b) Choose a basis $e_1, \ldots, e_{n-1}$ for $\ker f$, and extend it to a basis $e_1, \ldots, e_{n-1}, e_n$ for V. Let $\alpha^1, \ldots, \alpha^n$ be the dual basis for $V^\vee$. Write both f and g in terms of this dual basis.

3.3 We write temporarily α^I for $\alpha^{i_1} \otimes \cdots \otimes \alpha^{i_k}$ and e_J for $(e_{j_1}, \ldots, e_{j_k})$.

(a) Prove that $f = \sum f(e_I)\alpha^I$ by showing that both sides agree on all (e_J). This proves that the set $\{\alpha^I\}$ spans.
(b) Suppose $\sum c_I \alpha^I = 0$. Applying both sides to e_J gives $c_J = \sum c_I \alpha^I(e_J) = 0$. This proves that the set $\{\alpha^I\}$ is linearly independent.

3.9 To compute $\omega(v_1, \dots, v_n)$ for any $v_1, \dots, v_n \in V$, write $v_j = \sum_i e_i a^i_j$ and use the fact that ω is multilinear and alternating.

3.10* Linear independence of covectors
($\Rightarrow$) If $\alpha^1, \dots, \alpha^k$ are linearly dependent, then one of them is a linear combination of the others. Without loss of generality, we may assume that

$$\alpha^k = \sum_{i=1}^{k-1} c_i \alpha^i.$$

In the wedge product $\alpha^1 \wedge \cdots \wedge \alpha^{k-1} \wedge (\sum_{i=1}^{k-1} c_i \alpha^i)$, every term has a repeated α^i. Hence, $\alpha^1 \wedge \cdots \wedge \alpha^k = 0$.

($\Leftarrow$) Suppose $\alpha^1, \dots, \alpha^k$ are linearly independent. Then they can be extended to a basis $\alpha^1, \dots, \alpha^k, \dots, \alpha^n$ for $V^\vee$. Let $v_1, \dots, v_n$ be the dual basis for V. By Proposition 3.27,

$$(\alpha^1 \wedge \cdots \wedge \alpha^k)(v_1, \dots, v_k) = \det[\alpha^i(v_j)] = \det[\delta^i_j] = 1.$$

Hence, $\alpha^1 \wedge \cdots \wedge \alpha^k \neq 0$. ◊

3.11* Exterior multiplication
($\Leftarrow$) Clear because $\alpha \wedge \alpha = 0$.

($\Rightarrow$) Suppose $\alpha \wedge \omega = 0$. Extend α to a basis $\alpha^1, \dots, \alpha^n$ for $V^\vee$, with $\alpha^1 = \alpha$. Write $\omega = \sum c_J \alpha^J$. In the sum $\alpha \wedge \omega = \sum c_J \alpha \wedge \alpha^J$, all the terms $\alpha \wedge \alpha^J$ with $j_1 = 1$ vanish since $\alpha = \alpha^1$. Hence,

$$0 = \alpha \wedge \omega = \sum_{j_1 \neq 1} c_J \alpha \wedge \alpha^J.$$

Since $\{\alpha \wedge \alpha^J\}_{j_1 \neq 1}$ is a subset of a basis for $A_{k+1}(V)$, it is linearly independent and so all $c_J = 0$ if $j_1 \neq 1$. Thus,

$$\omega = \sum_{j_1 = 1} c_J \alpha^J = \alpha \wedge \left(\sum_{j_1 = 1} c_J \alpha^{j_2} \wedge \cdots \wedge \alpha^{j_k} \right).$$ ◊

4.1 $\omega(X) = yz$, $d\omega = -dx \wedge dz$.

4.2 Write $\omega = \sum_{i<j} c_{ij}\, dx^i \wedge dx^j$. Then $c_{ij}(p) = \omega_p(e_i, e_j)$, where $e_i = \partial/\partial x^i$. Calculate $c_{12}(p)$, $c_{13}(p)$, and $c_{23}(p)$. The answer is $\omega_p = p^3\, dx^1 \wedge dx^2$.

4.3 $dx = \cos\theta\, dr - r\sin\theta\, d\theta$, $dy = \sin\theta, dr + r\cos\theta\, d\theta$, $dx \wedge dy = r\, dr \wedge d\theta$.

4.4 $dx \wedge dy \wedge dz = \rho^2 \sin\phi\, d\rho \wedge d\phi \wedge d\theta$.

4.5 $\alpha \wedge \beta = (a_1 b_1 + a_2 b_2 + a_3 b_3)\, dx^1 \wedge dx^2 \wedge dx^3$.

5.3 The image $\phi_4(U_{14}) = \{(x,z) \mid -1 < z < 1,\ 0 < x < \sqrt{1-z^2}\}$.
The transition function $\phi_1 \circ \phi_4^{-1}(x,z) = \phi_1(x,y,z) = (y,z) = (-\sqrt{1-x^2-z^2}, z)$ is a C^∞ function of x, z.

Loring W. Tu

An Introduction to Manifolds

Second Edition

May 19, 2010

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Loring W. Tu
Department of Mathematics
Tufts University
Medford, MA 02155
loring.tu@tufts.edu

Appendices

Appendices

A

Point-Set Topology

Point-set topology, also called "general topology," is concerned with properties that remain invariant under homeomorphisms (continuous maps having continuous inverses). The basic development in the subject took place in the late nineteenth and early twentieth centuries. This appendix is a collection of basic results from point-set topology that are used throughout the book.

A.1 Topological Spaces

The prototype of a topological space is the Euclidean space $\mathbb{R}^n$. However, Euclidean space comes with many additional structures, such as a metric, coordinates, an inner product, and an orientation, that are extraneous to its topology. The idea behind the definition of a topological space is to discard all those properties of $\mathbb{R}^n$ that have nothing to do with continuous maps, thereby distilling the notion of continuity to its very essence.

In advanced calculus one learns several characterizations of a continuous map, among which is the following: a map f from an open subset of $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous if and only if the inverse image $f^{-1}(V)$ of any open set V in $\mathbb{R}^m$ is open in $\mathbb{R}^n$. This shows that continuity can be defined solely in terms of open sets.

To define open sets axiomatically, we look at properties of open sets in $\mathbb{R}^n$. Recall that in $\mathbb{R}^n$ the *distance* between two points p and q is given by

$$d(p,q) = \left[\sum_{i=1}^{n} (p^i - q^i)^2\right]^{1/2},$$

and the *open ball* $B(p,r)$ with center $p \in \mathbb{R}^n$ and radius $r > 0$ is the set

$$B(p,r) = \{x \in \mathbb{R}^n \mid d(x,p) < r\}.$$

A set U in $\mathbb{R}^n$ is said to be *open* if for every p in U, there is an open ball $B(p,r)$ with center p and radius r such that $B(p,r) \subset U$ (Figure A.1). It is clear that the union of

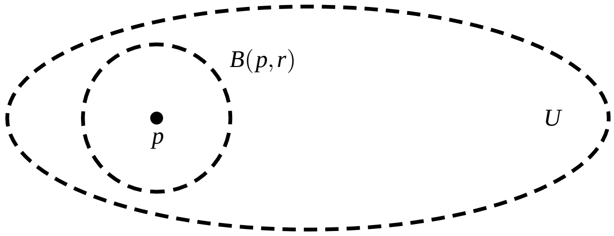


Fig. A.1. An open set in $\mathbb{R}^n$.

an arbitrary collection $\{U_\alpha\}$ of open sets is open, but the same need not be true of the intersection of infinitely many open sets.

Example. The intervals $]-1/n, 1/n[$, $n = 1,2,3,\ldots$, are all open in $\mathbb{R}^1$, but their intersection $\bigcap_{n=1}^{\infty}]-1/n, 1/n[$ is the singleton set $\{0\}$, which is not open.

What is true is that the intersection of a *finite* collection of open sets in $\mathbb{R}^n$ is open. This leads to the definition of a topology on a set.

Definition A.1. A *topology* on a set S is a collection $\mathcal{T}$ of subsets containing both the empty set $\varnothing$ and the set S such that $\mathcal{T}$ is closed under arbitrary unions and finite intersections; i.e., if $U_\alpha \in \mathcal{T}$ for all α in an index set A, then $\bigcup_{\alpha\in \mathrm{A}} U_\alpha \in \mathcal{T}$ and if $U_1,\ldots,U_n \in \mathcal{T}$, then $\bigcap_{i=1}^{n} U_i \in \mathcal{T}$.

The elements of $\mathcal{T}$ are called *open sets* and the pair $(S,\mathcal{T})$ is called a *topological space*. To simplify the notation, we sometimes simply refer to a pair $(S,\mathcal{T})$ as "the topological space S" when there is no chance of confusion. A *neighborhood* of a point p in S is an open set U containing p. If $\mathcal{T}_1$ and $\mathcal{T}_2$ are two topologies on a set S and $\mathcal{T}_1 \subset \mathcal{T}_2$, then we say that $\mathcal{T}_1$ is *coarser* than $\mathcal{T}_1$, or that $\mathcal{T}_2$ is *finer* than $\mathcal{T}_1$. A coarser topology has fewer open sets; conversely, a finer topology has more open sets.

Example. The open subsets of $\mathbb{R}^n$ as we understand them in advanced calculus form a topology on $\mathbb{R}^n$, the *standard topology* of $\mathbb{R}^n$. In this topology a set U is open in $\mathbb{R}^n$ if and only if for every $p \in U$, there is an open ball $B(p,\varepsilon)$ with center p and radius ε contained in U. Unless stated otherwise, $\mathbb{R}^n$ will always have its standard topology.

The criterion for openness in $\mathbb{R}^n$ has a useful generalization to a topological space.

Lemma A.2 (Local criterion for openness). *Let S be a topological space. A subset A is open in S if and only if for every $p \in A$, there is an open set V such that $p \in V \subset A$.*

Proof.
($\Rightarrow$) If A is open, we can take $V = A$.
($\Leftarrow$) Suppose for every $p \in A$ there is an open set V_p such that $p \in V_p \subset A$. Then

$$A \subset \bigcup_{p \in A} V_p \subset A,$$

so that equality $A = \bigcup_{p \in A} V_p$ holds. As a union of open sets, A is open. □

Example. For any set S, the collection $\mathcal{T} = \{\varnothing, S\}$ consisting of the empty set $\varnothing$ and the entire set S is a topology on S, sometimes called the *trivial* or *indiscrete topology*. It is the coarsest topology on a set.

Example. For any set S, let $\mathcal{T}$ be the collection of all subsets of S. Then $\mathcal{T}$ is a topology on S, called the *discrete topology*. A *singleton set* is a set with a single element. The discrete topology can also be characterized as the topology in which every singleton subset $\{p\}$ is open. A topological space having the discrete topology is called a *discrete space*. The discrete topology is the finest topology on a set.

The complement of an open set is called a *closed set*. By de Morgan's laws from set theory, arbitrary intersections and finite unions of closed sets are closed (Problem A.3). One may also specify a topology by describing all the closed sets.

Remark. When we say that a topology is *closed* under arbitrary union and finite intersection, the word "closed" has a different meaning from that of a "closed subset."

Example A.3 (*Finite-complement topology on* $\mathbb{R}^1$). Let $\mathcal{T}$ be the collection of subsets of $\mathbb{R}^1$ consisting of the empty set $\varnothing$, the line $\mathbb{R}^1$ itself, and the complements of finite sets. Suppose F_α and F_i are finite subsets of $\mathbb{R}^1$ for $\alpha \in$ some index set A and $i = 1, \dots, n$. By de Morgan's laws,

$$\bigcup_{\alpha} (\mathbb{R}^1 - F_\alpha) = \mathbb{R}^1 - \bigcap_{\alpha} F_\alpha \quad \text{and} \quad \bigcap_{i=1}^{n} (\mathbb{R}^1 - F_i) = \mathbb{R}^1 - \bigcup_{i=1}^{n} F_i.$$

Since the arbitrary intersection $\bigcap_{\alpha \in \mathrm{A}} F_\alpha$ and the finite union $\bigcup_{i=1}^{n} F_i$ are both finite, $\mathcal{T}$ is closed under arbitrary unions and finite intersections. Thus, $\mathcal{T}$ defines a topology on $\mathbb{R}^1$, called the *finite-complement topology*.

For the sake of definiteness, we have defined the finite-complement topology on $\mathbb{R}^1$, but of course, there is nothing specific about $\mathbb{R}^1$ here. One can define in exactly the same way the finite-complement topology on any set.

Example A.4 (*Zariski topology*). One well-known topology is the *Zariski topology* from algebraic geometry. Let K be a field and let S be the vector space K^n. Define a subset of K^n to be *Zariski closed* if it is the zero set $Z(f_1, \dots, f_r)$ of finitely many polynomials $f_1, \dots, f_r$ on K^n. To show that these are indeed the closed subsets of a topology, we need to check that they are closed under arbitrary intersections and finite unions.

Let $I = (f_1, \dots, f_r)$ be the ideal generated by $f_1, \dots, f_r$ in the polynomial ring $K[x_1, \dots, x_n]$. Then $Z(f_1, \dots, f_r) = Z(I)$, the zero set of *all* the polynomials in the ideal I. Conversely, by the Hilbert basis theorem [11, §9.6, Th. 21], any ideal in

$K[x_1,\dots,x_n]$ has a finite set of generators. Hence, the zero set of finitely many polynomials is the same as the zero set of an ideal in $K[x_1,\dots,x_n]$. If $I=(f_1,\dots,f_r)$ and $J=(g_1,\dots,g_s)$ are two ideals, then the *product ideal* IJ is the ideal in $K[x_1,\dots,x_n]$ generated by all products f_ig_j, $1\le i\le r$, $1\le j\le s$. If $\{I_\alpha\}_{\alpha\in A}$ is a family of ideals in $K[x_1,\dots,x_n]$, then their *sum* $\sum_\alpha I_\alpha$ is the smallest ideal in $K[x_1,\dots,x_n]$ containing all the ideals I_α.

Exercise A.5 (Intersection and union of zero sets). Let I_α, I, and J be ideals in the polynomial ring $K[x_1,\dots,x_n]$. Show that

(i) $$\bigcap_\alpha Z(I_\alpha)=Z\left(\sum_\alpha I_\alpha\right)$$

and

(ii) $$Z(I)\cup Z(J)=Z(IJ).$$

The complement of a Zariski closed subset of K^n is said to be *Zariski open*. If $I=(0)$ is the zero ideal, then $Z(I)=K^n$, and if $I=(1)=K[x_1,\dots,x_n]$ is the entire ring, then $Z(I)$ is the empty set $\varnothing$. Hence, both the empty set and K^n are Zariski open. It now follows from Exercise A.5 that the Zariski open subsets of K^n form a topology on K^n, called the *Zariski topology* on K^n. Since the zero set of a polynomial on $\mathbb{R}^1$ is a finite set, the Zariski topology on $\mathbb{R}^1$ is precisely the finite-complement topology of Example A.3.

A.2 Subspace Topology

Let $(S,\mathcal{T})$ be a topological space and A a subset of S. Define $\mathcal{T}_A$ to be the collection of subsets

$$\mathcal{T}_A=\{U\cap A \mid U\in\mathcal{T}\}.$$

By the distributive property of union and intersection,

$$\bigcup_\alpha (U_\alpha\cap A)=\left(\bigcup_\alpha U_\alpha\right)\cap A$$

and

$$\bigcap_i (U_i\cap A)=\left(\bigcap_i U_i\right)\cap A,$$

which shows that $\mathcal{T}_A$ is closed under arbitrary unions and finite intersections. Moreover, $\varnothing, A\in\mathcal{T}_A$. So $\mathcal{T}_A$ is a topology on A, called the *subspace topology* or the *relative topology* of A in S, and elements of $\mathcal{T}_A$ are said to be *open in* A. To emphasize the fact that an open set U in A need not be open in S, we also say that U is *open relative to* A or *relatively open in* A. The subset A of S with the subspace topology $\mathcal{T}_A$ is called a *subspace* of S.

Example. Consider the subset $A = [0,1]$ of $\mathbb{R}^1$. In the subspace topology, the half-open interval $[0,1/2[$ is open relative to A, because

$$\left[0,\tfrac{1}{2}\right[\;=\; \left]-\tfrac{1}{2},\tfrac{1}{2}\right[\,\cap A.$$

(See Figure A.2.)

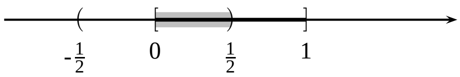


Fig. A.2. A relatively open subset of $[0,1]$.

A.3 Bases

It is generally difficult to describe directly all the open sets in a topology $\mathcal{T}$. What one can usually do is to describe a subcollection $\mathcal{B}$ of $\mathcal{T}$ so that any open set is expressible as a union of open sets in $\mathcal{B}$.

Definition A.6. A subcollection $\mathcal{B}$ of a topology $\mathcal{T}$ on a topological space S is a *basis for the topology* $\mathcal{T}$ if given an open set U and point p in U, there is an open set $B \in \mathcal{B}$ such that $p \in B \subset U$. We also say $\mathcal{B}$ *generates* the topology $\mathcal{T}$ or that $\mathcal{B}$ is a *basis for the topological space S*.

Example. The collection of all open balls $B(p,r)$ in $\mathbb{R}^n$, with $p \in \mathbb{R}^n$ and r a positive real number, is a basis for the standard topology of $\mathbb{R}^n$.

Proposition A.7. *A collection $\mathcal{B}$ of open sets of S is a basis if and only if every open set in S is a union of sets in $\mathcal{B}$.*

Proof.
($\Rightarrow$) Suppose $\mathcal{B}$ is a basis and U is an open set in S. For every $p \in U$, there is a basic open set $B_p \in \mathcal{B}$ such that $p \in B_p \subset U$. Therefore, $U = \bigcup_{p \in U} B_p$.

($\Leftarrow$) Suppose every open set in S is a union of open sets in $\mathcal{B}$. Given an open set U and a point p in U, since $U = \bigcup_{B_\alpha \in \mathcal{B}} B_\alpha$, there is a $B_\alpha \in \mathcal{B}$ such that $p \in B_\alpha \subset U$. Hence, $\mathcal{B}$ is a basis. □

The following proposition gives a useful criterion for deciding if a collection $\mathcal{B}$ of subsets is a basis for some topology.

Proposition A.8. *A collection $\mathcal{B}$ of subsets of a set S is a basis for some topology $\mathcal{T}$ on S if and only if*

(i) *S is the union of all the sets in $\mathcal{B}$, and*
(ii) *given any two sets B_1 and $B_2 \in \mathcal{B}$ and a point $p \in B_1 \cap B_2$, there is a set $B \in \mathcal{B}$ such that $p \in B \subset B_1 \cap B_2$ (Figure A.3).*

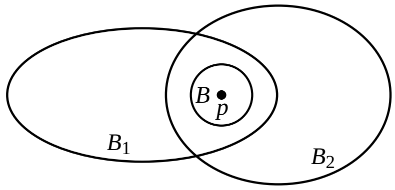


Fig. A.3. Criterion for a basis.

Proof.
($\Rightarrow$) (i) follows from Proposition A.7.

(ii) If $\mathcal{B}$ is a basis, then B_1 and B_2 are open sets and hence so is $B_1 \cap B_2$. By the definition of a basis, there is a $B \in \mathcal{B}$ such that $p \in B \subset B_1 \cap B_2$.

($\Leftarrow$) Define $\mathcal{T}$ to be the collection consisting of all sets that are unions of sets in $\mathcal{B}$. Then the empty set $\varnothing$ and the set S are in $\mathcal{T}$ and $\mathcal{T}$ is clearly closed under arbitrary union. To show that $\mathcal{T}$ is closed under finite intersection, let $U = \bigcup_\mu B_\mu$ and $V = \bigcup_\nu B_\nu$ be in $\mathcal{T}$, where $B_\mu, B_\nu \in \mathcal{B}$. Then

$$U \cap V = \left(\bigcup_\mu B_\mu\right) \cap \left(\bigcup_\nu B_\nu\right)$$
$$= \bigcup_{\mu,\nu} (B_\mu \cap B_\nu).$$

Thus, any p in $U \cap V$ is in $B_\mu \cap B_\nu$ for some μ, ν. By (ii) there is a set B_p in $\mathcal{B}$ such that $p \in B_p \subset B_\mu \cap B_\nu$. Therefore,

$$U \cap V = \bigcup_{p \in U \cap V} B_p \in \mathcal{T}.$$ □

Proposition A.9. *Let $\mathcal{B} = \{B_\alpha\}$ be a basis for a topological space S, and A a subspace of S. Then $\{B_\alpha \cap A\}$ is a basis for A.*

Proof. Let U' be any open set in A and $p \in U'$. By the definition of subspace topology, $U' = U \cap A$ for some open set U in S. Since $p \in U \cap A \subset U$, there is a basic open set B_α such that $p \in B_\alpha \subset U$. Then

$$p \in B_\alpha \cap A \subset U \cap A = U',$$

which proves that the collection $\{B_\alpha \cap A \mid B_\alpha \in \mathcal{B}\}$ is a basis for A. □

A.4 First and Second Countability

First and second countability of a topological space have to do with the countability of a basis. Before taking up these notions, we begin with an example. We say that a point in $\mathbb{R}^n$ is *rational* if all of its coordinates are rational numbers. Let $\mathbb{Q}$ be the set of rational numbers and $\mathbb{Q}^+$ the set of positive rational numbers. From real analysis, it is well known that every open interval in $\mathbb{R}$ contains a rational number.

Lemma A.10. *Every open set in $\mathbb{R}^n$ contains a rational point.*

Proof. An open set U in $\mathbb{R}^n$ contains an open ball $B(p,r)$, which in turn contains an open cube $\prod_{i=1}^n I_i$, where I_i is the open interval $]p^i - (r/\sqrt{n}), p^i + (r/\sqrt{n})[$ (see Problem A.4). For each i, let q^i be a rational number in I_i. Then $(q^1,\dots,q^n)$ is a rational point in $\prod_{i=1}^n I_i \subset B(p,r) \subset U$. □

Proposition A.11. *The collection $\mathcal{B}_{\text{rat}}$ of all open balls in $\mathbb{R}^n$ with rational centers and rational radii is a basis for $\mathbb{R}^n$.*

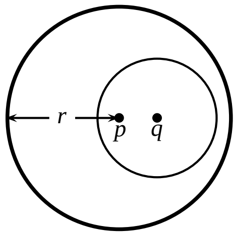


Fig. A.4. A ball with rational center q and rational radius $r/2$.

Proof. Given an open set U in $\mathbb{R}^n$ and point p in U, there is an open ball $B(p,r')$ with positive real radius r' such that $p \in B(p,r') \subset U$. Take a rational number r in $]0,r'[$. Then $p \in B(p,r) \subset U$. By Lemma A.10, there is a rational point q in the smaller ball $B(p,r/2)$. We claim that

$$p \in B\left(q, \frac{r}{2}\right) \subset B(p,r). \tag{A.1}$$

(See Figure A.4.) Since $d(p,q) < r/2$, we have $p \in B(q,r/2)$. Next, if $x \in B(q,r/2)$, then by the triangle inequality,

$$d(x,p) \le d(x,q) + d(q,p) < \frac{r}{2} + \frac{r}{2} = r.$$

So $x \in B(p,r)$. This proves the claim (A.1). Because $p \in B(q,r/2) \subset U$, the collection $\mathcal{B}_{\text{rat}}$ of open balls with rational centers and rational radii is a basis for $\mathbb{R}^n$. □

Both of the sets $\mathbb{Q}$ and $\mathbb{Q}^+$ are countable. Since the centers of the balls in $\mathcal{B}_{\text{rat}}$ are indexed by $\mathbb{Q}^n$, a countable set, and the radii are indexed by $\mathbb{Q}^+$, also a countable set, the collection $\mathcal{B}_{\text{rat}}$ is countable.

Definition A.12. A topological space is said to be *second countable* if it has a countable basis.

Example A.13. Proposition A.11 shows that $\mathbb{R}^n$ with its standard topology is second countable. With the discrete topology, $\mathbb{R}^n$ would not be second countable. More generally, any uncountable set with the discrete topology is not second countable.

Proposition A.14. *A subspace A of a second countable space S is second countable.*

Proof. By Proposition A.9, if $\mathcal{B} = \{B_i\}$ is a countable basis for S, then $\mathcal{B}_A := \{B_i \cap A\}$ is a countable basis for A. □

Definition A.15. Let S be a topological space and p a point in S. A *basis of neighborhoods at p* or a *neighborhood basis at p* is a collection $\mathcal{B} = \{B_\alpha\}$ of neighborhoods of p such that for any neighborhood U of p, there is a $B_\alpha \in \mathcal{B}$ such that $p \in B_\alpha \subset U$. A topological space S is *first countable* if it has a countable basis of neighborhoods at every point $p \in S$.

Example. For $p \in \mathbb{R}^n$, let $B(p, 1/n)$ be the open ball of center p and radius $1/n$ in $\mathbb{R}^n$. Then $\{B(p, 1/n)\}_{n=1}^\infty$ is a neighborhood basis at p. Thus, $\mathbb{R}^n$ is first countable.

Example. An uncountable discrete space is first countable but not second countable. Every second countable space is first countable (the proof is left to Problem A.18).

Suppose p is a point in a first countable topological space and $\{V_i\}_{i=1}^\infty$ is a countable neighborhood basis at p. By taking $U_i = V_1 \cap \cdots \cap V_i$, we obtain a countable descending sequence

$$U_1 \supset U_2 \supset U_3 \supset \cdots$$

that is also a neighborhood basis at p. Thus, in the definition of first countability, we may assume that at every point the countable neighborhood basis at the point is a descending sequence of open sets.

A.5 Separation Axioms

There are various separation axioms for a topological space. The only ones we will need are the Hausdorff condition and normality.

Definition A.16. A topological space S is *Hausdorff* if, given any two distinct points x, y in S, there exist disjoint open sets U, V such that $x \in U$ and $y \in V$. A Hausdorff space is *normal* if given any two disjoint closed sets F, G in S, there exist disjoint open sets U, V such that $F \subset U$ and $G \subset V$ (Figure A.5).

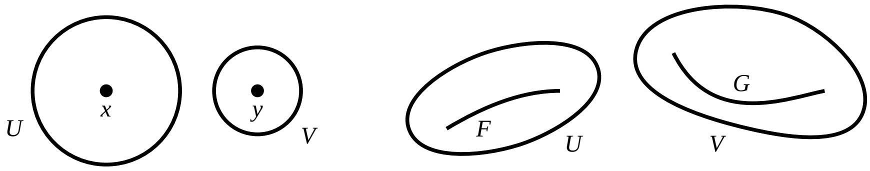


Fig. A.5. The Hausdorff condition and normality.

Proposition A.17. *Every singleton set* (*a one-point set*) *in a Hausdorff space* S *is closed.*

Proof. Let $x \in S$. For any $y \in S - \{x\}$, by the Hausdorff condition there exist an open set $U \ni x$ and an open set $V \ni y$ such that U and V are disjoint. In particular,

$$y \in V \subset S - U \subset S - \{x\}.$$

By the local criterion for openness (Lemma A.2), $S - \{x\}$ is open. Therefore, $\{x\}$ is closed. □

Example. The Euclidean space $\mathbb{R}^n$ is Hausdorff, for given distinct points x, y in $\mathbb{R}^n$, if $\varepsilon = \frac{1}{2}d(x,y)$, then the open balls $B(x,\varepsilon)$ and $B(y,\varepsilon)$ will be disjoint (Figure A.6).

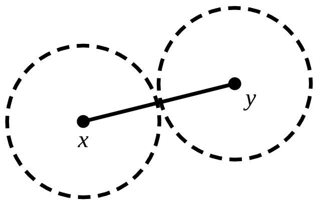


Fig. A.6. Two disjoint neighborhoods in $\mathbb{R}^n$.

Example A.18 (*Zariski topology*). Let $S = K^n$ be a vector space of dimension n over a field K, endowed with the Zariski topology. Every open set U in S is of the form $S - Z(I)$, where I is an ideal in $K[x_1, \ldots, x_n]$. The open set U is nonempty if and only if I is not the zero ideal. In the Zariski topology any two nonempty open sets intersect: if $U = S - Z(I)$ and $V = S - Z(J)$ are nonempty, then I and J are nonzero ideals and

$$\begin{aligned}
U \cap V &= (S - Z(I)) \cap (S - Z(J)) \\
&= S - (Z(I) \cup Z(J)) && \text{(de Morgan's law)} \\
&= S - Z(IJ), && \text{(Exercise A.5)}
\end{aligned}$$

which is nonempty because IJ is not the zero ideal. Therefore, K^n with the Zariski topology is not Hausdorff.

Proposition A.19. *Any subspace A of a Hausdorff space S is Hausdorff.*

Proof. Let x and y be distinct points in A. Since S is Hausdorff, there exist disjoint neighborhoods U and V of x and y respectively in S. Then $U \cap A$ and $V \cap A$ are disjoint neighborhoods of x and y respectively in A. □

A.6 Product Topology

The *Cartesian product* of two sets A and B is the set $A \times B$ of all ordered pairs (a,b) with $a \in A$ and $b \in B$. Given two topological spaces X and Y, consider the collection $\mathcal{B}$ of subsets of $X \times Y$ of the form $U \times V$, with U open in X and V open in Y. We will call elements of $\mathcal{B}$ *basic open sets* in $X \times Y$. If $U_1 \times V_1$ and $U_2 \times V_2$ are in $\mathcal{B}$, then

$$(U_1 \times V_1) \cap (U_2 \times V_2) = (U_1 \cap U_2) \times (V_1 \cap V_2),$$

which is also in $\mathcal{B}$ (Figure A.7). From this, it follows easily that $\mathcal{B}$ satisfies the conditions of Proposition A.8 for a basis and generates a topology on $X \times Y$, called the *product topology*. Unless noted otherwise, this will always be the topology we assign to the product of two topological spaces.

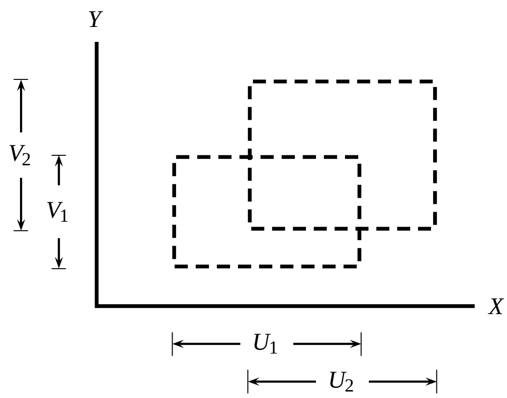


Fig. A.7. Intersection of two basic open subsets in $X \times Y$.

Proposition A.20. *Let $\{U_i\}$ and $\{V_j\}$ be bases for the topological spaces X and Y, respectively. Then $\{U_i \times V_j\}$ is a basis for $X \times Y$.*

Proof. Given an open set W in $X \times Y$ and point $(x,y) \in W$, we can find a basic open set $U \times V$ in $X \times Y$ such that $(x,y) \in U \times V \subset W$. Since U is open in X and $\{U_i\}$ is a basis for X,

$$x \in U_i \subset U$$

for some U_i. Similarly,

$$y \in V_j \subset V$$

for some V_j. Therefore,

$$(x,y) \in U_i \times V_j \subset U \times V \subset W.$$

By the definition of a basis, $\{U_i \times V_j\}$ is a basis for $X \times Y$. □

Corollary A.21. *The product of two second countable spaces is second countable.*

Proposition A.22. *The product of two Hausdorff spaces X and Y is Hausdorff.*

Proof. Given two distinct points $(x_1,y_1),(x_2,y_2)$ in $X \times Y$, without loss of generality we may assume that $x_1 \neq x_2$. Since X is Hausdorff, there exist disjoint open sets U_1,U_2 in X such that $x_1 \in U_1$ and $x_2 \in U_2$. Then $U_1 \times Y$ and $U_2 \times Y$ are disjoint neighborhoods of (x_1,y_1) and (x_2,y_2) (Figure A.8), so $X \times Y$ is Hausdorff. □

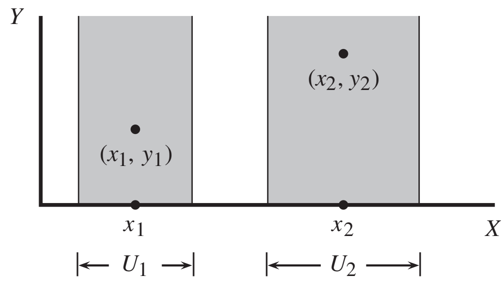


Fig. A.8. Two disjoint neighborhoods in $X \times Y$.

The product topology can be generalized to the product of an arbitrary collection $\{X_\alpha\}_{\alpha\in A}$ of topological spaces. Whatever the definition of the product topology, the projection maps $\pi_{\alpha_i}\colon \prod_\alpha X_\alpha \to X_{\alpha_i}$, $\pi_{\alpha_i}(\prod x_\alpha) = x_{\alpha_i}$ should all be continuous. Thus, for each open set U_{α_i} in X_{α_i}, the inverse image $\pi_{\alpha_i}^{-1}(U_{\alpha_i})$ should be open in $\prod_\alpha X_\alpha$. By the properties of open sets, a *finite* intersection $\bigcap_{i=1}^r \pi_{\alpha_i}^{-1}(U_{\alpha_i})$ should also be open. Such a finite intersection is a set of the form $\prod_{\alpha\in A} U_\alpha$, where U_α is open in X_α and $U_\alpha = X_\alpha$ for all but finitely many $\alpha \in A$. We define the *product topology* on the Cartesian product $\prod_{\alpha\in A} X_\alpha$ to be the topology with basis consisting of sets of this form. The product topology is the coarsest topology on $\prod_\alpha X_\alpha$ such that all the projection maps $\pi_{\alpha_i}\colon \prod_\alpha X_\alpha \to X_{\alpha_i}$ are continuous.

A.7 Continuity

Let $f: X \to Y$ be a function of topological spaces. Mimicking the definition from advanced calculus, we say that f is *continuous at a point* p in X if for every neighborhood V of $f(p)$ in Y, there is a neighborhood U of p in X such that $f(U) \subset V$. We say that f is *continuous on* X if it is continuous at every point of X.

Proposition A.23 (Continuity in terms of open sets). *A function* $f: X \to Y$ *is continuous if and only if the inverse image of any open set is open.*

Proof.
($\Rightarrow$) Suppose V is open in Y. To show that $f^{-1}(V)$ is open in X, let $p \in f^{-1}(V)$. Then $f(p) \in V$. Since f is assumed to be continuous at p, there is a neighborhood U of p such that $f(U) \subset V$. Therefore, $p \in U \subset f^{-1}(V)$. By the local criterion for openness (Lemma A.2), $f^{-1}(V)$ is open in X.

($\Leftarrow$) Let p be a point in X, and V a neighborhood of $f(p)$ in Y. By hypothesis, $f^{-1}(V)$ is open in X. Since $f(p) \in V$, $p \in f^{-1}(V)$. Then $U = f^{-1}(V)$ is a neighborhood of p such that $f(U) = f(f^{-1}(V)) \subset V$, so f is continuous at p. □

Example A.24 (*Continuity of an inclusion map*). If A is a subspace of X, then the inclusion map $i: A \to X$, $i(a) = a$ is continuous.

Proof. If U is open in X, then $i^{-1}(U) = U \cap A$, which is open in the subspace topology of A. □

Example A.25 (*Continuity of a projection map*). The projection $\pi: X \times Y \to X$, $\pi(x,y) = x$ is continuous.

Proof. Let U be open in X. Then $\pi^{-1}(U) = U \times Y$, which is open in the product topology on $X \times Y$. □

Proposition A.26. *The composition of continuous maps is continuous: if* $f: X \to Y$ *and* $g: Y \to Z$ *are continuous, then* $g \circ f: X \to Z$ *is continuous.*

Proof. Let V be an open subset of Z. Then

$$(g \circ f)^{-1}(V) = f^{-1}(g^{-1}(V)),$$

because for any $x \in X$,

$$x \in (g \circ f)^{-1}(V) \text{ iff } g(f(x)) \in V \text{ iff } f(x) \in g^{-1}(V) \text{ iff } x \in f^{-1}(g^{-1}(V)).$$

By Proposition A.23, since g is continuous, $g^{-1}(V)$ is open in Y. Similarly, since f is continuous, $f^{-1}(g^{-1}(V))$ is open in X. By Proposition A.23 again, $g \circ f: X \to Z$ is continuous. □

If A is a subspace of X and $f: X \to Y$ is a function, the *restriction* of f to A,

$$f|_A : A \to Y,$$

is defined by

$$(f|_A)(a) = f(a).$$

With $i: A \to X$ being the inclusion map, the restriction $f|_A$ is the composite $f \circ i$. Since both f and i are continuous (Example A.24) and the composition of continuous functions is continuous (Proposition A.26), we have the following corollary.

Corollary A.27. *The restriction $f|_A$ of a continuous function $f: X \to Y$ to a subspace A is continuous.*

Continuity may also be phrased in terms of closed sets.

Proposition A.28 (Continuity in terms of closed sets). *A function $f: X \to Y$ is continuous if and only if the inverse image of any closed set is closed.*

Proof. Problem A.9. □

A map $f: X \to Y$ is said to be *open* if the image of every open set in X is open in Y; similarly, $f: X \to Y$ is said to be *closed* if the image of every closed set in X is closed in Y.

If $f: X \to Y$ is a bijection, then its inverse map $f^{-1}: Y \to X$ is defined. In this context, for any subset $V \subset Y$, the notation $f^{-1}(V)$ a priori has two meanings. It can mean either the inverse image of V under the map f,

$$f^{-1}(V) = \{x \in X \mid f(x) \in V\},$$

or the image of V under the map f^{-1},

$$f^{-1}(V) = \{f^{-1}(y) \in X \mid y \in V\}.$$

Fortunately, because $y = f(x)$ iff $x = f^{-1}(y)$, these two meanings coincide.

A.8 Compactness

While its definition may not be intuitive, the notion of compactness is of central importance in topology. Let S be a topological space. A collection $\{U_\alpha\}$ of open subsets of S is said to *cover* S or to be *open cover* of S if $S \subset \bigcup_\alpha U_\alpha$. Of course, because S is the ambient space, this condition is equivalent to $S = \bigcup_\alpha U_\alpha$. A *subcover* of an open cover is a subcollection whose union still contains S. The topological space S is said to be *compact* if every open cover of S has a finite subcover.

With the subspace topology, a subset A of a topological space S is a topological space in its own right. The subspace A can be covered by open sets in A or by open sets in S. An *open cover of A in S* is a collection $\{U_\alpha\}$ of open sets in S that covers A. In this terminology, A is compact if and only if every open cover of A in A has a finite subcover.

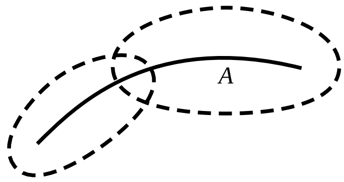


Fig. A.9. An open cover of A in S.

Proposition A.29. *A subspace A of a topological space S is compact if and only if every open cover of A in S has a finite subcover.*

Proof.
($\Rightarrow$) Assume A compact and let $\{U_\alpha\}$ be an open cover of A in S. This means $A \subset \bigcup_\alpha U_\alpha$. Hence,

$$A \subset \left(\bigcup_\alpha U_\alpha\right) \cap A = \bigcup_\alpha (U_\alpha \cap A).$$

Since A is compact, the open cover $\{U_\alpha \cap A\}$ has a finite subcover $\{U_{\alpha_i} \cap A\}_{i=1}^r$. Thus,

$$A \subset \bigcup_{i=1}^r (U_{\alpha_i} \cap A) \subset \bigcup_{i=1}^r U_{\alpha_i},$$

which means $\{U_{\alpha_i}\}_{i=1}^r$ is a finite subcover of $\{U_\alpha\}$.

($\Leftarrow$) Suppose every open cover of A in S has a finite subcover and let $\{V_\alpha\}$ be an open cover of A by in A. Then each $V_\alpha = U_\alpha \cap A$ for some open set U_α in S. Since

$$A \subset \bigcup_\alpha V_\alpha \subset \bigcup_\alpha U_\alpha,$$

by hypothesis, there are finitely many sets $\{U_{\alpha_i}\}$ such that $A \subset \bigcup_i U_{\alpha_i}$. Hence,

$$A \subset \left(\bigcup_i U_{\alpha_i}\right) \cap A = \bigcup_i (U_{\alpha_i} \cap A) = \bigcup_i V_{\alpha_i},$$

So $\{V_{\alpha_i}\}$ is a finite subcover of $\{V_\alpha\}$ that covers A. Therefore, A is compact. □

Proposition A.30. *A closed subset F of a compact topological space S is compact.*

Proof. Let $\{U_\alpha\}$ be an open cover of F in S. The collection $\{U_\alpha, S - F\}$ is then an open cover of S. By the compactness of S, there is a finite subcover $\{U_{\alpha_i}, S - F\}$ that covers S, so $F \subset \bigcup_i U_{\alpha_i}$. This proves that F is compact. □

Proposition A.31. *In a Hausdorff space S, it is possible to separate a compact subset K and a point p not in K by disjoint open sets; i.e., there exist an open set $U \supset K$ and an open set $V \ni p$ such that $U \cap V = \varnothing$.*

Proof. By the Hausdorff property, for every $x \in K$, there are disjoint open sets $U_x \ni x$ and $V_x \ni p$. The collection $\{U_x\}_{x\in K}$ is a cover of K by open subsets of S. Since K is compact, it has a finite subcover $\{U_{x_i}\}$.

Let $U = \bigcup_i U_{x_i}$ and $V = \bigcap_i V_{x_i}$. Then U is an open set of S containing K. Being the intersection of finitely many open sets containing p, V is an open set containing p. Moreover, the set

$$U \cap V = \bigcup_i (U_{x_i} \cap V)$$

is empty since each $U_{x_i} \cap V \subset U_{x_i} \cap V_{x_i}$, which is empty. □

Proposition A.32. *Every compact subset K of a Hausdorff space S is closed.*

Proof. By the preceding proposition, for every point p in $S - K$, there is an open set V such that $p \in V \subset S - K$. This proves that $S - K$ is open. Hence, K is closed. □

Exercise A.33 (Compact Hausdorff space).* Prove that a compact Hausdorff space is normal. (Normality was defined in Definition A.16.)

Proposition A.34. *The image of a compact set under a continuous map is compact.*

Proof. Let $f\colon X \to Y$ be a continuous map and K a compact subset of X. Suppose $\{U_\alpha\}$ is a cover of $f(K)$ by open subsets of Y. Since f is continuous, the inverse images $f^{-1}(U_\alpha)$ are all open. Moreover,

$$K \subset f^{-1}(f(K)) \subset f^{-1}\left(\bigcup_\alpha U_\alpha\right) = \bigcup_\alpha f^{-1}(U_\alpha).$$

So $\{f^{-1}(U_\alpha)\}$ is an open cover of K in X. By the compactness of K, there is a finite subcollection $\{f^{-1}(U_{\alpha_i})\}$ such that

$$K \subset \bigcup_i f^{-1}(U_{\alpha_i}) = f^{-1}\left(\bigcup_i U_{\alpha_i}\right).$$

Then $f(K) \subset \bigcup_i U_{\alpha_i}$. Thus, $f(K)$ is compact. □

Proposition A.35. *A continuous map $f\colon X \to Y$ from a compact space X to a Hausdorff space Y is a closed map.*

Proof. Let F be a closed subset of the compact space X. By Proposition A.30, F is compact. As the image of a compact set under a continuous map, $f(F)$ is compact in Y (Proposition A.34). As a compact subset of the Hausdorff space Y, $f(F)$ is closed (Proposition A.32). □

A continuous bijection $f\colon X \to Y$ whose inverse is also continuous is called a *homeomorphism*.

Corollary A.36. *A continuous bijection $f\colon X \to Y$ from a compact space X to a Hausdorff space Y is a homeomorphism.*

Proof. By Proposition A.28, to show that $f^{-1}: Y \to X$ is continuous, it suffices to prove that for every closed set F in X, the set $(f^{-1})^{-1}(F) = f(F)$ is closed in Y, i.e., that f is a closed map. The corollary then follows from Proposition A.35. □

Exercise A.37 (Finite union of compact sets). Prove that a finite union of compact subsets of a topological space is compact.

We mention without proof an important result. For a proof, see [29, Theorem 26.7, p. 167, and Theorem 37.3, p. 234].

Theorem A.38 (The Tychonoff theorem). *The product of any collection of compact spaces is compact in the product topology.*

A.9 Boundedness in $\mathbb{R}^n$

A subset A of $\mathbb{R}^n$ is said to be *bounded* if it is contained in some open ball $B(p,r)$; otherwise, it is *unbounded*.

Proposition A.39. *A compact subset of $\mathbb{R}^n$ is bounded.*

Proof. If A were an unbounded subset of $\mathbb{R}^n$, then the collection $\{B(0,i)\}_{i=1}^{\infty}$ of open balls with radius increasing to infinity would be an open cover of A in $\mathbb{R}^n$ that does not have a finite subcover. □

By Propositions A.39 and A.32, a compact subset of $\mathbb{R}^n$ is closed and bounded. The converse is also true.

Theorem A.40 (The Heine–Borel theorem). *A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

For a proof, see for example [29].

A.10 Connectedness

Definition A.41. A topological space S is *disconnected* if it is the union $S = U \cup V$ of two disjoint nonempty open subsets U and V (Figure A.10). It is *connected* if it is not disconnected. A subset A of S is *disconnected* if it is disconnected in the subspace topology.

Proposition A.42. *A subset A of a topological space S is disconnected if and only if there are open sets U and V in S such that*

(i) $U \cap A \neq \varnothing$, $V \cap A \neq \varnothing$,
(ii) $U \cap V \cap A = \varnothing$,
(iii) $A \subset U \cup V$.

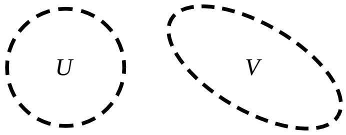


Fig. A.10. A disconnected space.

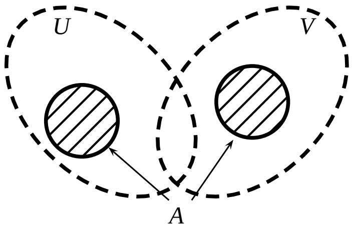


Fig. A.11. A separation of *A*.

A pair of open sets in S with these properties is called a separation *of A* (*Figure* A.11).

Proof. Problem A.15. □

Proposition A.43. *The image of a connected space X under a continuous map $f\colon X \to Y$ is connected.*

Proof. Suppose $f(X)$ is not connected. Then there is a separation $\{U,V\}$ of $f(X)$ in Y. By the continuity of f, both $f^{-1}(U)$ and $f^{-1}(V)$ are open in X. We claim that $\{f^{-1}(U), f^{-1}(V)\}$ is a separation of X.

(i) Since $U \cap f(X) \neq \varnothing$, the open set $f^{-1}(U) \neq \varnothing$.
(ii) If $x \in f^{-1}(U) \cap f^{-1}(V)$, then $f(x) \in U \cap V \cap f(X) = \varnothing$, a contradiction. Hence, $f^{-1}(U) \cap f^{-1}(V) = \varnothing$.
(iii) Since $f(X) \subset U \cup V$, we have $X \subset f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V)$.

The existence of a separation of X contradicts the connectedness of X. This contradiction proves that $f(X)$ is connected. □

Proposition A.44. *In a topological space S, the union of a collection of connected subsets A_α having a point p in common is connected.*

Proof. Suppose $\bigcup_\alpha A_\alpha = U \cup V$, where U and V are disjoint open subsets of $\bigcup_\alpha A_\alpha$. The point $p \in \bigcup_\alpha A_\alpha$ belongs to U or V. Assume without loss of generality that $p \in U$.

For each α,

$$A_\alpha = A_\alpha \cap (U \cup V) = (A_\alpha \cap U) \cup (A_\alpha \cap V).$$

The two open sets $A_\alpha \cap U$ and $A_\alpha \cap V$ of A_α are clearly disjoint. Since $p \in A_\alpha \cap U$, $A_\alpha \cap U$ is nonempty. By the connectedness of A_α, $A_\alpha \cap V$ must be empty for all α. Hence,

$$V = \left(\bigcup_\alpha A_\alpha\right) \cap V = \bigcup_\alpha (A_\alpha \cap V)$$

is empty. So $\bigcup_\alpha A_\alpha$ must be connected. □

A.11 Connected Components

Let x be a point in a topological space S. By Proposition A.44, the union C_x of all connected subsets of S containing x is connected. It is called the *connected component* of S containing x.

Proposition A.45. *Let C_x be a connected component of a topological space S. Then a connected subset A of S is either disjoint from C_x or is contained entirely in C_x.*

Proof. If A and C_x have a point in common, then by Proposition A.44, $A \cup C_x$ is a connected set containing x. Hence, $A \cup C_x \subset C_x$, which implies that $A \subset C_x$. □

Accordingly, the connected component C_x is the largest connected subset of S containing x in the sense that it contains every connected subset of S containing x.

Corollary A.46. *For any two points x, y in a topological space S, the connected components C_x and C_y either are disjoint or coincide.*

Proof. If C_x and C_y are not disjoint, then by Proposition A.45, they are contained in each other. In this case, $C_x = C_y$. □

As a consequence of Corollary A.46, the connected components of S partition S into disjoint subsets.

A.12 Closure

Let S be a topological space and A a subset of S.

Definition A.47. The *closure* of A in S, denoted $\overline{A}$, $\mathrm{cl}(A)$, or $\mathrm{cl}_S(A)$, is defined to be the intersection of all the closed sets containing A.

The advantage of the bar notation $\overline{A}$ is its simplicity, while the advantage of the $\mathrm{cl}_S(A)$ notation is its indication of the ambient space. If $A \subset B \subset S$, then the closure of A in B and the closure of A in S need not be the same. In this case, it is useful to have the notations $\mathrm{cl}_B(A)$ and $\mathrm{cl}_M(A)$ for the two closures.

As an intersection of closed sets, $\overline{A}$ is a closed set. It is the smallest closed set containing A in the sense that any closed set containing A contains $\overline{A}$.

Proposition A.48 (Local characterization of closure). *Let A be a subset of a topological space S. A point $p \in S$ is in the closure* $\mathrm{cl}(A)$ *if and only if every neighborhood of p contains a point of A* (Figure A.12).

Here by "local," we mean a property satisfied by a basis of neighborhoods at a point.

Proof. We will prove the proposition in the form of its contrapositive:

$$p \notin \mathrm{cl}(A) \quad \Longleftrightarrow \quad \text{there is a neighborhood of } p \text{ disjoint from } A.$$

($\Rightarrow$) Suppose

$$p \notin \mathrm{cl}(A) = \bigcap \{F \text{ closed in } S \mid F \supset A\}.$$

Then $p \notin$ some closed set F containing A. It follows that $p \in S - F$, an open set disjoint from A.

($\Leftarrow$) Suppose $p \in$ an open set U disjoint from A. Then the complement $F := S - U$ is a closed set containing A and not containing p. Therefore, $p \notin \mathrm{cl}(A)$. □

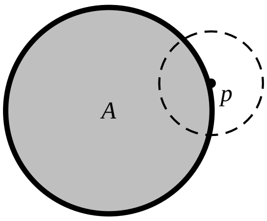


Fig. A.12. Every neighborhood of p contains a point of A.

Example. The closure of the open disk $B(\mathbf{0}, r)$ in $\mathbb{R}^2$ is the closed disk

$$\overline{B}(\mathbf{0}, r) = \{p \in \mathbb{R}^2 \mid d(p, \mathbf{0}) \le r\}.$$

Definition A.49. A point p in S is an *accumulation point* of A if every neighborhood of p in S contains a point of A other than p. The set of all accumulation points of A is denoted $\mathrm{ac}(A)$.

If U is a neighborhood of p in S, we call $U - \{p\}$ a *deleted neighborhood* of p. An equivalent condition for p to be an accumulation point of A is to require that every deleted neighborhood of p in S contain a point of A. In some books an accumulation point is called a *limit point*.

Example. If $A = [0,1[\cup \{2\}$ in $\mathbb{R}^1$, then the closure of A is $[0,1] \cup \{2\}$, but the set of accumulation points of A is only the closed interval $[0,1]$.

Proposition A.50. *Let A be a subset of a topological space S. Then*

$$\mathrm{cl}(A) = A \cup \mathrm{ac}(A).$$

Proof.
$(\supset)$ By definition, $A \subset \mathrm{cl}(A)$. By the local characterization of closure (Proposition A.48), $\mathrm{ac}(A) \subset \mathrm{cl}(A)$. Hence, $A \cup \mathrm{ac}(A) \subset \mathrm{cl}(A)$.

$(\subset)$ Suppose $p \in \mathrm{cl}(A)$. Either $p \in A$ or $p \notin A$. If $p \in A$, then $p \in A \cup \mathrm{ac}(A)$. Suppose $p \notin A$. By Proposition A.48, every neighborhood of p contains a point of A, which cannot be p, since $p \notin A$. Therefore, every deleted neighborhood of p contains a point of A. In this case,

$$p \in \mathrm{ac}(A) \subset A \cup \mathrm{ac}(A).$$

So $\mathrm{cl}(A) \subset A \cup \mathrm{ac}(A)$. □

Proposition A.51. *A set A is closed if and only if $A = \overline{A}$.*

Proof.
$(\Leftarrow)$ If $A = \overline{A}$, then A is closed because $\overline{A}$ is closed.

$(\Rightarrow)$ Suppose A is closed. Then A is a closed set containing A, so that $\overline{A} \subset A$. Because $A \subset \overline{A}$, equality holds. □

Proposition A.52. *If $A \subset B$ in a topological space S, then $\overline{A} \subset \overline{B}$.*

Proof. Since $\overline{B}$ contains B, it also contains A. As a closed subset of S containing A, it contains $\overline{A}$ by definition. □

Exercise A.53 (Closure of a finite union or finite intersection). Let A and B be subsets of a topological space S. Prove the following:

(a) $\overline{A \cup B} = \overline{A} \cup \overline{B}$,
(b) $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$.

The example of $A =]a, 0[$ and $B =]0, b[$ in the real line shows that, in general, $\overline{A \cap B} \neq \overline{A} \cap \overline{B}$.

A.13 Convergence

Let S be a topological space. A *sequence* in S is a map from the set $\mathbb{Z}^+$ of positive integers to S. We write a sequence as

$$\langle x_i \rangle \quad \text{or} \quad x_1, x_2, x_3, \ldots .$$

Definition A.54. The sequence $\langle x_i \rangle$ *converges* to p if for every neighborhood U of p, there is a positive integer N such that for all $i \geq N$, $x_i \in U$. In this case we say that p is a *limit* of the sequence $\langle x_i \rangle$ and write $x_i \to p$ or $\lim_{i \to \infty} x_i = p$.

Proposition A.55 (Uniqueness of the limit). *In a Hausdorff space S, if a sequence $\langle x_i \rangle$ converges to p and to q, then $p = q$.*

Proof. Problem A.19. □

Thus, in a Hausdorff space we may speak of *the* limit of a convergent sequence.

Proposition A.56 (The sequence lemma). *Let S be a topological space and A a subset of S. If there is a sequence $\langle a_i \rangle$ in A that converges to p, then $p \in \mathrm{cl}(A)$. The converse is true if S is first countable.*

Proof.
($\Rightarrow$) Suppose $a_i \to p$, where $a_i \in A$ for all i. By the definition of convergence, every neighborhood U of p contains all but finitely many of the points a_i. In particular, U contains a point in A. By the local characterization of closure (Proposition A.48), $p \in \mathrm{cl}(A)$.

($\Leftarrow$) Suppose $p \in \mathrm{cl}(A)$. Since S is first countable, we can find a countable basis of neighborhoods $\{U_n\}$ at p such that

$$U_1 \supset U_2 \supset \cdots .$$

By the local characterization of closure, in each U_i there is a point $a_i \in A$. We claim that the sequence $\langle a_i \rangle$ converges to p. If U is any neighborhood of p, then by the definition of a basis of neighborhoods at p, there is a U_N such that $p \in U_N \subset U$. For all $i \geq N$, we then have

$$U_i \subset U_N \subset U,$$

Therefore, for all $i \geq N$,

$$a_i \in U_i \subset U.$$

This proves that $\langle a_i \rangle$ converges to p. □

Problems

A.1. Set theory
If U_1 and U_2 are subsets of a set X, and V_1 and V_2 are subsets of a set Y, prove that

$$(U_1 \times V_1) \cap (U_2 \times V_2) = (U_1 \cap U_2) \times (V_1 \cap V_2).$$

A.2. Union and intersection
Suppose $U_1 \cap V_1 = U_2 \cap V_2 = \varnothing$ in a topological space S. Show that the intersection $U_1 \cap U_2$ is disjoint from the union $V_1 \cup V_2$. (*Hint*: Use the distributive property of an intersection over a union.)

A.3. Closed sets
Let S be a topological space. Prove the following two statements.

(a) If $\{F_i\}_{i=1}^n$ is a finite collection of closed sets in S, then $\bigcup_{i=1}^n F_i$ is closed.
(b) If $\{F_\alpha\}_{\alpha \in \mathrm{A}}$ is an arbitrary collection of closed sets in S, then $\bigcap_\alpha F_\alpha$ is closed.

A.4. Cubes versus balls
Prove that the open cube $]-a,a[^n$ is contained in the open ball $B(\mathbf{0},\sqrt{n}a)$, which in turn is contained in the open cube $]-\sqrt{n}a,\sqrt{n}a[^n$. Therefore, open cubes with arbitrary centers in $\mathbb{R}^n$ form a basis for the standard topology on $\mathbb{R}^n$.

A.5. Product of closed sets
Prove that if A is closed in X and B is closed in Y, then $A \times B$ is closed in $X \times Y$.

A.6. Characterization of a Hausdorff space by its diagonal
Let S be a topological space. The diagonal Δ in $S \times S$ is the set

$$\Delta = \{(x,x) \in S \times S\}.$$

Prove that S is Hausdorff if and only if the diagonal Δ is closed in $S \times S$. (*Hint*: Prove that S is Hausdorff if and only if $S \times S - \Delta$ is open in $S \times S$.)

A.7. Projection
Prove that if X and Y are topological spaces, then the projection $\pi : X \times Y \to X$, $\pi(x,y) = x$ is an open map.

A.8. The ε-δ criterion for continuity
Prove that a function $f\colon A \to \mathbb{R}^m$ is continuous at $p \in A$ if and only if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x \in A$ satisfying $d(x,p) < \delta$, one has $d(f(x), f(p)) < \varepsilon$.

A.9. Continuity in terms of closed sets
Prove Proposition A.28.

A.10. Continuity of a map into a product
Let X, Y_1, and Y_2 be topological spaces. Prove that a map $f = (f_1, f_2)\colon X \to Y_1 \times Y_2$ is continuous if and only if both components $f_i\colon X \to Y_i$ are continuous.

A.11. Continuity of the product map
Given two maps $f\colon X \to X'$ and $g\colon Y \to Y'$ of topological spaces, we define their *product* to be

$$f \times g\colon X \times Y \to X' \times Y', \quad (f \times g)(x,y) = (f(x), g(y)).$$

Note that if $\pi_1\colon X \times Y \to X$ and $\pi_2\colon X \times Y \to Y$ are the two projections, then $f \times g = (f \circ \pi_1, f \circ \pi_2)$. Prove that $f \times g$ is continuous if and only if both f and g are continuous.

A.12. Homeomorphism
Prove that if a continuous bijection $f\colon X \to Y$ is a closed map, then it is a homeomorphism (cf. Corollary A.36).

A.13.* The Lindelöf condition
Show that if a topological space is second countable, then it is Lindelöf; i.e., every open cover has a countable subcover.

A.14. Compactness
Prove that a finite union of compact sets in a topological space S is compact.

A.15.* Disconnected subset in terms of a separation
Prove Proposition A.42.

A.16. Local connectedness
A topological space S is said to be *locally connected at* $p \in S$ if for every neighborhood U of p, there is a connected neighborhood V of p such that $V \subset U$. The space S is *locally connected* if it is locally connected at every point. Prove that if S is locally connected, then the connected components of S are open.

A.17. Closure
Let U be an open subset and A an arbitrary subset of a topological space S. Prove that $U \cap \bar{A} \neq \varnothing$ if and only if $U \cap A \neq \varnothing$.

A.18. Countability
Prove that every second countable space is first countable.

A.19.* Uniqueness of the limit
Prove Proposition A.55.

A.20.* Closure in a product
Let S and Y be topological spaces and $A \subset S$. Prove that

$$\mathrm{cl}_{S\times Y}(A \times Y) = \mathrm{cl}_S(A) \times Y$$

in the product space $S \times Y$.

A.21. Dense subsets
A subset A of a topological space S is said to be *dense* in S if its closure $\mathrm{cl}(A) = S$.

(a) Prove that A is dense in S if and only if for every $p \in S$, every neighborhood U of p contains a point of A.
(b) Let K be a field. Prove that a Zariski open subset U of K^n is dense in K^n. (*Hint*: Example A.18.)

B

The Inverse Function Theorem on $\mathbb{R}^n$ and Related Results

This appendix reviews three logically equivalent theorems from real analysis, the inverse function theorem, the implicit function theorem, and the constant rank theorem, which describe the local behavior of a C^∞ map from $\mathbb{R}^n$ to $\mathbb{R}^m$. We will assume the inverse function theorem and deduce the other two, in the simplest cases, from the inverse function theorem. In Section 11 these theorems are applied to manifolds in order to clarify the local behavior of a C^∞ map when the map has maximal rank at a point or constant rank in a neighborhood.

B.1 The Inverse Function Theorem

A C^∞ map $f\colon U \to \mathbb{R}^n$ defined on an open subset U of $\mathbb{R}^n$ is *locally invertible* or a *local diffeomorphism* at a point p in U if f has a C^∞ inverse in some neighborhood of p. The inverse function theorem gives a criterion for a map to be locally invertible. We call the matrix $Jf = [\partial f^i/\partial x^j]$ of partial derivatives of f the *Jacobian matrix* of f and its determinant $\det[\partial f^i/\partial x^j]$ the *Jacobian determinant* of f.

Theorem B.1 (Inverse function theorem). *Let $f\colon U \to \mathbb{R}^n$ be a C^∞ map defined on an open subset U of $\mathbb{R}^n$. At any point p in U, the map f is invertible in some neighborhood of p if and only if the Jacobian determinant* $\det[\partial f^i/\partial x^j\,(p)]$ *is not zero.*

For a proof, see for example [35, Theorem 9.24, p. 221]. Although it apparently reduces the invertibility of f on an open set to a single number at p, because the Jacobian determinant is a continuous function, the nonvanishing of the Jacobian determinant at p is equivalent to its nonvanishing in a neighborhood of p.

Since the linear map represented by the Jacobian matrix $Jf(p)$ is the best linear approximation to f at p, it is plausible that f is invertible in a neighborhood of p if and only if $Jf(p)$ is also, i.e., if and only if $\det(Jf(p)) \neq 0$.

B.2 The Implicit Function Theorem

In an equation such as $f(x,y) = 0$, it is often impossible to solve explicitly for one of the variables in terms of the other. If we can show the existence of a function $y = h(x)$, which we may or may not be able to write down explicitly, such that $f(x,h(x)) = 0$, then we say that $f(x,y) = 0$ can be solved *implicitly* for y in terms of x. The implicit function theorem provides a sufficient condition on a system of equations $f^i(x^1,\dots,x^n) = 0$, $i = 1,\dots,m$ under which *locally* a set of variables can be solved implicitly as C^∞ functions of the other variables.

Example. Consider the equation

$$f(x,y) = x^2 + y^2 - 1 = 0.$$

The solution set is the unit circle in the xy-plane.

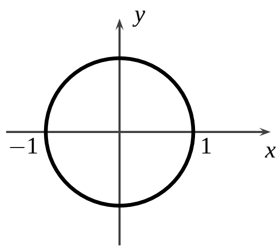


Fig. B.1. The unit circle.

From the picture we see that in a neighborhood of any point other than $(\pm 1,0)$, y is a function of x. Indeed,

$$y = \pm\sqrt{1-x^2},$$

and either function is C^∞ as long as $x \neq \pm 1$. At $(\pm 1,0)$, there is no neighborhood on which y is a function of x.

On a smooth curve $f(x,y) = 0$ in $\mathbb{R}^2$,

y can be expressed as a function of x in a neighborhood of a point (a,b)
$\iff$ the tangent line to $f(x,y) = 0$ at (a,b) is not vertical
$\iff$ the normal vector grad $f := \langle f_x, f_y\rangle$ to $f(x,y) = 0$ at (a,b) is not horizontal
$\iff$ $f_y(a,b) \neq 0$.

The implicit function theorem generalizes this condition to higher dimensions. We will deduce the implicit function theorem from the inverse function theorem.

Theorem B.2 (Implicit function theorem). *Let U be an open subset in $\mathbb{R}^n \times \mathbb{R}^m$ and $f\colon U \to \mathbb{R}^m$ a C^∞ map. Write $(x,y) = (x^1,\dots,x^n,y^1,\dots,y^m)$ for a point in U. At a point $(a,b) \in U$ where $f(a,b) = 0$ and the determinant* $\det[\partial f^i/\partial y^j(a,b)] \neq 0$, *there exist a neighborhood $A \times B$ of (a,b) in U and a unique function $h\colon A \to B$ such that in $A \times B \subset U \subset \mathbb{R}^n \times \mathbb{R}^m$,*

$$f(x,y) = 0 \quad \Longleftrightarrow \quad y = h(x).$$

Moreover, h is C^∞.

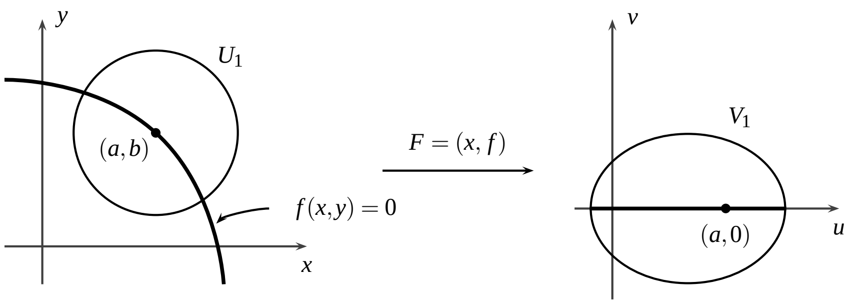


Fig. B.2. F^{-1} maps the u-axis to the zero set of f.

Proof. To solve $f(x,y) = 0$ for y in terms of x using the inverse function theorem, we first turn it into an inverse problem. For this, we need a map between two open sets of the same dimension. Since $f(x,y)$ is a map from an open set U in $\mathbb{R}^{n+m}$ to $\mathbb{R}^m$, it is natural to extend f to a map $F: U \to \mathbb{R}^{n+m}$ by adjoining x to it as the first n components:

$$F(x,y) = (u,v) = (x, f(x,y)).$$

To simplify the exposition, we will assume in the rest of the proof that $n = m = 1$. Then the Jacobian matrix of F is

$$JF = \begin{bmatrix} 1 & 0 \\ \partial f/\partial x & \partial f/\partial y \end{bmatrix}.$$

At the point (a,b),

$$\det JF(a,b) = \frac{\partial f}{\partial y}(a,b) \neq 0.$$

By the inverse function theorem, there are neighborhoods U_1 of (a,b) and V_1 of $F(a,b) = (a,0)$ in $\mathbb{R}^2$ such that $F: U_1 \to V_1$ is a diffeomorphism with C^∞ inverse F^{-1} (Figure B.2). Since $F: U_1 \to V_1$ is defined by

$$u = x,$$
$$v = f(x,y),$$

the inverse map $F^{-1}\colon V_1 \to U_1$ must be of the form

$$x = u,$$
$$y = g(u,v)$$

for some C^∞ function $g\colon V_1 \to \mathbb{R}$. Thus, $F^{-1}(u,v) = (u, g(u,v))$.

The two compositions $F^{-1} \circ F$ and $F \circ F^{-1}$ give

$$(x,y) = (F^{-1} \circ F)(x,y) = F^{-1}(x, f(x,y)) = (x, g(x, f(x,y))),$$
$$(u,v) = (F \circ F^{-1})(u,v) = F(u, g(u,v)) = (u, f(u, g(u,v))).$$

Hence,

$$y = g(x, f(x,y)) \quad \text{for all } (x,y) \in U_1, \tag{B.1}$$
$$v = f(u, g(u,v)) \quad \text{for all } (u,v) \in V_1. \tag{B.2}$$

If $f(x,y) = 0$, then (B.1) gives $y = g(x,0)$. This suggests that we define $h(x) = g(x,0)$ for all $x \in \mathbb{R}^1$ for which $(x,0) \in V_1$. The set of all such x is homeomorphic to $V_1 \cap (\mathbb{R}^1 \times \{0\})$ and is an open subset of $\mathbb{R}^1$. Since g is C^∞ by the inverse function theorem, h is also C^∞.

Claim. For $(x,y) \in U_1$ such that $(x,0) \in V_1$,

$$f(x,y) = 0 \iff y = h(x).$$

Proof (*of Claim*).
($\Rightarrow$) As we saw already, from (B.1), if $f(x,y) = 0$, then

$$y = g(x, f(x,y)) = g(x,0) = h(x). \tag{B.3}$$

($\Leftarrow$) If $y = h(x)$ and in (B.2) we set $(u,v) = (x,0)$, then

$$0 = f(x, g(x,0)) = f(x, h(x)) = f(x,y). \qquad \square$$

By the claim, in some neighborhood of $(a,b) \in U_1$, the zero set of $f(x,y)$ is precisely the graph of h. To find a product neighborhood of (a,b) as in the statement of the theorem, let $A_1 \times B$ be a neighborhood of (a,b) contained in U_1 and let $A = h^{-1}(B) \cap A_1$. Since h is continuous, A is open in the domain of h and hence in $\mathbb{R}^1$. Then $h(A) \subset B$,

$$A \times B \subset A_1 \times B \subset U_1, \quad \text{and} \quad A \times \{0\} \subset V_1.$$

By the claim, for $(x,y) \in A \times B$,

$$f(x,y) = 0 \iff y = h(x).$$

Equation (B.3) proves the uniqueness of h. $\square$

Replacing a partial derivative such as $\partial f/\partial y$ with a Jacobian matrix $[\partial f^i/\partial y^j]$, we can prove the general case of the implicit function theorem in exactly the same way. Of course, in the theorem $y^1, \ldots, y^m$ need not be the last m coordinates in $\mathbb{R}^{n+m}$; they can be any set of m coordinates in $\mathbb{R}^{n+m}$.

Theorem B.3. *The implicit function theorem is equivalent to the inverse function theorem.*

Proof. We have already shown, at least for one typical case, that the inverse function theorem implies the implicit function theorem. We now prove the reverse implication.

So assume the implicit function theorem, and let $f\colon U \to \mathbb{R}^n$ be a C^∞ map defined on an open subset U of $\mathbb{R}^n$ such that at some point $p \in U$, the Jacobian determinant $\det[\partial f^i/\partial x^j(p)]$ is nonzero. Finding a local inverse for $y = f(x)$ near p amounts to solving the equation

$$g(x,y) = f(x) - y = 0$$

for x in terms of y near $(p, f(p))$. Note that $\partial g^i/\partial x^j = \partial f^i/\partial x^j$. Hence,

$$\det\left[\frac{\partial g^i}{\partial x^j}(p, f(p))\right] = \det\left[\frac{\partial f^i}{\partial x^j}(p)\right] \neq 0.$$

By the implicit function theorem, x can be expressed in terms of y locally near $(p, f(p))$; i.e., there is a C^∞ function $x = h(y)$ defined in a neighborhood of $f(p)$ in $\mathbb{R}^n$ such that

$$g(x,y) = f(x) - y = f(h(y)) - y = 0.$$

Thus, $y = f(h(y))$. Since $y = f(x)$,

$$x = h(y) = h(f(x)).$$

Therefore, f and h are inverse functions defined near p and $f(p)$ respectively. □

B.3 Constant Rank Theorem

Every C^∞ map $f\colon U \to \mathbb{R}^m$ on an open set U of $\mathbb{R}^n$ has a *rank* at each point p in U, namely the rank of its Jacobian matrix $[\partial f^i/\partial x^j(p)]$.

Theorem B.4 (Constant rank theorem). *If $f\colon \mathbb{R}^n \supset U \to \mathbb{R}^m$ has constant rank k in a neighborhood of a point $p \in U$, then after a suitable change of coordinates near p in U and $f(p)$ in $\mathbb{R}^m$, the map f assumes the form*

$$(x^1, \dots, x^n) \mapsto (x^1, \dots, x^k, 0, \dots, 0).$$

More precisely, there are a diffeomorphism G of a neighborhood of p in U sending p to the origin in $\mathbb{R}^n$ and a diffeomorphism F of a neighborhood of $f(p)$ in $\mathbb{R}^m$ sending $f(p)$ to the origin in $\mathbb{R}^m$ such that

$$(F \circ f \circ G)^{-1}(x^1, \dots, x^n) = (x^1, \dots, x^k, 0, \dots, 0).$$

Proof (*for* $n = m = 2$, $k = 1$). Suppose $f = (f^1, f^2)\colon \mathbb{R}^2 \supset U \to \mathbb{R}^2$ has constant rank 1 in a neighborhood of $p \in U$. By reordering the functions f^1, f^2 or the variables x, y, we may assume that $\partial f^1/\partial x(p) \neq 0$. (Here we are using the fact that f has rank ≥ 1 at p.) Define $G\colon U \to \mathbb{R}^2$ by

$$G(x,y) = (u,v) = (f^1(x,y), y).$$

The Jacobian matrix of G is

$$JG = \begin{bmatrix} \partial f^1/\partial x & \partial f^1/\partial y \\ 0 & 1 \end{bmatrix}.$$

Since $\det JG(p) = \partial f^1/\partial x(p) \neq 0$, by the inverse function theorem there are neighborhoods U_1 of $p \in \mathbb{R}^2$ and V_1 of $G(p) \in \mathbb{R}^2$ such that $G\colon U_1 \to V_1$ is a diffeomorphism. By making U_1 a sufficiently small neighborhood of p, we may assume that f has constant rank 1 on U_1.

On V_1,

$$(u,v) = (G \circ G^{-1})(u,v) = (f^1 \circ G^{-1}, y \circ G^{-1})(u,v).$$

Comparing the first components gives $u = (f^1 \circ G^{-1})(u,v)$. Hence,

$$\begin{aligned}(f \circ G^{-1})(u,v) &= (f^1 \circ G^{-1}, f^2 \circ G^{-1})(u,v) \\ &= (u, f^2 \circ G^{-1}(u,v)) \\ &= (u, h(u,v)),\end{aligned}$$

where we set $h = f^2 \circ G^{-1}$.

Because $G^{-1}\colon V_1 \to U_1$ is a diffeomorphism and f has constant rank 1 on U_1, the composite $f \circ G^{-1}$ has constant rank 1 on V_1. Its Jacobian matrix is

$$J(f \circ G^{-1}) = \begin{bmatrix} 1 & 0 \\ \partial h/\partial u & \partial h/\partial v \end{bmatrix}.$$

For this matrix to have constant rank 1, $\partial h/\partial v$ must be identically zero on V_1. (Here we are using the fact that f has rank ≤ 1 in a neighborhood of p). Thus, h is a function of u alone and we may write

$$(f \circ G^{-1})(u,v) = (u, h(u)).$$

Finally, let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the change of coordinates $F(x,y) = (x, y - h(x))$. Then

$$(F \circ f \circ G^{-1})(u,v) = F(u,h(u)) = (u, h(u) - h(u)) = (u,0). \qquad \square$$

Example B.5. If a C^∞ map $f\colon \mathbb{R}^n \supset U \to \mathbb{R}^n$ defined on an open subset U of $\mathbb{R}^n$ has nonzero Jacobian determinant $\det(Jf(p)) \neq 0$ at a point $p \in U$, then by continuity it has nonzero Jacobian determinant in a neighborhood of p. Therefore, it has constant rank n in a neighborhood of p.

Problems

B.1.* The rank of a matrix
The *rank* of a matrix A, denoted $\operatorname{rk} A$, is defined to be the number of linearly independent columns of A. By a theorem in linear algebra, it is also the number of linearly independent rows of A. Prove the following lemma.

Lemma. *Let A be an $m \times n$ matrix (not necessarily square), and k a positive integer. Then $\operatorname{rk} A \geq k$ if and only if A has a nonsingular $k \times k$ submatrix. Equivalently, $\operatorname{rk} A \leq k-1$ if and only if all $k \times k$ minors of A vanish. (A $k \times k$* minor *of a matrix A is the determinant of a $k \times k$ submatrix of A.)*

B.2.* Matrices of rank at most *r*
For an integer $r \geq 0$, define D_r to be the subset of $\mathbb{R}^{m\times n}$ consisting of all $m \times n$ real matrices of rank at most r. Show that D_r is a closed subset of $\mathbb{R}^{m\times n}$. (*Hint*: Use Problem B.1.)

B.3.* Maximal rank
We say that the rank of an $m \times n$ matrix A is *maximal* if $\operatorname{rk} A = \min(m,n)$. Define $D_{\max}$ to be the subset of $\mathbb{R}^{m\times n}$ consisting of all $m \times n$ matrices of maximal rank r. Show that $D_{\max}$ is an open subset of $\mathbb{R}^{m\times n}$. (*Hint*: Suppose $n \leq m$. Then $D_{\max} = \mathbb{R}^{m\times n} - D_{n-1}$. Apply Problem B.2.)

B.4.* Degeneracy loci and maximal rank locus of a map
Let $F\colon S \to \mathbb{R}^{m\times n}$ be a continuous map from a topological space S to the space $\mathbb{R}^{m\times n}$. The *degeneracy locus of rank r* of F is defined to be

$$D_r(F) := \{x \in S \mid \operatorname{rk} F(x) \leq r\}.$$

(a) Show that the degeneracy locus $D_r(F)$ is a closed subset of S. (*Hint*: $D_r(F) = F^{-1}(D_r)$, where D_r was defined in Problem B.2.)
(b) Show that the *maximal rank locus* of F,

$$D_{\max}(F) := \{x \in S \mid \operatorname{rk} F(x) \text{ is maximal}\},$$

is an open subset of S.

B.5. Rank of a composition of linear maps
Suppose V, W, V', W' are finite-dimensional vector spaces.

(a) Prove that if the linear map $L\colon V \to W$ is surjective, then for any linear map $f\colon W \to W'$, $\operatorname{rk}(f \circ L) = \operatorname{rk} f$.
(b) Prove that if the linear map $L\colon V \to W$ is injective, then for any linear map $g\colon V' \to V$, $\operatorname{rk}(L \circ g) = \operatorname{rk} g$.

B.6. Constant rank theorem
Generalize the proof of the constant rank theorem (Theorem B.4) in the text to arbitrary n, m, and k.

B.7. Equivalence of the constant rank theorem and the inverse function theorem
Use the constant rank theorem (Theorem B.4) to prove the inverse function theorem (Theorem B.1). Hence, the two theorems are equivalent.

Suppose $b, b' \in B^k$ both map to c under j. Then $j(b - b') = jb - jb' = c - c = 0$. By the exactness at B^k, $b - b' = i(a'')$ for some $a'' \in A^k$.

With the choice of b as preimage, the element $d^*[c]$ is represented by a cocycle $a \in A^{k+1}$ such that $i(a) = db$. Similarly, with the choice of b' as preimage, the element $d^*[c]$ is represented by a cocycle $a' \in A^{k+1}$ such that $i(a') = db'$. Then $i(a - a') = d(b - b') = di(a'') = id(a'')$. Since i is injective, $a - a' = da''$, and thus $[a] = [a']$. This proves that $d^*[c]$ is independent of the choice of b.

The proof that the cohomology class of a is independent of the choice of c in the cohomology class $[c]$ can be summarized by the commutative diagram

$$\begin{array}{ccccc}
a - a' & \overset{i}{\rightarrowtail} & db - db' = 0 & & \\
 & & \uparrow{\scriptstyle d} & & \\
 & & b - b' & \overset{j}{\longmapsto} & c - c' \\
 & & \uparrow{\scriptstyle d} & & \uparrow{\scriptstyle d} \\
 & & b'' & \underset{j}{\longmapsto} & c''.
\end{array}$$

Suppose $[c] = [c'] \in H^k(\mathcal{C})$. Then $c - c' = dc''$ for some $c'' \in C^{k-1}$. By the surjectivity of $j \colon B^{k-1} \to C^{k-1}$, there is a $b'' \in B^{k-1}$ that maps to c'' under j. Choose $b \in B^k$ such that $j(b) = c$ and let $b' = b - db'' \in B^k$. Then $j(b') = j(b) - jdb'' = c - dj(b'') = c - dc'' = c'$. With the choice of b as preimage, $d^*[c]$ is represented by a cocycle $a \in A^{k+1}$ such that $i(a) = db$. With the choice of b' as preimage, $d^*[c]$ is represented by a cocycle $a' \in A^{k+1}$ such that $i(a') = db'$. Then

$$i(a - a') = d(b - b') = ddb'' = 0.$$

By the injectivity of i, $a = a'$, so $[a] = [a']$. This shows that $d^*[c]$ is independent of the choice of c in the cohomology class $[c]$. ◊

A.33 Compact Hausdorff space

Let S be a compact Hausdorff space, and A, B two closed subsets of S. By Proposition A.30, A and B are compact. By Proposition A.31, for any $a \in A$ there are disjoint open sets $U_a \ni a$ and $V_a \supset B$. Since A is compact, the open cover $\{U_a\}_{a \in A}$ for A has a finite subcover $\{U_{a_i}\}_{i=1}^n$. Let $U = \bigcup_{i=1}^n U_{a_i}$ and $V = \bigcap_{i=1}^n V_{a_i}$. Then $A \subset U$ and $B \subset V$. The open sets U and V are disjoint because if $x \in U \cap V$, then $x \in U_{a_i}$ for some i and $x \in V_{a_i}$ for the same i, contradicting the fact that $U_{a_i} \cap V_{a_i} = \varnothing$. ◊

$$\int_{\mathcal{H}^n} d\omega = (-1)^n \int_{\mathbb{R}^{n-1}} f(x^1,\dots,x^{n-1},0)\,dx^1\cdots dx^{n-1} = \int_{\partial\mathcal{H}^n} \omega$$

because $(-1)^n\mathbb{R}^{n-1}$ is precisely $\partial\mathcal{H}^n$ with its boundary orientation. So Stokes' theorem also holds in this case. $\Diamond$

23.5 Take the exterior derivative of $x^2+y^2+z^2=1$ to obtain a relation among the 1-forms dx, dy, and dz on S^2. Then show for example that for $x \neq 0$, one has $dx\wedge dy = (z/x)dy\wedge dz$.

24.1 Assume $\omega = df$. Derive a contradiction using Problem 8.10(b) and Proposition 17.2.

25.4* The snake lemma

If we view each column of the given commutative diagram as a cochain complex, then the diagram is a short exact sequence of cochain complexes

$$0 \to \mathcal{A} \to \mathcal{B} \to \mathcal{C} \to 0.$$

By the zig-zag lemma, it gives rise to a long exact sequence in cohomology. In the long exact sequence, $H^0(\mathcal{A}) = \ker\alpha$, $H^1(\mathcal{A}) = A^1/\operatorname{im}\alpha = \operatorname{coker}\alpha$, and similarly for $\mathcal{B}$ and $\mathcal{C}$. $\Diamond$

26.2 Define $d_{-1}=0$. Then the given exact sequence is equivalent to a collection of short exact sequences

$$0 \to \operatorname{im} d_{k-1} \to A^k \xrightarrow{d_k} \operatorname{im} d_k \to 0, \qquad k = 0,\dots,m-1$$

By the rank-nullity theorem,

$$\dim A^k = \dim(\operatorname{im} d_{k-1}) + \dim(\operatorname{im} d_k).$$

When we compute the alternating sum of the left-hand side, the right-hand side will cancel to 0. $\Diamond$

28.1 Let U be the punctured projective plane $\mathbb{R}P^2 - \{p\}$ and V a small disk containing p. Because U can be deformation retracted to the boundary circle, which after identification is in fact $\mathbb{R}P^1$, U has the homotopy type of $\mathbb{R}P^1$. Since $\mathbb{R}P^1$ is homeomorphic to S^1, $H^*(U) \simeq H^*(S^1)$. Apply the Mayer–Vietoris sequence. The answer is $H^0(\mathbb{R}P^2) = \mathbb{R}$, $H^k(\mathbb{R}P^2) = 0$ for $k > 0$.

28.2 $H^k(S^n) = \mathbb{R}$ for $k = 0, n$, and $H^k(S^n) = 0$ otherwise.

28.3 One way is to apply the Mayer–Vietoris sequence to $U = \mathbb{R}^2 - \{p\}$, $V = \mathbb{R}^2 - \{q\}$.

A.13* The Lindelöf condition

Let $\{B_i\}_{i\in I}$ be a countable basis and $\{U_\alpha\}_{\alpha\in \mathrm{A}}$ an open cover of the topological space S. For every $p \in U_\alpha$, there exists a B_i such that

$$p \in B_i \subset U_\alpha.$$

Since this B_i depends on p and α, we write $i = i(p,\alpha)$. Thus,

$$p \in B_{i(p,\alpha)} \subset U_\alpha.$$

Now let J be the set of all indices $j \in I$ such that $j = i(p,\alpha)$ for some p and some α. Then $\bigcup_{j\in J} B_j = S$ because every p in S is contained in some $B_{i(p,\alpha)} = B_j$.

For each $j \in J$, choose an $\alpha(j)$ such that $B_j \subset U_{\alpha(j)}$. Then $S = \bigcup_j B_j \subset \bigcup_j U_{\alpha(j)}$. So $\{U_{\alpha(j)}\}_{j\in J}$ is a countable subcover of $\{U_\alpha\}_{\alpha\in\mathrm{A}}$. $\Diamond$

A.15* Disconnected subset in terms of a separation
($\Rightarrow$) By (iii),
$$A = (U \cap V) \cap A = (U \cap A) \cup (V \cap A).$$
By (i) and (ii), $U \cap A$ and $V \cap A$ are disjoint nonempty open subsets of A. Hence, A is disconnected.

($\Leftarrow$) Suppose A is disconnected in the subspace topology. Then $A = U' \cup V'$, where U' and V' are two disjoint nonempty open subsets of A. By the definition of the subspace topology, $U' = U \cap A$ and $V' = V \cap A$ for some open sets U, V in S.

(i) holds because U' and V' are nonempty.
(ii) holds because U' and V' are disjoint.
(iii) holds because $A = U' \cup V' \subset U \cup V$. ♢

A.19* Uniqueness of the limit
Suppose $p \neq q$. Since S is Hausdorff, there exist disjoint open sets U_p and U_q such that $p \in U_p$ and $q \in U_q$. By the definition of convergence, there are integers N_p and N_q such that for all $i \geq N_p$, $x_i \in U_p$ and for all $i \geq N_q$, $x_i \in U_q$. This is a contradiction since $U_p \cap U_q$ is the empty set. ♢

A.20* Closure in a product
($\subset$) By Problem A.5, $\text{cl}(A) \times Y$ is a closed set containing $A \times Y$. By the definition of closure, $\text{cl}(A \times Y) \subset \text{cl}(A) \times Y$.

($\supset$) Conversely, suppose $(p,y) \in \text{cl}(A) \times Y$. If $p \in A$, then $(p,y) \in A \times Y \subset \text{cl}(A \times Y)$. Suppose $p \notin A$. By Proposition A.50, p is an accumulation of A. Let $U \times V$ be any basis open set in $S \times Y$ containing (p,y). Because $p \in \text{ac}(A)$, the open set U contains a point $a \in A$ with $a \neq p$. So $U \times V$ contains the point $(a,y) \in A \times Y$ with $(a,y) \neq (p,y)$. This proves that (p,y) is an accumulation point of $A \times Y$. By Proposition A.50 again, $(p,y) \in \text{ac}(A \times Y) \subset \text{cl}(A \times Y)$. This proves that $\text{cl}(A) \times Y \subset \text{cl}(A \times Y)$. ♢

B.1* The rank of a matrix
($\Rightarrow$) Suppose $\text{rk}A \geq k$. Then one can find k linearly independent columns, which we call a_1, $\dots$, a_k. Since the $m \times k$ matrix $[a_1 \ \cdots \ a_k]$ has rank k, it has k linearly independent rows b^1, $\dots$, b^k. The matrix B whose rows are b^1, $\dots$, b^k is a $k \times k$ submatrix of A, and $\text{rk}B = k$. In other words, B is nonsingular $k \times k$ submatrix of A.

($\Leftarrow$) Suppose A has a nonsingular $k \times k$ submatrix B. Let a_1, $\dots$, a_k be the columns of A such that the submatrix $[a_1 \ \cdots \ a_k]$ contains B. Since $[a_1 \ \cdots \ a_k]$ has k linearly independent rows, it also has k linearly independent columns. Thus, $\text{rk}A \geq k$. ♢

B.2* Matrices of rank at most r
Let A be an $m \times n$ matrix. By Problem B.1, $\text{rk}A \leq r$ if and only if all $(r+1) \times (r+1)$ minors $m_1(A), \dots, m_s(A)$ of A vanish. As the common zero set of a collection of continuous functions, D_r is closed in $\mathbb{R}^{m \times n}$. ♢

B.3* Maximal rank
For definiteness, suppose $n \leq m$. Then the maximal rank is n and every matrix $A \in \mathbb{R}^{m \times n}$ has rank $\leq n$. Thus,
$$D_{\max} = \{A \in \mathbb{R}^{m \times n} \mid \text{rk}A = n\} = \mathbb{R}^{m \times n} - D_{n-1}.$$
Since D_{n-1} is a closed subset of $\mathbb{R}^{m \times n}$ (Problem B.2), $D_{\max}$ is open in $\mathbb{R}^{m \times n}$. ♢

B.4* Degeneracy loci and maximal rank locus of a map

(a) Let D_r be the subset of $\mathbb{R}^{m\times n}$ consisting of matrices of rank at most r. The degeneracy locus of rank r of the map $F\colon S\to\mathbb{R}^{m\times n}$ may be described as

$$D_r(F)=\{x\in S \mid F(x)\in D_r\}=F^{-1}(D_r).$$

Since D_r is a closed subset of $\mathbb{R}^{m\times n}$ (Problem B.2) and F is continuous, $F^{-1}(D_r)$ is a closed subset of S.

(b) Let $D_{\max}$ be the subset of $\mathbb{R}^{m\times n}$ consisting of all matrices of maximal rank. Then $D_{\max}(F)=F^{-1}(D_{\max})$. Since $D_{\max}$ is open in $\mathbb{R}^{m\times n}$ (Problem B.3) and F is continuous, $F^{-1}(D_{\max})$ is open in S. ◊

B.7 Use Example B.5.