SMOOTH MANIFOLDS

IVO TEREK

The current text grew out of my lecture notes when teaching a Differential Topology course at Williams College, during the Fall 2024 term. The name "Differential Topology" here is a bad misnomer—it was a short introductory course on smooth manifolds (not assuming even point-set topology as a pre-requisite), spanning barely 12 weeks, and not covering advanced topics such as characteristic classes, Chern-Weil theory, Morse theory, or cobordism. To make matters worse, halfway through the the semester I had an accident and broke my dominant hand, which inevitably slowed down the course even more. All of this is to say that these notes are not meant to be comprehensive, nor make any effort to be so. That being said, I have done my best to include at least the material I would have liked to cover under normal circumstances. The homework problems assigned in that course will be scattered here as exercises, at the places and moments where I believe the reader would maximize their benefit from working them out. Problems I would have liked to assign but never did are included as well; there is a grand total of 122 exercises (and 71 figures). The sections are not divided by lectures. I roughly followed the order of the presentation in [23] in this course, but would often explain and prove things my own way.

Some mathematical maturity may be required of the reader, but these notes will be hopefully adequate for self-study (perhaps as a workbook). It also seems appropriate to me to thank the half-a-dozen students who did take or audit that course—their questions and observations were certainly extremely helpful not only when preparing these lectures, but also allowed me to think of new ways (at least to me) of explaining more subtle aspects of the theory.

Williamstown, October 2024

it3@williams.edu
Last updated: May 2, 2025.

Some relevant basic notation

- $\varnothing$ denotes the empty set.
- $A \subseteq B$ means that the set A is contained in the set B (that is, A is a subset of B), and the equality case $A = B$ is allowed; strict inclusions would be denoted (if at all) by $\subset$ or $\subsetneq$.
- For a function $f\colon X \to Y$ between sets, and any subset $B \subseteq Y$, we denote by $f^{-1}(B) = \{x \in X : f(x) \in B\}$ the inverse image of B under f. When $B = \{b\}$ is a singleton (i.e., has only one element), $f^{-1}(B)$ is written simply as $f^{-1}(b)$ instead of the more technically correct $f^{-1}(\{b\})$, even when f is not an invertible function; these are called the fibers of f.
- If X is any set and $\mathscr{C}$ is a collection of subsets of X, we write

  $$\bigcup \mathscr{C} = \bigcup_{C \in \mathscr{C}} C = \{x \in X : \text{ there is } C \in \mathscr{C} \text{ such that } x \in C\}$$

  for the union of the collection $\mathscr{C}$. And similarly for the intersection

  $$\bigcap \mathscr{C} = \bigcap_{C \in \mathscr{C}} C = \{x \in X : \text{ for every } C \in \mathscr{C} \text{ it holds that } x \in C\}$$

  of $\mathscr{C}$.
- The set-difference of two sets A and B is defined as $A \smallsetminus B = \{a \in A : a \notin B\}$.
- If X is any set, the identity function of X is denoted by $\mathrm{Id}_X\colon X \to X$, and it is given by $\mathrm{Id}_X(x) = x$ for every $x \in X$.
- If $f\colon X \to Y$ is any function between sets, $\mathrm{Gr}(f) = \{(x, y) \in X \times Y : y = f(x)\}$ is the graph of f.
- $\mathbb{R}^{n\times m}$ denotes the vector space of all size $n \times m$ matrices with real entries. It is of course isomorphic to $\mathbb{R}^{nm}$, but we reserve the former notation for when treating its elements as matrices.

Contents

List of Figures

1 A crash course in general topology

1.1 Introduction

Before understanding *differential topology*, let's first understand *topology*. Consider the definition of continuity for a function: $f\colon \mathbb{R} \to \mathbb{R}$ at a point $a \in \mathbb{R}$:

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \,:\, \forall\, x \in \mathbb{R},\ |x-a| < \delta \implies |f(x) - f(a)| < \varepsilon. \tag{1.1}$$

What is (1.1) effectively saying? The distance between $f(x)$ and $f(a)$ gets smaller as the distance between x and a gets smaller. So it is the *distance* that matters here, and not the algebraic structure of $\mathbb{R}$!

Definition 1 (Metric space)

A **metric space** is a pair (X, d), where X is a set and $\mathrm{d}\colon X \times X \to [0, \infty)$ is a **distance function** on X, i.e, d satisfies the following properties, for all $x, y, z \in X$:

(i) $\mathrm{d}(x,y) = 0 \iff x = y$,

(ii) $\mathrm{d}(x,y) = \mathrm{d}(y,x)$,

(iii) $\mathrm{d}(x,y) \leq \mathrm{d}(x,z) + \mathrm{d}(z,y)$.

Condition (iii) is called the **triangle inequality** because it says that the length of a side of a triangle is less or equal to the sum of the lengths of the other two sides. The definition of continuity for a function $f\colon X \to Y$ between metric spaces (X, d_X) and (Y, d_Y), at a point $a \in X$, should read:

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \,:\, \forall\, x \in X,\ \mathrm{d}_X(x,a) < \delta \implies \mathrm{d}_Y(f(x), f(a)) < \varepsilon. \tag{1.2}$$

Of course, $\mathrm{d}\colon \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$ given by $\mathrm{d}(x,y) = \|x-y\|$, where $\|\cdot\|\colon \mathbb{R}^n \to [0,\infty)$ defined by $\|(x_1, \ldots, x_n)\| = (x_1^2 + \cdots + x_n^2)^{1/2}$ is the standard Euclidean norm, makes $(\mathbb{R}^n, \mathrm{d})$ a metric space. For $(X, \mathrm{d}) = (Y, \mathrm{d}) = (\mathbb{R}^n, \mathrm{d})$ and $n = 1$, (1.2) reduces to (1.1).

In metric spaces, just like in the real line, it makes sense to talk about open sets. We just need to replace open intervals with something appropriate.

Definition 2 (Open balls and open sets in metric spaces)

Let (X, d) be a metric space.

- The **open ball** centered at $a \in X$ with radius $r > 0$ is defined as the set $B_r(a) = \{x \in X : \mathrm{d}(x,a) < r\}$. It replaces the open interval $(a-r, a+r)$ in the real line. The **closed ball** is $B_r[a] = \{x \in X : \mathrm{d}(x,a) \leq r\}$.
- A subset $U \subseteq X$ is called d-**open** if for every point $a \in U$ there is $\varepsilon > 0$ such that $B_\varepsilon(a) \subseteq U$.

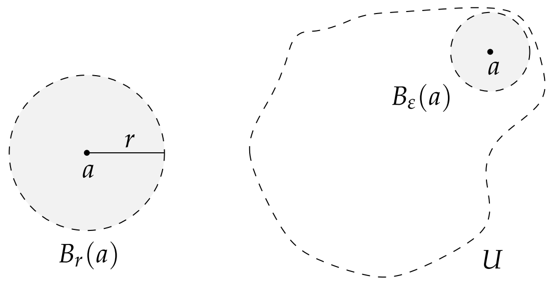


Figure 1: An open ball and a d-open set in a metric space.

Note that open balls $B_r(a)$ are always d-open, or else these definitions would do us no good, but make sure you understand that this does require some proof: the terms "open ball" and "d-open" were defined separately and you should not be fooled by the similar terminology. In any case, the proof consists in noting that whenever $x \in B_r(a)$, we have that $B_{r-\mathrm{d}(x,a)}(x) \subseteq B_r(a)$ as a consequence of the triangle inequality: if $y \in B_{r-\mathrm{d}(x,a)}(x)$, we have that

$$\begin{aligned} \mathrm{d}(y,a) &\leq \mathrm{d}(y,x) + \mathrm{d}(x,a) \\ &< r - \mathrm{d}(x,a) + \mathrm{d}(x,a) \\ &= r, \end{aligned}$$

as claimed.

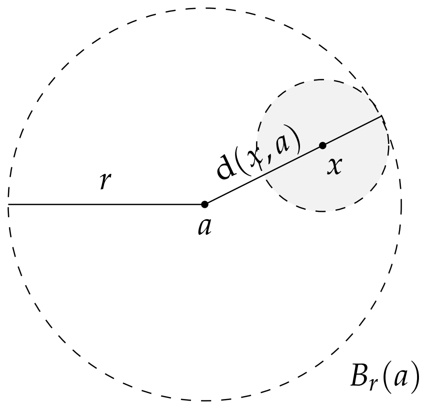


Figure 2: Illustrating the set-inclusion $B_{r-\mathrm{d}(x,a)}(x) \subseteq B_r(a)$.

The collection τ_d of all d-open subsets of X satisfies the following properties:

(i) $\varnothing$ and X are d-open.

(ii) arbitrary unions of d-open sets are d-open. (1.3)

(iii) finite intersections of d-open sets are d-open.

Exercise 1

Check that (1.3) holds.

The definition of continuity can be completely rephrased in terms of d-open sets.

Proposition 1

For a function $f\colon X \to Y$ between metric spaces (X, d_X) and (Y, d_Y), the following conditions are equivalent:

(a) f is continuous (at all points), in the sense of (1.2).

(b) For every $U \subseteq Y$ which is d_Y-open, the inverse image $f^{-1}(U) \subseteq X$ is d_X-open.

Proof:

- First we check that (a) implies (b). Let $a \in f^{-1}(U)$, so that $f(a) \in U$. As U is d_Y-open, there is $\varepsilon > 0$ such that $B_\varepsilon(f(a)) \subseteq U$. By continuity of f at a, there is $\delta > 0$ such that, for any $x \in X$, $\mathrm{d}_X(x,a) < \delta \implies \mathrm{d}_Y(f(x), f(a)) < \varepsilon$. This means that $x \in B_\delta(a) \implies f(x) \in B_\varepsilon(f(a)) \subseteq U$. But $x \in B_\delta(a) \implies f(x) \in U$ is equivalent to $B_\delta(a) \subseteq f^{-1}(U)$, and thus $f^{-1}(U)$ is d_X-open.

- Now we check that (b) implies (a). Let $a \in X$ and $\varepsilon > 0$. Then $B_\varepsilon(f(a)) \subseteq Y$ is d_Y-open. By assumption, $f^{-1}(B_\varepsilon(f(a))) \subseteq X$ is d_X-open, and it clearly contains a. Then there is $\delta > 0$ such that $B_\delta(a) \subseteq f^{-1}(B_\varepsilon(f(a)))$. This means that, for any $x \in X$, $\mathrm{d}_X(x,a) < \delta \implies \mathrm{d}_Y(f(x), f(a)) < \varepsilon$, as required. Since $a \in X$ was arbitrary, f is continuous and we are done. $\square$

The above equivalence means that we don't really need even distances to make sense of continuity, as long as we have a collection of "open sets" with the correct properties. With this in mind, we axiomatize (1.3):

Definition 3 (Topology)

A **topology** on a set X is a collection τ of subsets of X such that:

(i) $\varnothing, X \in \tau$.

(ii) τ is closed under arbitrary unions.

(iii) τ is closed under finite intersections.

The elements of τ are then called **open sets**, and the pair (X, τ) is called a **topological space**.

With this in place, we may turn item (b) of Proposition 1 into a definition:

Definition 4 (Continuity)

A function $f\colon X \to Y$ between topological spaces (X, τ) and (Y, τ') is **continuous** if whenever $U \in \tau'$, we have $f^{-1}(U) \in \tau$. In words, if inverse images of open sets are open. In addition, f is called a **homeomorphism** if it is continuous, bijective, and its inverse is also continuous.

Example 1 (Metric spaces are topological spaces, but not conversely)

If (X, d) is a metric space, the collection τ_{d} of all d-open sets is a topology on X, and so (X, τ_{d}) is a topological space. As we will eventually see, there are topologies who are not of the form τ_{d} for any distance function d—the ones who are are called **metrizable**, and have some nicer properties that general topologies might not have.

Here is the punchline:

> A topology is the "minimum" amount of structure needed on a set to be able to talk about continuity. *Differential* topology is about the minimum amount of structure needed to talk about *differentiability*.

This will ultimately lead us to the notion of a smooth manifold, which we present in Section 3: they are a special type of topological space, "locally modeled" on Euclidean space, on which we can develop Calculus.

1.2 More examples of topological spaces and continuous functions

Let's start with some less obvious examples of topological spaces and continuous functions, so you can get a feeling for what Definitions 3 and 4 allow. Keep in mind that a topology can be seen as a way of capturing "nearness" without relying on distances: two points $x, y \in X$ can be considered morally "close to each other" if there is an open subset U of X with $x, y \in U$, and the smaller U is, the closer x and y are—for example, taking $U = X$ is always possible in view of condition (i) in Definition 3, but this doesn't tell us much. If you don't find this interpretation helpful, don't think too hard about it (at this stage this is just heuristics anyway)

Example 2 (The extreme cases)

If X is any set, the **discrete topology** on X is $\tau_{\text{disc.}} = 2^X$ (that is, all subsets[a] of X are declared to be open). The **chaotic topology** on X, in turn, is $\tau_{\text{ch.}} = \{\varnothing, X\}$ (the reason for this name will be justified in Example 23). Conditions (i)-(iii) in Definition 3 are easily verified to hold for both $\tau_{\text{disc.}}$ and $\tau_{\text{ch.}}$. More interestingly, if (Y, τ') is a second topological space, then:

- every function $f\colon (X, \tau_{\text{disc.}}) \to (Y, \tau')$ is continuous, since $f^{-1}(U) \in \tau_{\text{disc.}}$ no matter what $U \in \tau'$ is, and;
- every function $f\colon (Y, \tau') \to (X, \tau_{\text{ch.}})$ is continuous, since only $f^{-1}(\varnothing) = \varnothing$ and $f^{-1}(X) = Y$ need to be tested, and they are both in τ'.

Note that, for the discrete topology $\tau_{\text{disc.}}$, all singletons (i.e., sets with a single element) are open. And in fact, this is enough: if you have a topology for which all singletons are open, then this topology necessarily equals $\tau_{\text{disc.}}$: every set is the union of the singletons of its points, while arbitrary unions of open sets are open, cf. condition (ii) in Definition 3.

[a]Formally, A^B denotes the set of all functions $B \to A$. As $2 = \{0, 1\}$ (set-theoretically), 2^X denotes the set of all functions $X \to \{0, 1\}$. Then 2^X is in bijective correspondence with the collection of all subsets of X, justifying the notation. Namely, to each subset $A \subseteq X$ we assign its **characteristic function** $\chi_A \in 2^X$, defined by $\chi_A(x) = 1$ if $x \in A$, and $\chi_A(x) = 0$ if $x \notin A$. The inverse of the assignment $A \mapsto \chi_A$ takes $f \in 2^X$ to $f^{-1}(1) \subseteq X$.

Example 3 (The Sierpinski space)

Let $X = \{a, b\}$ be a set with two elements. Apart from the topologies $\tau_{\text{disc.}}$ and $\tau_{\text{ch.}}$ on X, there are just two more topologies on X, which are "equivalent" under relabeling $a \leftrightarrow b$. So let $\tau = \{\varnothing, \{a\}, \{a, b\}\}$ be one of them, cf. Figure 3.

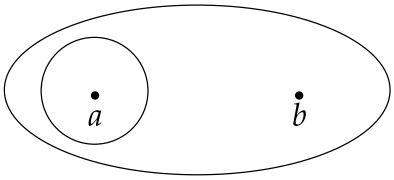


Figure 3: The Sierpinski topology on a two-point set.

To check that τ is indeed a topology, we need to check that conditions (i)-(iii) in Definition 3 are satisfied. For (i), there is nothing to do. For (ii) and (iii), brute force does it: all possible unions and intersections of elements in τ are again in τ.

$\cup$	$\varnothing$	$\{a\}$	$\{a,b\}$
$\varnothing$		$\{a\}$	$\{a,b\}$
$\{a\}$			$\{a,b\}$
$\{a,b\}$			

$\cap$	$\varnothing$	$\{a\}$	$\{a,b\}$
$\varnothing$		$\varnothing$	$\varnothing$
$\{a\}$			$\{a\}$
$\{a,b\}$			

(1.4)

Note that every open subset of X containing b also contains a. This means that the topology τ doesn't really distinguish the poins a and b, and it ought to be considered something "pathological". We will make the concept of "distinguishing two points" more precise in Section 1.7.

Finally, if (Y, τ') is another topological space, a function $f : (Y, \tau') \to (X, \tau)$ is continuous if and only if $f^{-1}(a) \in \tau'$, since there are no other nontrivial inverse images to check for.

The above situation illustrates a little useful trick: when X (or more generally, τ) is finite, to show that τ satisfies (ii) and (iii) in Definition 3, it is enough to check that τ is closed under unions and intersections of *two* sets—induction takes care of the rest. We may also ignore $\varnothing$ and X in this process from the start once we know that they are in τ, since $\varnothing \cup A = X \cap A = A$, $\varnothing \cap A = \varnothing$, and $X \cup A = X$ for every subset $A \subseteq X$. In addition, $A \cup A = A \cap A = A$ for every $A \subseteq X$, so there is no need to address these cases. In other words, whenever you want to organize your work using tables such as the ones in (1.4), *it suffices to fill in the entries strictly above the diagonal*. To better appreciate this shortcut, you should work out a concrete example yourself:

Exercise 2

Show that the collection

$$\tau = \{\varnothing, X, \{a\}, \{c,d\}, \{a,c,d\}, \{b,c,d,e\}\}$$

of subsets of $X = \{a,b,c,d,e\}$ is a topology on X. Represent the resulting topological space (X,τ) with a picture, as we have done in Example 3.

Fun fact: the number of topologies on a set with n elements, where n ranges from 1 to 7, is $1, 4, 29, 355, 6942, 209527, 9535241$ [11, p. 43]. (https://oeis.org/A000798)

Not-so-fun fact: it doesn't look like it is known how many of such possible topologies have particularly nice or specific properties (that we'll encounter later, such as metrizability, Hausdorffness, connectedness, etc.).

Often, the sets in a topology are not given explictly like in the situations above, but instead are given through some rule.

Example 4 (Particular-point topology)

Let X be any nonempty set, and fix an element $z \in X$. Then, we claim that the collection $\tau_z = \{\varnothing\} \cup \{U \subseteq X : z \in U\}$ is a topology on X. Condition (i) in Definition 3 holds because $\varnothing \in \tau_z$ by definition of τ_z, and $X \in \tau_z$ because $z \in X$. Next, condition (ii) is satisfied because any union of subsets of X containing z also contains z (and clearly, just one of them containing z would be enough). Finally, condition (iii) is verified because any intersection (not just the finite ones) of subsets of X containing z again contains z.

Now, let Y be a second nonempty set equipped with the particular-point topology τ_w built using an element $w \in Y$. We conclude by observing that a function $f\colon (X,\tau_z) \to (Y,\tau_w)$ is continuous if and only if $f(z) = w$. Indeed, if f is continuous, then $\{w\} \in \tau_w$ implies that $f^{-1}(w) \in \tau_z$, so that $z \in f^{-1}(w)$ and finally $f(z) = w$. And conversely, if $f(z) = w$, we argue that f is continuous: for $U \in \tau_w \smallsetminus \{\varnothing\}$ we have that $w \in U$, and so $z \in f^{-1}(U)$ because $f(z) = w \in U$, leading to $f^{-1}(U) \in \tau_z$, as required.

Exercise 3

Let X be any set, and τ_1 and τ_2 be two topologies on X. True or false? Prove or give a counter-example:

(a) $\tau_1 \cup \tau_2$ is also a topology on X.

(b) $\tau_1 \cap \tau_2$ is also a topology on X.

The next situation is often presented as one of the first examples of topological spaces, beyond metric spaces (that is, that is not metrizable):

Exercise 4 (Cofinite and cocountable topologies)

Let X be any infinite set, and let $\tau_{\text{cof.}} = \{\varnothing\} \cup \{U \subseteq X \mid X \smallsetminus U \text{ is finite}\}$.

(a) Show that $\tau_{\text{cof.}}$ is a topology on X. It is called the *cofinite topology on* X.

Similarly, if X is uncountable, $\tau_{\text{coco.}} = \{\varnothing\} \cup \{U \subseteq X \mid X \smallsetminus U \text{ is countable}\}$ is a topology, called the *cocountable topology on* X. Assume that this is the case. Meanwhile, we also have the discrete and chaotic topologies on X, $\tau_{\text{disc.}}$ and $\tau_{\text{ch.}}$.

(b) Sort these four topologies according to inclusion: $\ldots \subseteq \ldots \subseteq \ldots \subseteq \ldots$.

Warning: it is not true in general that two different topologies on a same set are necessarily comparable via inclusion.

Example 5 (Euclidean topologies on abstract vector spaces)

Let V be a finite-dimensional real vector space, and $T\colon V \to \mathbb{R}^n$ be a linear isomorphism. If we define a subset $U \subseteq V$ to be open if the image $T(U)$ is an open subset of $\mathbb{R}^n$ (equipped with its standard Euclidean topology), it is easy to check that we obtain a topology on V. What is remarkable here is that

$$\text{the resulting topology on } V \text{ does not depend on the choice of } T. \tag{1.5}$$

Indeed, if $S\colon V \to \mathbb{R}^n$ is a second isomorphism, then $T \circ S^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$ is a homeomorphism (since both $T \circ S^{-1}$ and its inverse $S \circ T^{-1}$ are continuous, being linear), and so $T(U) \subseteq \mathbb{R}^n$ is open if and only if $S(U) \subseteq \mathbb{R}^n$ is open, as a consequence of the obvious relation $S(U) = (S \circ T^{-1})(T(U))$.

We call the topology in (1.5) the **Euclidean topology of** V. By design, every linear isomorphism between V and $\mathbb{R}^n$ becomes a homeomorphism, so V is topologically indistinguishable from $\mathbb{R}^n$ (as it should be). However, we have now freed ourselves from the constraint dictating that elements of V should be n-tuples of real numbers; the ultimate gain here is one of abstraction.

1.3 Subspaces

Relevant to the next part of the discussion is the next easy result, which we have neglected to mention so far:

Lemma 1

The composition of continuous functions between topological spaces is also continuous.

Proof: Omitting the topologies from the notation (for the first time in many to come), let $f\colon X \to Y$ and $g\colon Y \to Z$ be continuous. Then, whenever $U \subseteq Z$ is open, the inverse image $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U)) \subseteq X$ is open, being the inverse image under the continuous function f of $g^{-1}(U) \subseteq Y$ (itself open by continuity of g and openness of U). $\square$

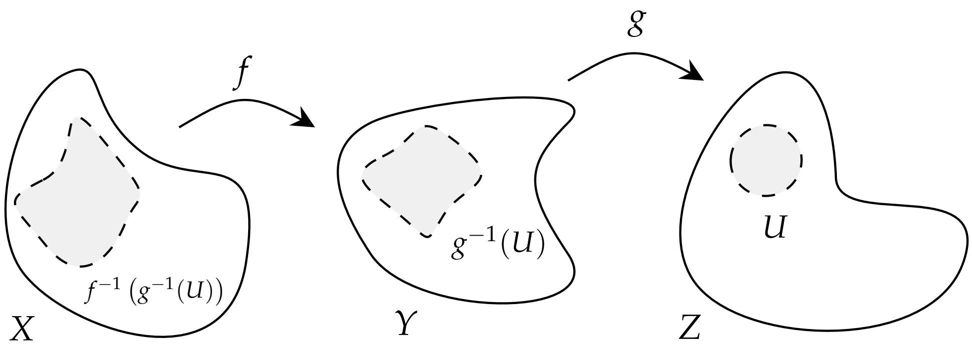


Figure 4: Continuity of the composition of continuous functions.

Our main result here is:

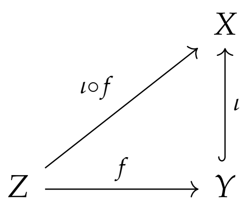
Proposition 2 (Setting up the subspace topology, and its characteristic property)

Let (X, τ) be a topological space, and $Y \subseteq X$ be any subset. Then the collection

$$\begin{aligned} \tau_Y &= \{V \subseteq Y : \text{there is } U \in \tau \text{ such that } V = U \cap Y\} \\ &= \{U \cap Y : U \in \tau\} \end{aligned} \tag{1.6}$$

is a topology on Y, and the inclusion mapping $\iota\colon Y \hookrightarrow X$ is continuous.

In addition, if (Z, τ') is a third topological space and $f\colon Z \to Y$ is any function, then f is continuous if and only if the composition $\iota \circ f\colon Z \to X$ is continuous, cf. the next diagram. That is, τ_Y allows for codomains to be restricted when discussing continuity.



Proof: Once more we return to Definition 3. For condition (i), note that $\varnothing = \varnothing \cap Y$ and $Y = X \cap Y$ with $\varnothing, X \in \tau$, so that $\varnothing, Y \in \tau_Y$. Now, for condition (ii), consider any collection $\{V_\alpha\}_{\alpha \in A} \subseteq \tau_Y$: for each $\alpha \in A$, fix $U_\alpha \in \tau$ with $V_\alpha = U_\alpha \cap Y$; now take unions to obtain $\bigcup_{\alpha \in A} V_\alpha = (\bigcup_{\alpha \in A} U_\alpha) \cap Y$ with $\bigcup_{\alpha \in A} U_\alpha \in \tau$, so that $\bigcup_{\alpha \in A} V_\alpha \in \tau_Y$. Finally, if $V_1, V_2 \in \tau_Y$ are written as $V_1 = U_1 \cap Y$ and $V_2 = U_2 \cap Y$ for some $U_1, U_2 \in \tau$, we have that $V_1 \cap V_2 = (U_1 \cap U_2) \cap Y$ with $U_1 \cap U_2 \in \tau$, showing that $V_1 \cap V_2 \in \tau_Y$ and establishing (iii). This proves that τ_Y is a topology on Y.

It remains to verify the claims about the inclusion mapping. For the first, it suffices to note that for every subset $U \subseteq X$ we have that $\iota^{-1}(U) = U \cap Y$, so that $U \in \tau$ implies that $\iota^{-1}(U) \in \tau_Y$. Now, let (Z, τ') and f be as in the result statement. If f is continuous, then $\iota \circ f$ is continuous, being the composition of continuous functions (cf. Lemma 1). On the other hand, assuming that $\iota \circ f$ is continuous, we show that f is continuous: if $V \in \tau_Y$ is written as $V = U \cap Y$ for some $U \in \tau$, we have that $f^{-1}(V) = (\iota \circ f)^{-1}(U) \in \tau'$ by continuity of $\iota \circ f$, as required. $\square$

See Figure 5:

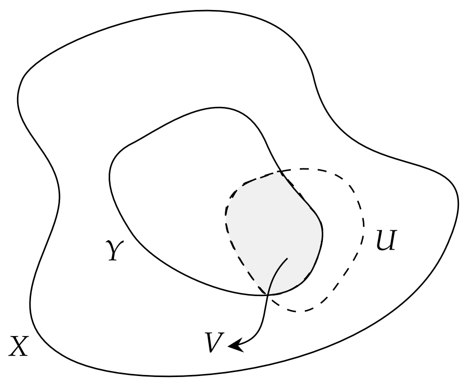


Figure 5: Definition of subspace topology.

We are now allowed to write the next definition:

Definition 5 (Subspace topology)

Let (X, τ) be a topological space, and $Y \subseteq X$ be any subset. The topology τ_Y defined in (1.6) is called the **subspace topology in** Y **induced by** X. If nothing else is said, this is the topology taken on Y by default.

Example 6 (Euclidean topologies on subsets of abstract vector spaces)

If V is a finite-dimensional real vector space, equipped with its Euclidean topology (Example 5), and $Y \subseteq V$ is any subset, the subspace topology on Y induced by V is called the **Euclidean topology** of Y.

The next exercises, aimed at the readers more concerned with subtler points of the theory, have long statements but consist merely on unwinding the relevant definitions.

Exercise 5 (Paranoia or consistency?)

Let (X, τ) be a topological space, and $Y, Z \subseteq X$ two subsets such that $Z \subseteq Y$. On Z, we have two possible topologies: the first one is τ_Z, the subspace topology in Z induced by X, and the second one is $(\tau_Y)_Z$, the subspace topology in Z induced by Y (which is itself equipped with its subspace topology induced by X).

Show that $\tau_Z = (\tau_Y)_Z$.

Exercise 6 (Paranoia or consistency? 2: electric boogaloo)

Let (X, d) be a metric space, and $Y \subseteq X$ be any subset. Write d' for the restriction of d to Y, so that (Y, d') is a metric space on its own right. As we have seen in Example 1, there is a topology τ_{d} on X, and a topology $\tau_{\mathrm{d}'}$ on Y. On the other hand, there is a subspace topology $(\tau_{\mathrm{d}})_Y$ on Y.

Show that $(\tau_{\mathrm{d}})_Y = \tau_{\mathrm{d}'}$.

Exercise 7 (Equivalent metrics induce the same topology)

Let X be any set, and let $\mathrm{d}, \mathrm{d}' \colon X \times X \to [0.\infty)$ be two distance functions on X. Assume that d and d' are **Lipschitz-equivalent**, that is, there are constants $a, b > 0$ such that $a\mathrm{d}(x,y) \leq \mathrm{d}'(x,y) \leq b\mathrm{d}(x,y)$ for all $x, y \in X$. Show that $\tau_{\mathrm{d}} = \tau_{\mathrm{d}'}$.

1.4 Bases for topological spaces, and product spaces

We start with one more very instructive example of a topological space:

Exercise 8 (The Sorgenfrey line)

In the real line $\mathbb{R}$, say that a subset $U \subseteq \mathbb{R}$ is *Sorgenfrey-open* if for every $x \in U$ there is $\varepsilon > 0$ such that $[x, x+\varepsilon) \subseteq U$.

(a) Show that the collection τ_S of all Sorgenfrey-open sets is a topology on $\mathbb{R}$.

(b) Exhibit a subset of $\mathbb{R}$ which is Sorgenfrey-open, but not open in the standard Euclidean topology of $\mathbb{R}$ (and prove that this is the case).

(c) Verify in detail that a function $f \colon (\mathbb{R}, \tau_S) \to (\mathbb{R}, \tau_S)$ is continuous if and only if whenever $a \in \mathbb{R}$, for every $\varepsilon > 0$ there is $\delta > 0$ such that for any $x \in \mathbb{R}$, $a \leq x < a + \delta$ implies that $f(a) \leq f(x) < f(a) + \varepsilon$. Compare it with (1.1).

This τ_S is called the Sorgenfrey topology on $\mathbb{R}$, and it serves as a counter-example to many false statements you may come across when studying general topology.

As you may have already noticed, it is often convenient to define topologies in terms of a "special" collection of subsets (who become open themselves, *ex post facto*): open balls in metric spaces, the half-intervals in the Sorgenfrey line, or even singletons for the discrete topology!

What properties are common to open balls and metric spaces, that ultimately allow us to check that the collections of d-open sets and of Sorgenfrey-open sets do satisfy conditions (i)-(iii) of Definition 3, and hence are topologies?

How to make sense of what should be the "building blocks" for a topology?

Definition 6 (Bases for topological spaces)

A **basis** for a topological space (X, τ) is a subcollection $\mathscr{B} \subseteq \tau$ (i.e., $\mathscr{B}$ consists of open subsets of X) with the following property: whenever $U \in \tau$ and $x \in U$, there is an element $B \in \mathscr{B}$ such that $x \in B \subseteq U$.

See Figure 6. In Linear Algebra, the fact that every vector space has a basis is not quite trivial and requires the use of Zorn's Lemma. In general topology, this is not an issue: if worst comes to worst, we can always take $\mathscr{B} = \tau$ and call it a day. Of course, this is not only not helpful, but also defeats the purpose of introducing bases to begin with. In other words, the smaller the basis we can find, the better. Hoping for finite bases is obviously wishful thinking here, but countability is the next best thing. For this reason we introduce the following notion:

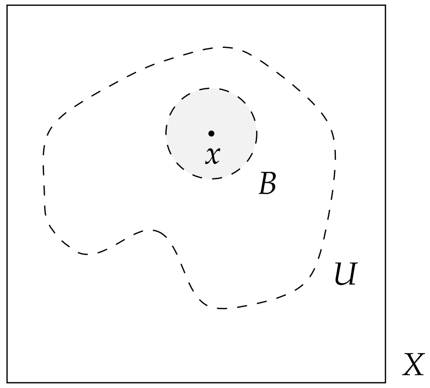


Figure 6: A basic open neighborhood B of x in X, contained in U.

Definition 7 (Second-countable spaces)

A topological space is **second-countable** if it admits a countable basis.

Remark. The above definition begs for an obvious question: if we're talking about *second*-countable spaces, what would *first*-countable spaces be? There is the concept of a **local basis at a point** a in a topological space (X, τ)—it is a collection $\mathscr{B}_a \subseteq \tau$ with the properties (i) $a \in B$ for all $B \in \mathscr{B}_a$, and (ii) whenever $U \in \tau$ has $a \in U$, there exists $B \in \mathscr{B}_a$ such that $B \subseteq U$. Then (X, τ) is called **first-countable** if at each point of X there is a countable local basis. Note that $\bigcup_{a \in X} \mathscr{B}_a$ is a basis for (X, τ) (in the sense of the definition above), but this union is in general uncountable.

In any case, here are some examples:

Example 7 (Second-countability of the standard real line)

The real line with its standard Euclidean topology is second-countable: the collection $\mathscr{B} = \{(a - \varepsilon, a + \varepsilon) : a \in \mathbb{Q} \text{ and } \varepsilon \in (0, \infty) \cap \mathbb{Q}\}$ is a basis, countable since $\mathbb{Q}$ is countable. Indeed, let $U \subseteq \mathbb{R}$ be any open subset, and $x \in U$, so that there is $\varepsilon' > 0$ such that $(x - \varepsilon', x + \varepsilon') \subseteq U$. Now, choosing $\varepsilon \in (0, \varepsilon'/2) \cap \mathbb{Q}$ and $a \in \mathbb{Q}$ such that $|a - x| < \varepsilon$, it follows that $(a - \varepsilon, a + \varepsilon) \subseteq (x - \varepsilon', x + \varepsilon')$, and hence $(a - \varepsilon, a + \varepsilon) \subseteq U$.

Exercise 9 (Finite-dimensional vector spaces are second-countable)

Show that $\mathbb{R}^n$ (and hence any abstract finite-dimensional vector space) with its standard Euclidean topology is second-countable.

Example 8 (Non-uniqueness of bases)

Whenever (X, d) is a metric space, the collection $\mathscr{B}_{\mathrm{O}} = \{B_r(p) : p \in X \text{ and } r > 0\}$ is a basis for (X, τ_{d})—this is a direct consequence of the very definition of d-open sets (Definition 2). This, in general, is not the only possibility. For the plane $X = \mathbb{R}^2$ equipped with its standard Euclidean (metric) topology, we could take the basis consisting of all bounded open rectangles, or all open triangles, or all open kites, to name a few:

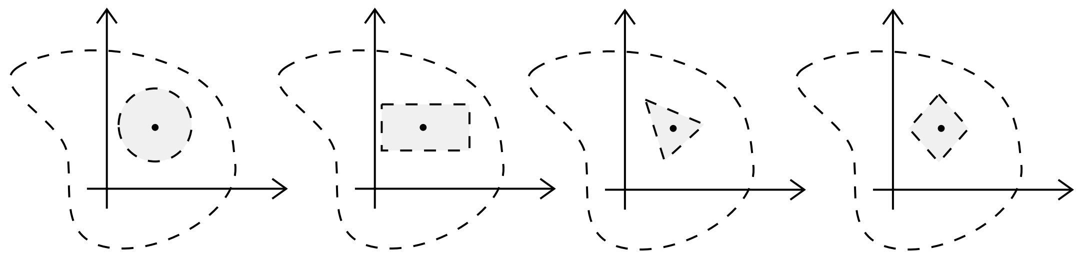


Figure 7: Several bases for the standard Euclidean topology in $\mathbb{R}^2$.

Such bases are "equivalent" because every open ball contains an open rectangle (or triangle, or kite), and vice-versa.

Example 9 (The optimal basis for a discrete space)

Let $(X, \tau_{\mathrm{disc.}})$ be a discrete topological space. Then $\{\{a\} : a \in X\}$ is a basis for $(X, \tau_{\mathrm{disc.}})$, since all subsets of X are open and whenever $U \subseteq X$ has $a \in U$, we necessarily have $a \in \{a\} \subseteq U$. In fact, it is not hard to see that this is the smallest possible basis, so that $(X, \tau_{\mathrm{disc.}})$ is second-countable if and only if X is countable.

Example 10 (The Sorgenfrey line is not second-countable)

Consider the Sorgenfrey line $(\mathbb{R}, \tau_S)$ form Exercise 8: one basis for it, by design, is the collection $\{[x, x+r) : x \in \mathbb{R} \text{ and } r > 0\}$. We claim, however, that $(\mathbb{R}, \tau_S)$ is *not* second-countable. The strategy to directly show that some topological space is not second-countable is almost always the same: start with an arbitrary basis for the topology, and produce an injective function from $\mathbb{R}$ to such basis. Since all bases are uncountable, the given space cannot be second-countable. Here's how such an argument would go in this case: let $\mathscr{B}$ be any basis for $(\mathbb{R}, \tau_S)$, and for each $x \in \mathbb{R}$ choose $B_x \in \mathscr{B}$ such that $x \in B_x \subseteq [x, x+1)$. Then note that $\mathbb{R} \ni x \mapsto B_x \in \mathscr{B}$ is injective: if $x, y \in \mathbb{R}$ are such that $x < y$, then $B_x \neq B_y$ because $x \in B_x$ but $x \notin B_y$.

Example 11 (Second-countability is a hereditary property)

Let (X, τ) be a second-countable topological space, and $Y \subseteq X$ be any subset. Then Y, equipped with its subspace topology τ_Y induced from X, is also second-countable. It suffices to note that if $\mathscr{B}$ is a basis for (X, τ), then the collection $\mathscr{B}_Y = \{B \cap Y : B \in \mathscr{B}\}$ is a basis for (Y, τ_Y) (and $\mathscr{B}_Y$ is countable whenever $\mathscr{B}$ is). Indeed, if $V \subseteq Y$ is open and $x \in V$, there is $U \in \tau$ such that $V = U \cap Y$; from $x \in U$ and $\mathscr{B}$ being a basis for (X, τ), there is $B \in \mathscr{B}$ so that $x \in B \subseteq U$, and hence $x \in B \cap Y \subseteq V$.

Back to the comparison with linear algebra: every vector may be uniquely expressed as a linear combination of basis vectors, so every open set should be expressed as the union of basic open sets. Of course, here we must drop the uniqueness requirement, since there's no obvious notion analogue to "linear independence" for open sets (disjointness surely won't be it).

Lemma 2

Let (X, τ) be a topological space, and $\mathscr{B}$ be a basis for (X, τ). Then every open subset of X may be written as a union of elements of $\mathscr{B}$.

Proof: Let $U \subseteq X$ be open and, for each $x \in U$, let $B_x \in \mathscr{B}$ be such that $x \in B_x \subseteq U$. In other words, $\{x\} \subseteq B_x \subseteq U$. Taking the union of everything over all points in U and using that $\bigcup_{x \in U}\{x\} = U$, it follows that $U \subseteq \bigcup_{x \in U} B_x \subseteq U$, and hence $U = \bigcup_{x \in U} B_x$, as required. □

We are ready to state and prove our main result about bases:

Theorem 1 (Creating topologies from proposed bases)

Let X be a set, and $\mathscr{B}$ be a collection of subsets of X. Then $\mathscr{B}$ is a basis for a topology in X if and only if the conditions below hold:

(i) $X = \bigcup \mathscr{B}$, that is, X is the union of all elements in $\mathscr{B}$; and

(ii) Whenever $B_1, B_2 \in \mathscr{B}$ and $x \in B_1 \cap B_2$ are given, there is a third set $B_3 \in \mathscr{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

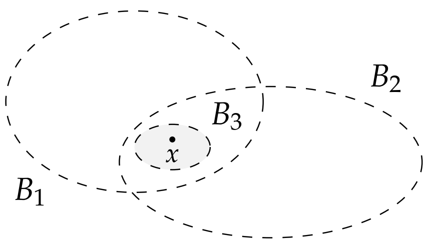


Figure 8: The intersection property of a basis.

Proof: Assume that there exists a topology τ on X for which $\mathscr{B}$ is a basis. Then (i) holds as a consequence of Lemma 2 applied to $U = X$, while (ii) follows from the very definition of basis: if $B_1, B_2 \in \mathscr{B}$, in particular B_1 and B_2 are open, so that $B_1 \cap B_2$ is open as well; if $x \in B_1 \cap B_2$, the definition of basis yields a third set $B_3 \in \mathscr{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$ as required.

Conversely, assume that $\mathscr{B}$ satisfies (i) and (ii), and let $\tau = \{\bigcup \mathscr{C} : \mathscr{C} \subseteq \mathscr{B}\}$. In words, we let τ consists of all possible unions of the sets in $\mathscr{B}$. We claim that τ is a topology on X. First, we have that $\varnothing \in \tau$ because $\varnothing = \bigcup \varnothing$ and $\varnothing \subseteq \mathscr{B}$, while $X \in \tau$ because $X = \bigcup \mathscr{B}$ (by (i)) and $\mathscr{B} \subseteq \mathscr{B}$. Next, we establish closure of τ under unions: if $\{U_i\}_{i \in I} \subseteq \tau$, for some index set I, we may select subsets $\mathscr{C}_i \subseteq \mathscr{B}$ such that $U_i = \bigcup \mathscr{C}_i$, and note that $\bigcup_{i \in I} U_i = \bigcup (\bigcup_{i \in I} \mathscr{C}_i)$ with $\bigcup_{i \in I} \mathscr{C}_i \subseteq \mathscr{B}$, so that $\bigcup_{i \in I} U_i \in \tau$. Finally, assuming that $U_1, U_2 \in \tau$ are given, we select subsets $\mathscr{C}_1, \mathscr{C}_2 \subseteq \mathscr{B}$, and write $U_1 \cap U_2 = \bigcup (\mathscr{C}_1 \cap \mathscr{C}_2)$ with $\mathscr{C}_1 \cap \mathscr{C}_2 \subseteq \mathscr{B}$, showing that $U_1 \cap U_2 \in \tau$. This completes the proof that τ is a topology. Of course, there is one remaining loose end: showing that $\mathscr{B}$ is indeed a basis for τ. If $U \in \tau$ and $x \in U$, we select $\mathscr{C} \subseteq \mathscr{B}$ such that $U = \bigcup \mathscr{C}$, and use the definition of union to obtain $B \in \mathscr{C}$ (hence $B \in \mathscr{B}$) such that $x \in B$; as $B \subseteq \bigcup \mathscr{C} = U$, we are done.

□

Exercise 10

In the setting of the above proof, show that two collections $\mathscr{B}$ and $\mathscr{B}'$ satistying (i) and (ii) give rise to the same topology on X if and only if for every $B \in \mathscr{B}$ and $x \in B$ there is $B' \in \mathscr{B}'$ such that $x \in B' \subseteq B$, and similarly if we switch the roles of $\mathscr{B}$ and $\mathscr{B}'$; compare this with the situation in Example 8.

See next how one applies the above result in practice:

Example 12 (The "vertical" topology)

Let's say that a subset $U \subseteq \mathbb{R}^2$ is **v-open** if for each $(a, b) \in U$ there is $\varepsilon > 0$ such that $\{a\} \times (b - \varepsilon, b + \varepsilon) \subseteq U$. See Figure 9.

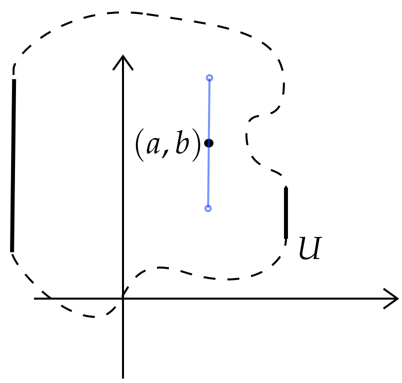


Figure 9: An open subset in the "vertical" topology of the plane.

We claim that the collection $\tau_{\text{vert.}}$ of all v-open sets is a topology on $\mathbb{R}^2$. Instead of proving this directly using the given definition of a v-open set, consider the collection

$$\mathscr{B}_{\text{vert.}} = \{\{x\} \times J : x \in \mathbb{R} \text{ and } I \subseteq \mathbb{R} \text{ is an open interval}\}$$

of all "open vertical segments" in the plane. Clearly $\mathbb{R}^2 = \bigcup_{x\in\mathbb{R}}(\{x\} \times \mathbb{R})$, so that $\mathscr{B}_{\text{vert}}$ saisfies (i). As for condition (ii), note that if $(\{x_1\} \times J_1) \cap (\{x_2\} \times J_2) \neq \varnothing$, so that $(\{x_1\} \cap \{x_2\}) \times (J_1 \cap J_2) \neq \varnothing$, and (x, y) is a point in such intersection, then $x_1 = x_2 = x$ and $y \in J_1 \cap J_2$. Then, $\{x\} \times (J_1 \cap J_2) \in \mathscr{B}_{\text{vert.}}$ (as the intersection of two open intervals is an open interval) has

$$(x, y) \in \{x\} \times (J_1 \cap J_2) \subseteq (\{x_1\} \times J_1) \cap (\{x_2\} \times J_2),$$

as required. By Theorem 1, $\tau_{\text{vert.}}$ is a topology.

Exercise 11

Show that $(\mathbb{R}^2, \tau_{\text{vert.}})$ is *not* second-countable.
Hint: For every $x \in \mathbb{R}$, the vertical line $\{x\} \times \mathbb{R}$ is v-open. Revisit Example 10.

The exercise below contains an example which will be relevant for us later, when we start discussing manifolds.

Exercise 12 (The line with two origins)

Let $X = (\mathbb{R} \smallsetminus \{0\}) \cup \{z_1, z_2\}$, where $z_1, z_2 \notin \mathbb{R}$ are distinct, and consider the collection $\mathscr{B}$ consisting of all open intervals in $\mathbb{R}$ not containing zero, together with all sets of the form $(-a, 0) \cup \{z_i\} \cup (0, a)$, $i = 1, 2$.

(a) Show that $\mathscr{B}$ is a basis for a topology τ on X.

The topological space (X, τ) is called the **line with two origins**. See Figure 10.

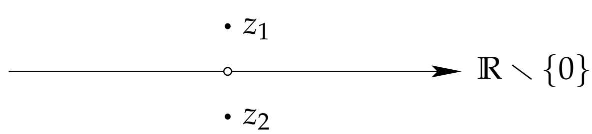


Figure 10: The line with two origins.

Proceeding, show that:

(b) every open set containing z_1 intersects every open set containing z_2.

(c) (X, τ) is second-countable.

Here is one of the reasons bases make our lives easier:

Lemma 3 (Continuity can be tested on a basis)

Let (X, τ) and (Y, τ') be topological spaces, $\mathscr{B}$ be a basis for (Y, τ'), and $f\colon X \to Y$ be any function. If $f^{-1}(B) \in \tau$ for every $B \in \mathscr{B}$, then f is continuous.

Proof: Let $V \in \tau'$ be arbitrary, and consider the inverse image $f^{-1}(V) \subseteq X$. For every point $x \in f^{-1}(V)$ there is $B_x \in \mathscr{B}$ such that $f(x) \in B_x \subseteq V$. We then claim that $f^{-1}(V) = \bigcup_{x\in f^{-1}(V)} f^{-1}(B_x)$ is an union of open sets, and hence open: applying f^{-1} to both sides of $\bigcup_{x\in f^{-1}(V)} B_x \subseteq V$ and using that inverse images distributes over sums yields one inclusion; for the reverse inclusion, just note that if $x \in f^{-1}(V)$, then $f(x) \in B_x$ implies that $x \in f^{-1}(B_x)$. □

With the language of bases, we can introduce another rather important class of examples.

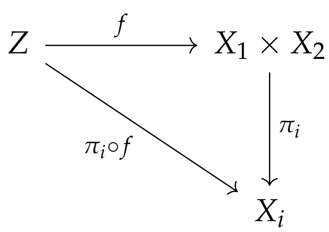
Proposition 3 (Setting up the product topology, and its characteristic property)

Let (X_1, τ_1) and (X_2, τ_2) be two topological spaces. The collection

$$\tau_1 \times \tau_2 = \{U_1 \times U_2 : U_1 \in \tau_1 \text{ and } U_2 \in \tau_2\} \tag{1.7}$$

is a basis for a topology $\tau_1 \otimes \tau_2$ on the cartesian product $X_1 \times X_2$, and both projections $\pi_1\colon X_1 \times X_2 \to X_1$ and $\pi_2\colon X_1 \times X_2 \to X_2$ are continuous.

In addition, if (Z, τ') is a third topological space and $f\colon Z \to X_1 \times X_2$ is any function, then f is continuous if and only if both compositions $\pi_i \circ f\colon Z \to X_i$ are continuous. That is, functions valued in product spaces are continuous if and only if their components are continuous.



Remark. Technically, denoting by $\tau_1 \times \tau_2$ the collection in (1.7) is an abuse of notation; we are not referring to the cartesian product of τ_1 and τ_2. The point here is that $\tau_1 \times \tau_2$ itself, in general, is *not* a topology on $X_1 \times X_2$, instead only being a basis for one.

Proof: We will check that $\tau_1 \times \tau_2$ in (1.7) satisfies both conditions in Theorem 1. Condition (i) is trivially satisfied, as $X \times Y \in \tau_1 \times \tau_2$. Condition (ii), in turn, is immediate from

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

valid for any subsets $U_1, V_1 \subseteq X_1$ and $U_2, V_2 \subseteq X_2$. See also Figure 11.

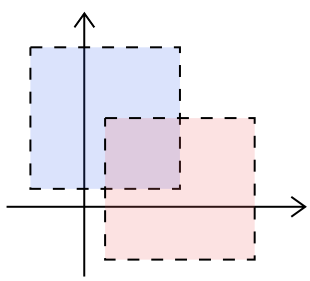


Figure 11: The intersection of two rectangles.

The natural projections are continuous, since both preimages $\pi_1^{-1}(U_1) = U_1 \times Y$ and $\pi_2^{-1}(U_2) = X \times U_2$ are open in $X \times Y$ whenever $U_1 \subseteq X_1$ and $U_2 \subseteq X_2$ are open.

Finally, consider a function $f\colon Z \to X_1 \times X_2$. If f is continuous, then both $\pi_1 \circ f$ and $\pi_2 \circ f$ are continuous by Lemma 1; conversely, if both $\pi_1 \circ f$ and $\pi_2 \circ f$ are continuous, and $U_1 \times U_2 \subseteq X_1 \times X_2$ is a product of open sets, we have that the inverse image $f^{-1}(U_1 \times U_2) = (\pi_1 \circ f)^{-1}(U_1) \cap (\pi_2 \circ f)^{-1}(U_2)$ is an intersection of open sets, and hence open as well. By Lemma 3, f is continuous. □

Definition 8 (Product topology)

If (X_1, τ_1) and (X_2, τ_2) are topological spaces, the topology $\tau_1 \otimes \tau_2$ defined in Proposition 3 is called the **product topology** on $X_1 \times X_2$.

The construction of the product topology can be done inductively for any finite number $(X_1, \tau_1), \ldots, (X_n, \tau_n)$ of topological spaces to obtain a topology $\bigotimes_{i=1}^n \tau_i$ on the product $X = \prod_{i=1}^n X_i$, all projections $\pi_i\colon X \to X_i$ become continuous, and the obvious analogue of the characteristic property holds. Repeating this construction for an infinite number of spaces actually yields something called the **box topology** on the cartesian product, with the definition of product topology being more subtle. We will not worry about this and work only on finite products here.

Exercise 13 (Paranoia or consistency? 3)

Let (X_i, τ_i), with $i = 1, 2, 3$, be topological spaces. We may consider the product space $(X_1 \times X_2, \tau_1 \otimes \tau_2)$, and then consider its product with (X_3, τ_3), obtaining $((X_1 \times X_2) \times X_3, (\tau_1 \otimes \tau_2) \otimes \tau_3)$. In a similar manner, we may consider the other product $(X_1 \times (X_2 \times X_3), \tau_1 \otimes (\tau_2 \otimes \tau_3))$. Show that both topologies on the triple product $X_1 \times X_2 \times X_3$ in fact agree (that is, $((x_1, x_2), x_3) \mapsto (x_1, (x_2, x_3))$ is a homeomorphism).

Example 13 (Tori)

For any integer $n \geq 1$, the ***n*-torus** is the cartesian product $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ of n copies of the circle $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, equipped with its product topology. This product topology in fact agrees with the subspace topology induced from $\mathbb{R}^{2n}$, if we write

$$\mathbb{T}^n = \{(x_1, y_1, \ldots, x_n, y_n) \in \mathbb{R}^{2n} : x_1^2 + y_1^2 = \cdots = x_n^2 + y_n^2 = 1\}.$$

There is a third way to describe tori, which we will see ahead in Example 14.

Exercise 14

The "vertical" topology from Example 12 is in fact equal to a certain product topology on $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Can you identify it?

Hint: What is the subspace topology in the x-axis $\mathbb{R} \times \{0\}$?

Exercise 15 (Bases for product spaces)

Let (X_1, τ_1) and (X_2, τ_2) be topological spaces with bases $\mathscr{B}_1$ and $\mathscr{B}_2$. Show that $\{B_1 \times B_2 : B_1 \in \mathscr{B}_1 \text{ and } B_2 \in \mathscr{B}_2\}$ is a basis for the product topology on $X_1 \times X_2$. Conclude that the product of two second-countable spaces is second-countable.

Exercise 16 (Paranoia or consistency? 3)

Let V be a finite-dimensional vector space, equipped with its Euclidean topology, and assume that there is a direct-sum decomposition $V = V_1 \oplus V_2$, for some non-trivial vector subspaces $V_1, V_2 \subseteq V$. These subspaces have Euclidean topologies of their own. Identifying V with the cartesian product $V_1 \times V_2$, show that the product topology on $V_1 \times V_2$ agrees with the Euclidean topology of V. (In other words, show that the sum map $V_1 \times V_2 \ni (v_1, v_2) \mapsto v_1 + v_2 \in V$ is a homeomorphism.)

We will encounter more properties of product spaces later, and finish this section with the following result:

Proposition 4 (The algebra $C^0(X)$)

Let (X, τ) be a topological space, and let $f, g\colon X \to \mathbb{R}$ be continuous functions, where $\mathbb{R}$ is equipped with its standard Euclidean topology. Then the functions

$$f + g\colon X \to \mathbb{R}, \quad fg\colon X \to \mathbb{R}, \quad \text{and} \quad \frac{f}{g}\colon X \smallsetminus g^{-1}(0) \to \mathbb{R}$$

are also continuous. In particular, the set $C^0(X)$ of all real-valued continuous functions on X, equipped with pointwise operations, is an $\mathbb{R}$-algebra[a].

[a] If $\mathbb{K}$ is any field, a $\mathbb{K}$-algebra is a vector field A over $\mathbb{K}$ equipped with a $\mathbb{K}$-bilinear operation $A \times A \to A$. We will see this concept again when discussing tangent spaces to manifolds later.

Proof: The addition $a\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, multiplication $m\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, and division $d\colon \mathbb{R} \times (\mathbb{R} \smallsetminus \{0\}) \to \mathbb{R}$, given by

$$a(u + v) = u + v, \quad m(u, v) = uv, \quad \text{and} \quad d(u, v) = u/v,$$

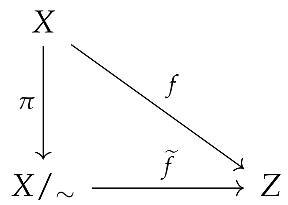
are all continuous functions (verify it!). Then

$$f + g = a \circ (f, g), \quad fg = m \circ (f, g), \quad \text{and} \quad f/g = d \circ (f, g)$$

must also be continuous, being compositions of continuous functions (cf. Lemma 1). That the pointwise operations in $C^0(X)$ satisfy the axioms required in the definition of an $\mathbb{R}$-algebra trivially follows from the corresponding axioms satisfied by the operations a and m in $\mathbb{R}$. $\square$

1.5 Quotient spaces

Recall that when X is any set and $\sim$ is an **equivalence relation**[1] on X, we may consider the quotient set $X/_\sim = \{[x]_\sim : x \in X\}$—its elements are the equivalence classes of $\sim$, defined as $[x]_\sim = \{y \in X : x \sim y\}$. Some basic facts are that $[x]_\sim = [y]_\sim$ if and only if $x \sim y$, and $[x]_\sim \cap [y]_\sim = \varnothing$ if $x \not\sim y$; hence X is a disjoint union of equivalence classes of $\sim$. If Z is a second set and $f\colon X \to Z$ is any function which is constant on each equivalence class of $\sim$, there is a unique function $\widetilde{f}\colon X/_\sim \to Z$ such that $\widetilde{f} \circ \pi = f$, where $\pi\colon X \to X/_\sim$ is the **quotient projection**, defined by $\pi(x) = [x]_\sim$. As usual, we describe this situation with a commutative diagram. As π is surjective, the images of f and $\widetilde{f}$ coincide. Moreover, if for every point $x \in X$ the equality $[x]_\sim = \{y \in X : f(y) = f(x)\}$ holds, then $\widetilde{f}$ is injective. (Constancy of f on the equivalence classes of $\sim$ only guarantees one inclusion.)



(1.8)

Here, however, we are interested in topology. If X starts with a topology, does the quotient $X/_\sim$ inherit a topology in any natural way? Is the projection π continuous? If the function f is continuous, does $\widetilde{f}$ come out continuous as well?

Very fortunately, the answers to all of these questions is *yes*, and $\pi\colon X \to X/_\sim$ turns out to be the protagonist of the story. Quotient constructions are a very rich source of examples in topology.

Consider the situation where $f\colon X \to Y$ is any surjective function between sets, and we define a binary relation $\sim$ on X by saying that $x_1 \sim x_2$ if $f(x_1) = f(x_2)$. Then $\sim$ is an equivalence relation on X and, by construction, the induced function $\widetilde{f}\colon X/_\sim \to Y$ is a bijection. In fact, *any* equivalence relation is of this form, with $Y = X/_\sim$ and $f = \pi$.

In view of the above, we may phrase things in a slightly more general manner, replacing $X/_\sim$ with a generic set Y. This is because topology "does not really care" about the particular equivalence relation $\sim$ under consideration.

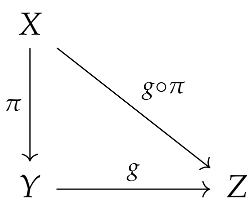
Proposition 5 (Setting up the quotient topology, and its characteristic property)

Let (X, τ) be a topological space, Y be any set, and $\pi\colon X \to Y$ be any function. Then, the collection

(1.9) $$\tau_\pi = \{U \subseteq Y : \pi^{-1}(U) \in \tau\}$$

is a topology on Y, and π itself is continuous.

In addition, if (Z, τ') is a third topological space and $g\colon Y \to Z$ is any function, then g is continuous if and only if $g \circ \pi\colon X \to Z$ is continuous, cf. the next diagram.



[1]That is, $\sim$ is a binary relation on X which is reflexive ($x \sim x$), symmetric ($x \sim y$ implies $y \sim x$), and transitive ($x \sim y$ and $y \sim z$ implies that $x \sim z$).

Proof: We verify the three conditions in Definition 3. First, $\varnothing \in \tau_\pi$ simply because $\pi^{-1}(\varnothing) = \varnothing \in \tau$. Secondly, if $\{U_\alpha\}_{\alpha\in A} \subseteq \tau_\pi$ is any collection of open sets, we have that $\pi^{-1}(\bigcup_{\alpha\in A} U_\alpha) = \bigcup_{\alpha\in A} \pi^{-1}(U_\alpha) \in \tau$, as each $\pi^{-1}(U_\alpha)$ is in τ by assumption, so that $\bigcup_{\alpha\in A} U_\alpha \in \tau_\pi$. In a similar manner, $\pi^{-1}(U_1 \cap U_2) = \pi^{-1}(U_1) \cap \pi^{-1}(U_2) \in \tau$ whenever $U_1, U_2 \subseteq \tau_\pi$, implying that $U_1 \cap U_2 \in \tau_\pi$. Hence, τ_π is a topology on Y. Continuity of $\pi\colon X \to Y$ is obvious.

Now, consider the characteristic property. If g is continuous, then $g \circ \pi$ is continuous by Lemma 1. Conversely, observe that $\pi^{-1}(g^{-1}(U)) = (g \circ \pi)^{-1}(U)$ for every subset $U \subseteq Z$. If $U \in \tau'$ and $g \circ \pi$ is continuous, the previous formula implies that $\pi^{-1}(g^{-1}(U)) \in \tau$, and thus $g^{-1}(U) \in \tau_\pi$, by definition of τ_π, meaning that g is continuous. ☐

Definition 9 (Quotient topology)

If (X, τ) is a topological space, Y is any set, and $\pi\colon X \to Y$ is any function, the topology τ_π in Proposition 5 is called the **quotient topology** on Y induced by π.

The characteristic property of the quotient topology is, in some categorical sense which can be made precise, "dual" to the characteristic property of the subspace topology. Here are some examples:

Example 14 (Circles as quotients)

In the real line $\mathbb{R}$ equipped with its standard Euclidean topology, define an equivalence relation $\sim$ by $x \sim y$ if and only if $x - y \in \mathbb{Z}$. The quotient space $\mathbb{R}/_\sim$ is usually denoted by $\mathbb{R}/\mathbb{Z}$ since, for each $x \in \mathbb{R}$, its equivalence class is given by $[x]_\sim = x + \mathbb{Z} = \{x + n : n \in \mathbb{Z}\}$. We always equip $\mathbb{R}/\mathbb{Z}$ with the quotient topology induced by $\pi\colon \mathbb{R} \to \mathbb{R}/\mathbb{Z}$. Imagine that we curl up the real line onto itself in such a way that all integer points lay on top of each other. Alternatively, note that every $x \in \mathbb{R}$ has a representative in the interval $[0, 1)$ (namely $x - \lfloor x \rfloor$, where $\lfloor \cdot \rfloor$ is the floor[a] function), so one can regard $\mathbb{R}/\mathbb{Z}$ as the closed interval $[0, 1]$ with its endpoints glued together.

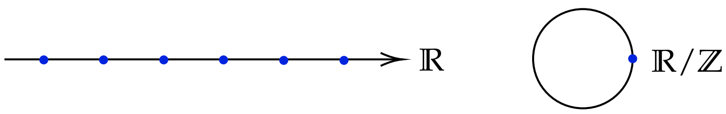


Figure 12: Visualizing the quotient space $\mathbb{R}/\mathbb{Z}$.

We will see ahead in Example 34 that $\mathbb{R}/\mathbb{Z}$ is homeomorphic to the circle $\mathbb{S}^1$, equipped with its standard Euclidean topology. One can make sense of $\mathbb{R}^n/\mathbb{Z}^n$, which turns out to be homeomorphic to the torus $\mathbb{T}^n$ from Example 13.

By the characteristic property of the quotient topology, if (Z, τ') is any topological space, continuous functions $\mathbb{R}/\mathbb{Z} \to Z$ are in one-to-one correspondence with continuous functions $f\colon \mathbb{R} \to Z$ such that $f(x + 1) = f(x)$ for all $x \in \mathbb{R}$.

[a] For example, $\lfloor 10.2 \rfloor = 10$, and $\lfloor -2.3 \rfloor = -3$.

Example 15 (Real projective space)

Let $n \geq 2$ be an integer, and consider in $\mathbb{R}^{n+1} \smallsetminus \{0\}$ the binary relation given by $p \sim q$ if there is $\lambda \in \mathbb{R} \smallsetminus \{0\}$ such that $q = \lambda p$. It is not hard to see that this is an equivalence relation—the quotient set $(\mathbb{R}^{n+1} \smallsetminus \{0\})/_\sim$ is instead denoted by $\mathbb{RP}^n$, and called the **n-dimensional real projective space**. Concretely, $\mathbb{RP}^n$ is the space of all lines passing through the origin of $\mathbb{R}^{n+1}$.

By the characteristic property of the quotient topology, if (Z, τ') is any topological space, continuous functions $\mathbb{RP}^n \to Z$ are in one-to-one correspondence with continuous functions $f\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to Z$ with the property that $f(\lambda p) = f(p)$ for all $p \in \mathbb{R}^{n+1} \smallsetminus \{0\}$ and $\lambda \in \mathbb{R} \smallsetminus \{0\}$.

Example 16 (The topology on quotients by group actions)

Let (X, τ) be a topological space, and G be any group acting on X. This means that there is a mapping $G \times X \ni (g, x) \mapsto g \cdot x \in X$ such that $e \cdot x = x$ and $g \cdot (h \cdot x) = (gh) \cdot x$ for every $g, h \in G$ and $x \in X$; recall that setting $x \sim y$ if there is $g \in G$ such that $y = g \cdot x$ defines an equivalence relation on X, and the quotient set $X/_\sim$ is instead denoted by X/G. The equivalence class of $x \in X$ is denoted by $G \cdot x$, and it is called the **orbit** of x. The quotient projection $\pi\colon X \to X/G$, given by $\pi(x) = G \cdot x$, induces a quotient topology on X/G. When talking about topologies on quotients of topological spaces under group actions, this is the topology chosen by default.

By the characteristic property of the quotient topology, if (Z, τ') is any topological space, continuous functions $X/G \to Z$ are in one-to-one correspondence with G-**invariant** continuous functions $f\colon X \to Z$ (that is, such that $f(g \cdot x) = f(x)$ for all $x \in X$ and $g \in G$.)

Exercise 17

Some spaces we have encountered before may also be realized as quotients. Consider $Z = \mathbb{R} \times \{0, 1\}$ equipped with the subspace topology induced by $\mathbb{R}^2$, and define an equivalence relation $\sim$ on Z by setting $(x, t) \sim (y, s)$ if $x = y \neq 0$ (and, otherwise, every element is $\sim$-related just to itself). Equip $Z/_\sim$ with its quotient topology, induced by $\pi\colon Z \to Z/_\sim$. Show that $Z/_\sim$ is homeomorphic to the line with two origins (Exercise 12).

Hint: There are three types of equivalence classes. Namely, $\{(x, 0), (x, 1)\}$ for $x \neq 0$, then $\{(0, 0)\}$, and $\{(0, 1)\}$. Where should each of them get mapped to, in the line with two origins? Use the characteristic property of the quotient topology to show that your guess does produce a continuous function $Z/_\sim \to X$.

We introduce some more terminology:

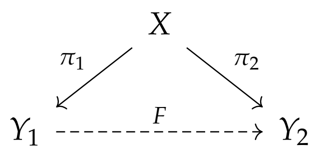
Definition 10 (Quotient mappings)

Let (X, τ) and (Y, τ') be topological spaces, and $f\colon X \to Y$ be a surjective continuous function. We call f a **quotient mapping** if the quotient topology on Y induced by f in fact equals τ'.

In the next result, recall from set-theory that inverse images of singletons are usually called "fibers".

Proposition 6 (Uniqueness of quotients)

Let (X, τ), (Y_1, τ_1), and (Y_2, τ_2) be topological spaces, and $\pi_1\colon X \to Y_1$ and $\pi_2\colon X \to Y_2$ be (surjective) quotient mappings. If π_1 and π_2 are constant along each other's fibers, there is a unique homeomorphism $F\colon Y_1 \to Y_2$ such that the diagram commutes:



Proof: With a moment of thought, we see that there is no other option to define F: if $y_1 \in Y_1$, we choose $x \in X$ such that $\pi_1(x) = y_1$, and then set $F(y_1) = \pi_2(x)$. This definition is correct (that is, F is well-defined) since π_2 restricted to $\pi_1^{-1}(y)$ is constant. Continuity of F follows from continuity of π_2 through the characteristic property of π_1, cf. Proposition 5. Repeating this argument, reversing the roles of (Y_1, τ_1) and (Y_2, τ_2), as well as the ones of π_1 and π_2, yields the definition and continuity of F^{-1}. Hence, F is a homeomorphism. □

There is one more concept which goes hand-in-hand with quotient mappings:

Definition 11 (Open mappings)

Let (X, τ) and (Y, τ') be topological spaces, and $f\colon X \to Y$ be any function. We say that f is an **open mapping** if $f(U) \in \tau'$ whenever $U \in \tau$, i.e., if direct images of open sets are again open.

There is one main example we should keep in mind, for reasons which will be clear much later:

Example 17 (Cartesian projections between Euclidean spaces are open)

The projection $\pi\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ given by $\pi(x, y) = x$ is an open mapping. Indeed, let $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ be open, and consider $\pi(U) \subseteq \mathbb{R}^n$. If $x \in \pi(U)$, there is $y \in \mathbb{R}^k$ such that $(x, y) \in U$, and we may find $\varepsilon > 0$ such that $B_\varepsilon(x, y) \subseteq U$. But it is clear that $B_\varepsilon(x) \subseteq \pi(B_\varepsilon(x, y))$, and so $B_\varepsilon(x) \subseteq \pi(U)$. Hence, $\pi(U)$ is open.

Moreover, open mappings allow us to "transfer" a basis from the first space onto the second.

Proposition 7 (Open images of 2^{nd}-countable spaces are 2^{nd}-countable)

Let (X, τ) and (Y, τ') be topological spaces, and $f\colon X \to Y$ be a surjective, continuous, and open mapping. If (X, τ) is second-countable, so is (Y, τ').

Proof: We claim that if $\mathscr{B}$ is a basis for (X, τ), then $f(\mathscr{B}) = \{f(B) : B \in \mathscr{B}\}$ is a basis for (Y, τ'). Indeed, let $V \subseteq Y$ be open, and $y \in V$ be any point. By surjectivity of f there is $x \in X$ such that $f(x) = y$, and we necessarily have $x \in f^{-1}(V)$. But $f^{-1}(V) \subseteq X$ is open and $\mathscr{B}$ is a basis, so there is $B \in \mathscr{B}$ such that $x \in B \in f^{-1}(V)$. It follows that $y = f(x) \in f(B) \subseteq f(f^{-1}(V)) = V$, as required (the last equality uses surjectivity of f again). Whenever $\mathscr{B}$ is countable, so is $f(\mathscr{B})$. $\square$

Continuous mappings are not necessarily open mappings, and open mappings are not necessarily continuous: the two notions are logically independent. Even quotient mappings are not necessarily open—but the ones who happen to be are particularly nice. So, when are quotient mappings open?

With the same notation from the previous two definitions, a subset $S \subseteq X$ is called **saturated** if $S = \pi^{-1}(\pi(S))$. This is just another way of saying that S must be a union of fibers. Similarly, if $A \subseteq X$ is any subset, the **saturation of** A is defined as $\mathrm{Sat}(A) = \pi^{-1}(\pi(A))$. It is clear that $A \subseteq \mathrm{Sat}(A)$ always holds, and $\mathrm{Sat}(A)$ is the smallest saturated subset of X containing A. Geometrically, $\mathrm{Sat}(A)$ is just the union of all the fibers of π which intersect A, cf. Figure 13.

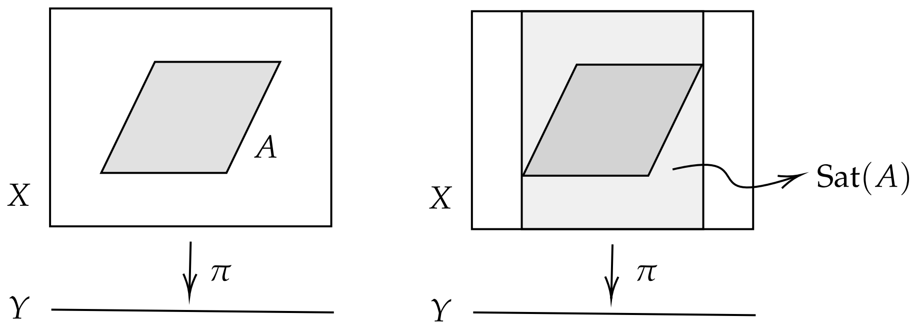


Figure 13: The saturation of a set.

Proposition 8

Let (X, τ) and (Y, τ') be topological spaces, and $\pi\colon X \to Y$ be a quotient mapping. Then π is an open mapping if and only if $\mathrm{Sat}(U) \in \tau$ whenever $U \in \tau$, that is, if the saturation of any open set is open.

Proof: If π is an open mapping and $U \subseteq X$ is open, then $\pi(U) \subseteq Y$ is open, and then continuity of π implies that $\mathrm{Sat}(U) = \pi^{-1}(\pi(U))$ is open.

Conversely, let $U \subseteq X$ be open. To show that $\pi(U) \subseteq Y$ is open, via the definition of quotient topology, we must check that $\mathrm{Sat}(U) = \pi^{-1}(\pi(U)) \subseteq X$ is open. But this was our assumption. $\square$

Example 18

Consider the unit sphere $\mathbb{S}^n \subseteq \mathbb{R}^{n+1}$, and let $\mathbb{Z}_2$ act on $\mathbb{S}^n$ by setting $1 \cdot p = p$ and $(-1) \cdot p = -p$, for each $p \in \mathbb{S}^n$. We equip the quotient set $\mathbb{S}^n/\mathbb{Z}_2$ with the quotient topology induced by the projection $\pi\colon \mathbb{S}^n \to \mathbb{S}^n/\mathbb{Z}_2$, which is given by $\pi(p) = \{p, -p\}$. For each open subset $U \subseteq \mathbb{S}^n$, we have that $\mathrm{Sat}(U) = U \cup (-U)$ is the union of two open sets, and hence open. Here, we write $-U = \{p \in \mathbb{S}^n : -p \in U\}$ (it is clearly homeomorphic to U). By Proposition 8, π is an open mapping.

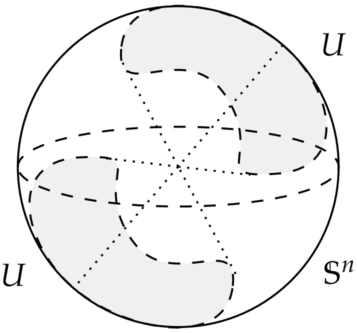


Figure 14: The $\mathbb{Z}_2$-saturation of an open subset $U \subseteq \mathbb{S}^n$.

The situation from the above example has a great generalization:

Exercise 18

Let (X, τ) be a topological space, and G be a group acting on X by homeomorphisms (that is, each $g \in G$ regarded as a mapping $g\colon X \to X$ is a homeomorphism). Show that the quotient projection $\pi\colon X \to X/G$ is an open mapping.

Finally, here is one last property we will need later.

Proposition 9

Let (X, τ) and (Y, τ') be topological spaces, and $\pi\colon X \to Y$ be a quotient mapping. Then, whenever $U \subseteq X$ is open and saturated, the restriction $\pi|_U\colon U \to \pi(U)$ is also a quotient mapping. (Here, U and $\pi(U)$ are equipped with their subspace topologies induced by X and Y.)

For the (slightly technical) proof, see [20, Theorem 22.1].

1.6 Some more point-set concepts

Here, we introduce some useful terminology for when discussing aspects of subspaces of topological spaces. We begin with the notion "dual" to the one of an open subset:

Exercise 19 (Definition and basic properties of closed sets)

Let (X, τ) be a topological space. Let us say that a subset $C \subseteq X$ is **closed** if $X \smallsetminus C \in \tau$, that is, if the complement $X \smallsetminus C$ is open. Show that:

(a) $\varnothing$ and X are closed.

(b) If $(C_i)_{i\in I}$ is a family of closed sets, then the intersection $\bigcap_{i\in I} C_i$ is also closed.

(c) If $C_1, C_2 \subseteq X$ are closed, then $C_1 \cup C_2$ is closed.

This means that to prescribe a topology on a set X, it suffices to prescribe a collection of subsets of X satisfying (a)–(c) above, call them closed, and define open subsets as being the complements of the closed ones.

Remark. There is a very important **warning** to be made here. If a set is not open, it does not mean that it is closed! If it is not closed, it does not mean that it is open either! A set may be both open *and* closed (such as the empty set, or the entire space X), or it might be neither (such as $[0,1) \subseteq \mathbb{R}$, with the Euclidean topology).

Definition 12 (Interior and closure)

Let (X,τ) be a topological space, $A \subseteq X$ be any subset, and $x \in X$ be any point.

(i) x is an **interior point** of A if there is $U \in \tau$ such that $x \in U \subseteq A$.

(ii) x is a **closure point** of A if for every $U \in \tau$ with $x \in U$, we have $U \cap A \neq \varnothing$.

The set of all interior points of A is called the **interior of** A and is denoted by $\mathring{A}$ (or $\operatorname{int}_X(A)$), while the set of all closure points of A is called the **closure of** A and is denoted by $\overline{A}$ (or $\operatorname{cl}_X(A)$).

Clearly, the inclusions $\mathring{A} \subseteq A$ and $A \subseteq \overline{A}$ always hold.

The same idea behind Lemma 2 shows that $\mathring{A}$ is always open: for every $x \in \mathring{A}$ there is $U_x \in \tau$ such that $x \in U_x \subseteq A$, but U_x itself (being open) is a witness that all of its points are interior to A, so that $U_x \subseteq \mathring{A}$. Hence, $\mathring{A} = \bigcup_{x\in\mathring{A}} U_x$ is an union of open sets.

Dually, $\overline{A}$ is always closed. Indeed, if $x \in X \smallsetminus \overline{A}$, there is $U \in \tau$ such that $x \in U$ but $U \cap A = \varnothing$. This last condition implies that $U \subseteq A$ or $U \subseteq X \smallsetminus A$, but the former possibility is ruled out by $x \in A$. No point in U is in $\overline{A}$, so in fact $U \subseteq X \smallsetminus \overline{A}$. This means—by the very definition of interior point—that $x \in (X \smallsetminus \overline{A})^\circ$, and hence $X \smallsetminus \overline{A} \subseteq (X \smallsetminus \overline{A})^\circ$. We conclude that $X \smallsetminus \overline{A} = (X \smallsetminus \overline{A})^\circ$ is open.

But there is more we can say:

Proposition 10

Let (X,τ) be a topological space, and $A \subseteq X$. Then:

(i) $\mathring{A} = \bigcup\{U \subseteq X : U \text{ is open and } U \subseteq A\}$;

(ii) $\overline{A} = \bigcap\{C \subseteq X : C \text{ is closed and } A \subseteq C\}$.

In other words, $\mathring{A}$ is the largest open subset of X contained in A, while $\overline{A}$ is the smallest closed subset of X containing A.

Proof: We prove (ii) and leave (i) as an exercise.

If $x \in \overline{A}$ but there is a closed subset $C \subseteq X$ such that $A \subseteq C$ but $x \notin C$, then $x \in X \smallsetminus C$. As $X \smallsetminus C$ is open, the definition of closure implies that $(X \smallsetminus C) \cap A \neq \varnothing$, contradicting $A \subseteq C$ and establishing that $\overline{A} \subseteq \bigcap\{C \subseteq X : C \text{ is closed and } A \subseteq C\}$.

For the reverse inclusion, let $C \subseteq X$ and assume that $A \subseteq C$. We show that $C \subseteq \overline{A}$: if there were $x \in C$ with $x \notin \overline{A}$, we would be able to find an open set U with $x \in U$ but $U \cap A = \varnothing$, implying that $x \notin A$ and contradicting that $A \subseteq C$. $\square$

Corollary 1 (Open and closedness in terms of interiors and closures)

Let (X, τ) be a topological space, and $A \subseteq X$ be any subset. Then:

(i) A is open if and only if $A = \mathring{A}$.

(ii) A is closed if and only if $A = \overline{A}$.

In particular, $\mathring{\mathring{A}} = \mathring{A}$ and $\overline{\overline{A}} = \overline{A}$.

Exercise 20

Prove (i) in Proposition 10, as well as Corollary 1.

We are due for some examples. There is one last concept that will help us form a full picture of what is happening here:

Definition 13 (Topological boundary)

Let (X, τ) be a topological space, and $A \subseteq X$ be any subset. A point $x \in X$ is called a **boundary point** of A if for every $U \in \tau$ with $x \in X$, we have $U \cap A \neq \varnothing$ and $U \cap (X \smallsetminus A) \neq \varnothing$. The set of all boundary points of A is called the **boundary of** A, and is denoted by ∂A (or $\mathrm{bd}_X(A)$).

Note that $\partial A = \overline{A} \cap \overline{X \smallsetminus A}$, by definition of closure. Hence, ∂A is always closed, and the relation $\partial A = \partial(X \smallsetminus A)$ holds.

See Figure 15.

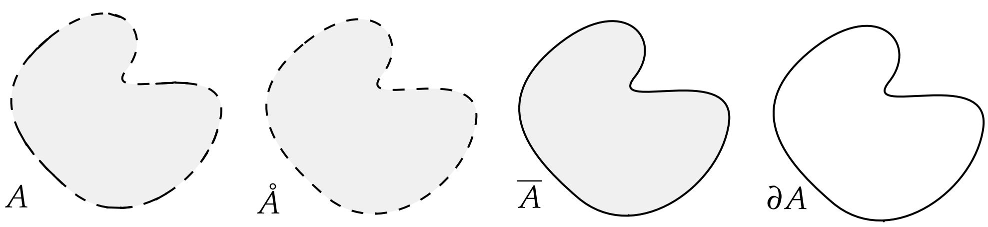


Figure 15: Visualizing the interior, closure, and boundary of a planar region.

Example 19 (In the real line)

In the real line $\mathbb{R}$ equipped with its standard Euclidean topology, consider the subset $A = [0,1) \cup \{2\}$. Then $\mathring{A} = (0,1)$, $\overline{A} = [0,1] \cup \{2\}$, and $\partial A = \{0,1,2\}$.

Example 20 (The extreme cases)

Let $(X, \tau_{\text{disc.}})$ be a discrete topological space, and $A \subseteq X$ be any subset. Then $\mathring{A} = \overline{A} = A$, while $\partial A = \varnothing$. At the other extreme, if we consider $\tau_{\text{ch.}}$ instead of $\tau_{\text{disc.}}$ on X, we have $\mathring{A} = \varnothing$ unless $A = X$, as well as $\overline{A} = X$ unless $A = \varnothing$, and then $\partial A = X$ whenever $A \notin \{\varnothing, X\}$.

Example 21 (Closures of balls)

Let (X, d) be a metric space, $a \in X$, and $r > 0$. Then $\overline{B_r(a)} \subseteq B_r[a]$, since $B_r[a]$ is closed and contains $B_r(a)$ (here we use Proposition 10). The reverse inclusion, however, need not hold in general: with the discrete metric, consider for instance $B_{1/2}(a) = B_{1/2}[a] = \{a\}$, with $\overline{B_r(a)} = X$. If X is a vector space and d is induced by a norm $\|\cdot\|$ on X, the equality $\overline{B_r(a)} = B_r[a]$ holds: if $x \in B_r(a)$ and $U \subseteq X$ is any open subset containing x, there is $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$, and then $a + (\delta/r)(x-a) \in U \cap B_r(a)$ whenever $r - \varepsilon < \delta < r$, showing that $x \in \overline{B_r(a)}$.

Exercise 21

Let $W \subseteq \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^n$, and assume that $\mathbb{R}^n$ is equipped with its standard Euclidean topology. Show that W is closed and, if $W \neq \mathbb{R}^n$, then $\mathring{W} = \varnothing$.

Example 22

In the Sorgenfrey line $(\mathbb{R}, \tau_S)$, let us compute $(0,1]^\circ$ and $\overline{(0,1]}$.

Interior: for each $n \geq 1$, we have that $[1/n, 1) \in \tau_S$ and $[1/n, 1) \subseteq (0,1]$. Hence $(0,1) = \bigcup_{n\geq 1}[1/n,1) \subseteq (0,1]$, showing that $(0,1) \subseteq (0,1]^\circ$. It remains to decide whether $x = 1$ is an interior point of $(0,1]$. But for every $\varepsilon > 0$, we have that $[1, 1+\varepsilon) \cap (\mathbb{R} \smallsetminus (0,1]) \neq \varnothing$, so that $1 \notin (0,1]^\circ$, and thus $(0,1]^\circ = (0,1)$.

Closure: for each $x > 1$, then $[x, \infty) \cap (0,1] = \varnothing$, so that $[x,\infty) \in \tau_S$ implies that $x \notin \overline{(0,1]}$; similarly, if $x < 0$, we have that $[x, x/2) \cap (0,1] = \varnothing$ with $[x, x/2) \in \tau_S$ implies that $x \notin \overline{(0,1]}$. Finally, for every $\varepsilon > 0$ we have that $[0,\varepsilon) \cap (0,1] \neq \varnothing$, so that $0 \in \overline{(0,1]}$. We conclude that $\overline{(0,1]} = [0,1]$.

There are other ways to justify the above, e.g., showing directly that $[0,1]$ is Sorgenfrey-closed.

Exercise 22

Let X be any infinite set, equipped with the cofinite topology $\tau_{\mathrm{cof.}}$, and $A \subseteq X$ be any subset. Show that $\overline{A} = X$ if and only if A is infinite.

As Figure 15 suggests, it is possible to detect whether subsets are open or not via their boundary.

Proposition 11

Let (X, τ) be a topological space, and $A \subseteq X$ be any subset. Then A is open if and only if $A \cap \partial A = \varnothing$.

Proof: In general, $\mathring{A} \subseteq X \smallsetminus \partial A$, so $\mathring{A} \cap \partial A = \varnothing$. If A is open, then $A = \mathring{A}$ and hence $A \cap \partial A = \varnothing$. Conversely, assume that $A \cap \partial A = \varnothing$, and let $x \in A$. As $x \notin \partial A$, there is an open subset $U \subseteq X$ such that $x \in U$ and with either $U \subseteq A$ or $U \subseteq X \smallsetminus A$—the latter case cannot happen as $x \in U \cap A$. □

Here is how interiors and closures behave under unions and intersections:

Proposition 12

Let (X, τ) be a topological space, and $A_1, A_2 \subseteq X$ be any subsets. Then:

(i) $\mathring{A}_1 \cap \mathring{A}_2 = (A_1 \cap A_2)^\circ$;

(ii) $\mathring{A}_1 \cup \mathring{A}_2 \subseteq (A_1 \cup A_2)^\circ$;

(iii) $\overline{A_1 \cup A_2} = \overline{A_1} \cup \overline{A_2}$;

(iv) $\overline{A_1 \cap A_2} \subseteq \overline{A_1} \cap \overline{A_2}$.

Proof: We will use that taking interiors and taking closures are inclusion-preserving operations, that is, if $A \subseteq B$ then $\mathring{A} \subseteq \mathring{B}$ and $\overline{A} \subseteq \overline{B}$—this follows immediately from Proposition 10.

As we have that $A_1 \cap A_2 \subseteq A_i$ for $i = 1, 2$, we have that $(A_1 \cap A_2)^\circ \subseteq \mathring{A}_i$, and so $(A_1 \cap A_2)^\circ \subseteq \mathring{A}_1 \cap \mathring{A}_2$. For the reverse inclusion, we note that $\mathring{A}_1 \cap \mathring{A}_2$ is open and has $\mathring{A}_1 \cap \mathring{A}_2 \subseteq A_1 \cap A_2$, so that $\mathring{A}_1 \cap \mathring{A}_2 \subseteq (A_1 \cap A_2)^\circ$. This establishes (i). Item (ii) is dealt with similarly: as $A_i \subseteq A_1 \cup A_2$, it follows that $\mathring{A}_i \subseteq (A_1 \cup A_2)^\circ$, and so $\mathring{A}_1 \cup \mathring{A}_2 \subseteq (A_1 \cup A_2)^\circ$.

Now we consider closures. As above, $\overline{A_1} \cup \overline{A_2} \subseteq \overline{A_1 \cup A_2}$, while for the reverse inclusion follows from noting that $\overline{A_1} \cup \overline{A_2}$ is closed and contains $A_1 \cup A_2$, so that $\overline{A_1 \cup A_2} \subseteq \overline{A_1} \cup \overline{A_2}$; this proves (iii). Finally, we return to $A_1 \cap A_2 \subseteq A_i$ for $i = 1, 2$, so that $\overline{A_1 \cap A_2} \subseteq \overline{A_i}$, and hence $\overline{A_1 \cap A_2} \subseteq \overline{A_1} \cap \overline{A_2}$, establishing (iv). □

Exercise 23

Show by examples that the inclusions in items (ii) and (iv) of Proposition 12 can indeed be strict.

In the next exercise, you can check how unions and closures behave relative to taking complements:

Exercise 24 (Interior-closure duality)

Let (X, τ) be a topological space, and $A \subseteq X$ be any subset. Show that:

(a) $\overline{X \setminus A} = X \setminus \mathring{A}$.

(b) $(X \setminus A)^\circ = X \setminus \overline{A}$.

Proposition 13

Let (X, τ) be a topological space, and $A \subseteq X$ be any subset. Then

$$X = \mathring{A} \,\dot{\cup}\, \partial A \,\dot{\cup}\, (X \setminus A)^\circ. \tag{1.10}$$

Here, the dots over the union symbols mean that we have a disjoint union.

Proof: Let $x \in X$. If $x \in \mathring{A}$, we are done. Otherwise, for every open subset $U \subseteq X$ such that $x \in U$, we have that $U \cap (X \setminus A) \neq \varnothing$. If some such U is entirely contained in $X \setminus A$, then $x \in (X \setminus A)^\circ$. Else, every such U has $U \cap A \neq \varnothing$ and $U \cap (X \setminus A) \neq \varnothing$, so that $x \in \partial A$. □

For the last exercise in this section, we use the following definition: a subset D of a topological space (X, τ) is called **dense** if $\overline{D} = X$. This can be equivalently rephrased as saying that $D \cap U \neq \varnothing$ for every non-empty open subset $U \subseteq X$.

Exercise 25

Unfortunately, metric spaces are not necessarily second-countable (cf. Example 9, with the **discrete metric**[a]), but when arguing in Example 7 that the collection $\{(x - \varepsilon, x + \varepsilon) : x \in \mathbb{Q} \text{ and } \varepsilon \in (0, \infty) \cap \mathbb{Q}\}$ is a basis for the Euclidean topology in $\mathbb{R}$, we used very strongly that $\mathbb{Q}$ is dense in $\mathbb{R}$. So, let's say that a topological space (X, τ) is **separable** if it contains a dense countable subset.

(a) Show that every second-countable topological space is separable.

(b) Show that the converse holds for metric spaces: every separable metric space is second-countable.

[a] That is, $\mathrm{d}\colon X \times X \to [0, \infty)$ is given by $\mathrm{d}(x, y) = 1$ if $x \neq y$ and $\mathrm{d}(x, y) = 0$ if $x = y$. It indeed satisfies the conditions in Definition 1 and we have that $\tau_{\mathrm{d}} = \tau_{\text{disc.}}$ (verify it!).

1.7 Convergence and Hausdorff spaces

One of the main concepts studied in real analysis is the one of convergence. Recall that for a sequence $(x_n)_{n\geq 1}$ of real numbers, we say that $x_n \to x$ if

$$\forall\, \varepsilon > 0,\, \exists\, n_0 \in \mathbb{N}_{>0} \,:\, \forall\, n \in \mathbb{N},\, n \geq n_0 \implies |x_n - x| < \varepsilon. \tag{1.11}$$

The generalization to convergence in a metric space (X, d) is immediate, if you remember how we started the entire discussion on topology from the definition of continuity:

$$\forall\, \varepsilon > 0,\, \exists\, n_0 \in \mathbb{N}_{>0} \,:\, \forall\, n \in \mathbb{N},\, n \geq n_0 \implies \mathrm{d}(x_n, x) < \varepsilon. \tag{1.12}$$

The next definition should not be too surprising:

Definition 14 (Convergence in topological spaces)

Let (X, τ) be a topological space, and $(x_n)_{n\geq 1}$ be a sequence of points in X. We say that $(x_n)_{n\geq 1}$ **converges** to a point $x \in X$ if for every $U \in \tau$ with $x \in U$ there is an integer $n_0 \geq 1$ such that whenever $n \geq n_0$, we have $x_n \in U$.

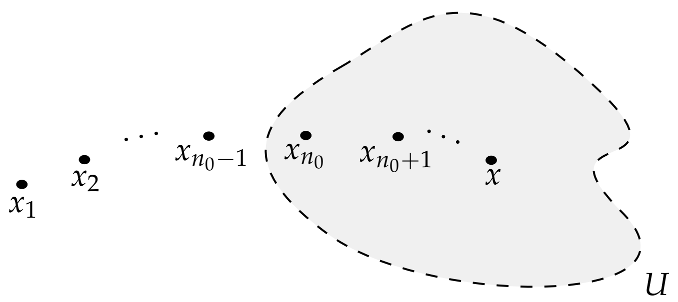


Figure 16: All but finitely many terms of $(x_n)_{n\geq 1}$ must enter U.

We denote this by $x_n \to x$.

There is an important warning to be made here, when working with this level of generality: we are only allowed to write $x = \lim_{n\to+\infty} x_n$ once we know that limits are unique, otherwise the symbol $\lim_{n\to+\infty} x_n$ could be ambiguous, referring to more than one point in the space X.

The problem is that limits of sequences are not necessarily unique.

Example 23 (The true reason for the name "chaotic topology")

Let X be any set, and equip it with the chaotic topology $\tau_{\text{ch.}} = \{\varnothing, X\}$. If $(x_n)_{n\geq 1}$ is *any* sequence in X, and $x \in X$ is *any* point, then $x_n \to x$. Indeed, the only open set we need to consider when testing for Definition 14 is $U = X$, which always contains all terms x_n. In other words, any sequence converges to all points of X at the same time!

Luckily, the spaces where limits fail to be unique are rather pathological—the majority of the spaces we encounter "in nature" are well-behaved.

Proposition 14

Limits of sequences are unique in metric spaces.

Proof: Let (X, d) be a metric space, and $(x_n)_{n\geq 1}$ be a sequence in X such that $x_n \to x$ and $x_n \to y$, for some points $x, y \in X$. We will show that $x = y$. Let $\varepsilon > 0$ be arbitrary and choose integers $n_1, n_2 \geq 1$ such that (i) $x_n \in B_{\varepsilon/2}(x)$ for every $n \geq n_1$, and (ii) $x_n \in B_{\varepsilon/2}(y)$ for every $n \geq n_2$. If we set $n_0 = \max\{n_1, n_2\}$, then for every $n \geq n_0$ we have that

$$\mathrm{d}(x,y) \leq \mathrm{d}(x,x_n) + \mathrm{d}(x_n,y) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

If $0 \leq \mathrm{d}(x,y) < \varepsilon$ for every $\varepsilon > 0$, then it must be $\mathrm{d}(x,y) = 0$, and thus $x = y$.

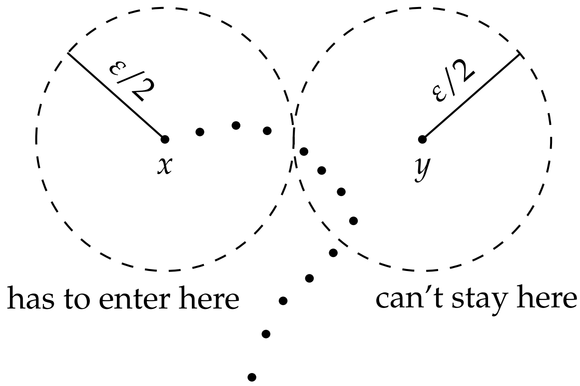


Figure 17: Uniqueness of limits in metric spaces.

In other words, if a sequence must enter arbitrarily small balls around a point, it cannot stay in arbitrarily small balls centered at different points. □

What really mattered in the above proof was not the fact that we had a distance function, but instead that we were able to separate the points x and y with disjoint open sets. We turn this condition into a definition.

Definition 15 (Hausdorff spaces)

A topological space (X, τ) is called a **Hausdorff space** if whenever $x, y \in X$ are distinct points, there are disjoint sets $U, V \in \tau$ with $x \in U$ and $y \in V$.

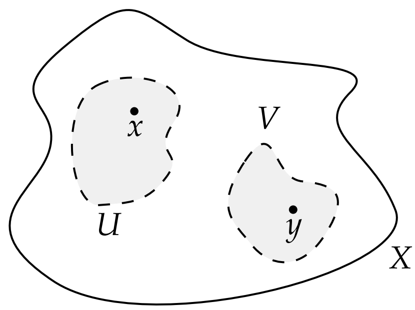


Figure 18: The Hausdorff condition.

Figure 18, as expected, is what one usually has in mind when thinking about Hausdorff spaces.

Proposition 15

Limits of sequences are unique in Hausdorff spaces.

Proof: Let (X, τ) be a Hausdorff space, and $(x_n)_{n\geq 1}$ be a sequence in X. If $x, y \in X$ are such that $x_n \to x$ and $x \neq y$, we show that $x_n \not\to y$. Namely, if this were to be the case, and $U, V \subseteq X$ are disjoint open sets with $x \in U$ and $y \in V$, the definition of convergence would provide us with integers $n_1, n_2 \geq 1$ such that (i) $x_n \in U$ for every $n \geq n_1$, and (ii) $x_n \in V$ for any $n \geq n_2$. Thus, if $n_0 = \max\{n_1, n_2\}$ and $n \geq n_0$, it follows that $x_n \in U \cap V$, which is a contradiction. □

Derivatives and integrals—the protagonists of Calculus—are, by definition, certain limits. If manifolds should be topological spaces on which we can perform Calculus, we should certainly require them to be Hausdorff spaces.

Some examples:

Example 24

Metric spaces are Hausdorff: if $x, y \in X$ are distinct, we let $r = \mathrm{d}(x, y)/2 > 0$, so that $B_r(x) \cap B_r(y) = \varnothing$; clearly $x \in B_r(x)$ and $y \in B_r(y)$.

Example 25

Discrete spaces are always Hausdorff: we can take U and V to be singletons.

Example 26

The two-point Sierpinski space $(\{a, b\}, \{\varnothing, \{a\}, \{a, b\}\})$ (Example 3) is not Hausdorff. For example, the constant sequence equal to a converges to both a and b.

Example 27

You have shown in Exercise 12 that the line with two origins is not Hausdorff: the two origins z_1 and z_2 cannot be separated by disjoint open sets. Alternatively, note that the sequence given by $x_n = 1/n$ converges to both z_1 and z_2.

Exercise 26 (Hausdorffness of products and subspaces)

Show that any countably infinite set equipped with its cofinite topology (cf. Exercise 4) is not Hausdorff.

Exercise 27

Show that:

(a) subspaces of Hausdorff spaces are also Hausdorff.

(b) the product of two Hausdorff spaces is also Hausdorff.

(c) quotients of Hausdorff spaces need *not* be Hausdorff.

The proof of the result below is very helpful to see how the definition of convergence plays along the definition of continuity:

Proposition 16

Let (X, τ) and (Y, τ') be topological spaces, $f\colon X \to Y$ be a continuous function, and $(x_n)_{n\geq 1}$ be a sequence in X such that $x_n \to x$, with $x \in X$. Then, $f(x_n) \to f(x)$.

Proof: The argument is very direct, applying the relevant definitions. Let $V \in \tau'$ be such that $f(x) \in V$, and consider $f^{-1}(V) \in \tau$ (here is where continuity of f enters). As $x \in f^{-1}(V)$, the definition of convergence gives us an integer $n_0 \geq 1$ such that $x_n \in f^{-1}(V)$ whenever $n \geq n_0$. Then $f(x_n) \in V$ whenever $n \geq n_0$, showing that $f(x_n) \to f(x)$ as required. □

Here is one more example:

Example 28 ($\mathbb{S}^n/\mathbb{Z}_2$ is Hausdorff)

Consider again the quotient $\mathbb{S}^n/\mathbb{Z}_2$, from Example 18. We claim that $\mathbb{S}^n/\mathbb{Z}_2$ is Hausdorff. Let $\{p, -p\}, \{q, -q\} \in \mathbb{S}^n/\mathbb{Z}_2$ be distinct, so that $p \neq q$ and $p \neq -q$. As $\mathbb{S}^n$ is Hausdorff, there are mutually disjoint open subsets $U_1, U_2, V_1, V_2 \subseteq \mathbb{S}^n$ such that $p \in U_1$, $-p \in U_2$, $q \in V_1$, and $-q \in V_2$. Then the images of $U = U_1 \cup U_2$ and $V = V_1 \cup V_2$ under the projection $\mathbb{S}^n \to \mathbb{S}^n/\mathbb{Z}_2$ are open, and in fact disjoint: if not, we may write $y = \pm x$ for some choice of sign $\pm$, for some x in either U_1 or U_2 and y in either V_1 or V_2, with all possibilities leading to contradictions.

We conclude this section with an alternative characterization of the Hausdorff property. Given any set X, its **diagonal** is defined as $\Delta = \{(x, y) \in X \times X : x = y\}$.

Proposition 17

Let (X, τ) be a topological space. Then, (X, τ) is Hausdorff if and only if the diagonal subspace $\Delta \subseteq X \times X$ is closed. (Here, $X \times X$ has the product topology.)

Proof: Assuming that (X, τ) is Hausdorff, we show that $(X \times X) \smallsetminus \Delta$ is open. If $(x, y) \in (X \times X) \smallsetminus \Delta$, then $x \neq y$, and the Hausdorff condition allows us to choose disjoint open subsets $U, V \subseteq X$ with $x \in U$ and $y \in V$. Then $(x, y) \in U \times V$ and

$U \times V \subseteq (X \times X) \setminus \Delta$ (this is equivalent to $U \cap V = \varnothing$), showing that (x,y) is interior to $(X \times X) \setminus \Delta$. Thus $(X \times X) \setminus \Delta$ is open, making Δ closed. See Figure 19.

Conversely, assuming that Δ is closed, we show that (X,τ) is Hausdorff. If $x,y \in X$ are distinct, $(x,y) \notin \Delta$. As the complement $(X \times X) \setminus \Delta$ is open, there exists an open subset $W \subseteq (X \times X) \setminus \Delta$ such that $(x,y) \in W$. By definition of product topology, there are open subsets $U,V \subseteq W$ such that $(x,y) \in U \times V \subseteq W$. As seen above, $U \times V \subseteq (X \times X) \setminus \Delta$ means that U and V are disjoint, as required.

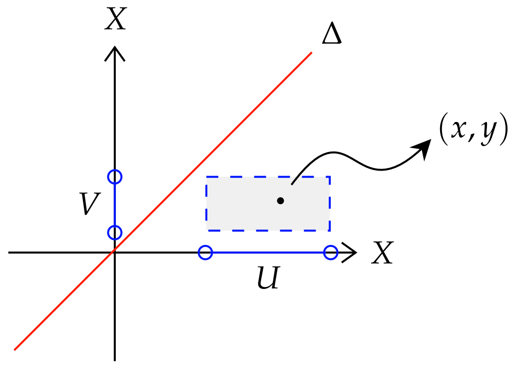


Figure 19: The Hausdorff condition and the diagonal subspace.

□

1.8 Compactness

In real analysis, the closed and bounded subsets of $\mathbb{R}$ play a central role—they are called "compact" and can be characterized in terms of open coverings.

In a metric space, we define a subset to be bounded if it is contained in some open ball (possibly with very large radius; the center of the ball doesn't matter). As boundedness doesn't make sense in arbitrary topological spaces, since there is no distance function available, we must take the latter characterization as the definition of compactness as we move on:

Definition 16 (Compactness)

A topological space (X,τ) is called **compact** if every open cover of X admits a finite subcover, that is, whenever $\mathcal{U} \subseteq \tau$ is such that $\bigcup \mathcal{U} = X$, there is a finite subcollection $\mathcal{F} \subseteq \mathcal{U}$ such that $\bigcup \mathcal{F} = X$.

Remark. While phrasing the definition of compactness in terms of subcollections $\mathcal{U}$ of τ is certainly clean, the reader not used to working with such set-theoretical language might have a difficult time parsing it. You may think of it in terms of indexed collections: whenever $\{U_\alpha\}_{\alpha \in A} \subseteq \tau$ has $\bigcup_{\alpha \in A} U_\alpha = X$, there is a finite subset $F \subseteq A$ such that $\bigcup_{\alpha \in F} U_\alpha = X$. Here, $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ and $\mathcal{F} = \{U_\alpha\}_{\alpha \in F}$. Rewrite some of the proofs coming next in terms of such indexed families and see for yourself: often the index set A and the indices $\alpha \in A$ amount to just extra notational baggage.

This definition of compactness can be difficult to digest at a first read. In my experience, it pays off to think about it "operationally", that is, what does this definition "do" and how do we use it in proofs? Say there is some process or algorithm we want to run on the space X, and we need it to "end in finite time". If we are able to achieve this on enough open subsets to cover X, and X is compact, then we are able to achieve it on X itself. This is vague, but later in this section we present a series of propositions containing basic properties of compact spaces—their proofs might help you understand what I am trying to get at here.

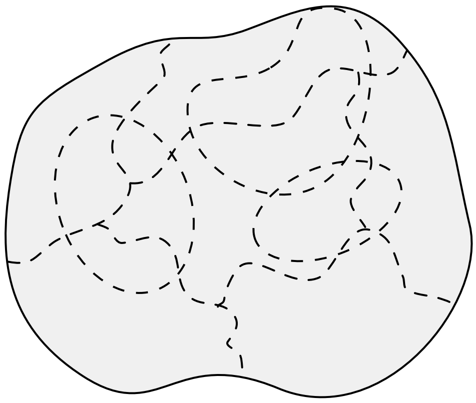


Figure 20: Definition of compactness via open covers.

Compact spaces should be thought of as "small", without being finite. Proving that a space is not compact is often much easier than proving that it is compact: the former requires just exhibiting *one* open cover without a finite subcover, while the latter requires showing that *any* open cover has a finite subcover. To renew your intuition for compactness, it is helpful to also see examples which are not compact.

Example 29 (Unbounded metric spaces are not compact)

The Euclidean space $\mathbb{R}^n$ is not compact: the open cover $\mathcal{U} = \{B_r(0) : r > 0\}$ admits no finite subcover—the union of any finite subcover of $\mathcal{U}$ is always equal to some open ball centered at the origin, and not the entire $\mathbb{R}^n$. This argument shows, more generally, that unbounded metric spaces are not compact; see also Proposition 21 ahead.

Example 30 ($[0,1)$ is not compact)

The half-open interval $[0,1)$, with its standard Euclidean topology, is not compact: the open cover $\mathcal{U} = \{[0,r) : r \in (0,1)\}$ of $[0,1)$ has no finite subcover, since the union of any finite subcollection of $\mathcal{U}$ is of the form $[0,R)$ for some $R \in (0,1)$, and thus $[0,R) \subsetneq [0,1)$.

The idea here is that if there is any point which "should be" in the given set, but is not there, we can try to set up an open cover $\mathcal{U}$ which "approaches" said point in such a way that removing even a single set from $\mathcal{U}$ makes it no longer be a cover. This is of course related to whether our given set is closed or not—we will revisit this idea ahead in Proposition 20.

The next example is particularly important, playing a role in the proof of the Heine-Borel Theorem (we will prove it later in this section), and so we phrase it as a lemma:

Lemma 4

Every bounded closed interval $[a,b] \subseteq \mathbb{R}$, equipped with its standard Euclidean topology, is compact.

Proof: We may assume without loss of generality that $[a,b] = [0,1]$. Let $\mathcal{U}$ be an open cover of $[0,1]$, and consider the set

$$I = \{t \in (0,1] : [0,t] \text{ is covered by finitely many elements of } \mathcal{U}\}.$$

The proof will be concluded once we show that $1 \in I$. First, note that $0 \in I$, so that I is non-empty, and therefore $b = \sup I$ exists. We will show that:

$$\begin{aligned} &\text{(i) } I \text{ is closed under taking limits of increasing sequences, implying that } b = \max I \text{ is actually an element of } I\text{, and;} \\ &\text{(ii) } b = 1. \end{aligned} \tag{1.13}$$

If $(t_n)_{n\geq 1} \subseteq I$ is an increasing sequence and $t_n \to t$, we will show that $t \in I$ as follows: let $U \in \mathcal{U}$ be such that $t \in U$, and use the definition of convergence to select $n_0 \geq 1$ such that $t_n \in U$ for every $n \geq n_0$; fix one such $t_n \in U$. Since $t_n \in I$, there is a finite subset $\mathcal{F} \subseteq \mathcal{U}$ such that $[0,t_n] \subseteq \bigcup \mathcal{F}$—it follows that $[0,t] \subseteq \bigcup(\mathcal{F} \cup \{U\})$, and this establishes (1.13-i).

As for (1.13-ii), assume by contradiction that $b < 1$, and let $U \in \mathcal{U}$ be such that $b \in U$. Then, take $b' \in U \cap (b,1]$, and note that whenever $\mathcal{F} \subseteq \mathcal{U}$ is a finite subset such that $[0,b] \subseteq \bigcup \mathcal{F}$, then $[0,b'] \subseteq \bigcup(\mathcal{F} \cup \{U\})$. This means that $b' > b$ and $b' \in I$, against maximality of b. ☐

Example 31 (Closed and bounded sets need not be compact in metric spaces)

Since $(a,b) \subseteq \mathbb{R}$ gets replaced with $B_r(p) \subseteq X$ in a metric space (X, d), and the closure of (a,b) equals $[a,b]$, which is compact, the reader might be led to think that (i) the closure of $B_r(p)$ is the closed ball $B_r[p] = \{x \in X : \mathrm{d}(x,p) \leq r\}$, and (ii) that $B_r[p]$ is compact (being obviously closed and bounded). Both claims are **false**, and it is a serious mistake to think so.

If X is any set with the discrete metric, then for any point $x \in X$ we have that $\overline{B_1(x)} = B_1(x) = \{x\}$, while $B_1[x] = X$ is not compact as soon as the set X is infinite (since the open cover $\{\{a\} : a \in X\}$ would have no finite open subcover).

In the next few results, we will often consider compactness of subspaces Y of a given topological space (X,τ). Technically, the sets V in open coverings of the topological space (Y, τ_Y) are in τ_Y and so must be written as $V = U \cap Y$ with $U \subseteq X$ open—exploring the eventual compactness of (X,τ) (or of subspaces larger than Y)

requires us to use Us instead of Vs. Going back and forth between Us and Vs is tiresome and potentially obfuscates the main ideas behind the argument being carried out; the collection of the Us should by all means be considered a valid open cover of Y even though their union in general strictly contains Y, as opposed to being equal to it as the union of the Vs would be.

Exercise 28 (Compactness is absolute)

Let (X, τ) be a topological space, and $Y \subseteq X$ be a subset. Let's say that Y is **compact in** X if for every collection $\mathscr{V}$ of open subsets of X such that $Y \subseteq \bigcup \mathscr{V}$, there is a finite subset $\mathscr{F} \subseteq \mathscr{V}$ such that $Y \subseteq \bigcup \mathscr{F}$. Show that Y is compact in X (in the above sense) if and only if the topological space (Y, τ_Y) is compact. This eliminates any ambiguity in saying "Y is compact".

The exercise below is a great opportunity for you to check if you understood the technical definition of compactness, and appreciate what Exercise 28 is doing (even if you have not solved it yet).

Exercise 29

Let (X, τ) be a topological space, and $K_1, \ldots, K_n \subseteq X$ be compact subsets. Show that the union $K = \bigcup_{i=1}^n K_i$ is compact.

Remark. If you have the time to spare, try solving this exercise both using and not using the result of Exercise 28, and compare both solutions.

Proposition 18 (Continuous images of compact spaces are compact)

Let (X, τ) and (Y, τ') be topological spaces, and $f \colon X \to Y$ be continuous and surjective. If (X, τ) is compact, then (Y, τ') must also be compact.

Proof: Let $\mathscr{V} \subseteq \tau'$ be an open cover of Y, and note that $f^{-1}(\mathscr{V}) = \{f^{-1}(V) : V \in \mathscr{V}\}$ is an open cover of X, since $\bigcup f^{-1}(\mathscr{V}) = f^{-1}(\bigcup \mathscr{V}) = f^{-1}(Y) = X$ and each such $f^{-1}(V)$ is open by continuity of f. By compactness of (X, τ), there exists a finite subset $\mathscr{F}' \subseteq f^{-1}(\mathscr{V})$ such that $\bigcup \mathscr{F}' = X$. Now we use surjectivity of f to show that the collection $\mathscr{F} = \{V \in \mathscr{V} : f^{-1}(V) \in \mathscr{F}'\}$ is a finite subcover of Y. Indeed, by surjectivity of f the assignment $\mathscr{F} \ni V \mapsto f^{-1}(V) \in \mathscr{F}'$ is injective[2], so that finiteness of $\mathscr{F}$ follows from the one of $\mathscr{F}'$; if $y \in Y$, we choose $x \in X$ so that $f(x) = y$, and then select some $f^{-1}(V) \in \mathscr{F}'$ with $x \in f^{-1}(V)$—now $y \in V$ with $V \in \mathscr{F}$ shows that $\bigcup \mathscr{F} = Y$, as required. □

[2] Namely, we use that if $f \colon X \to Y$ is a surjective function between sets, then $f(f^{-1}(B)) = B$ for every subset $B \subseteq Y$. Without surjectivity, we may only say that $f(f^{-1}(B)) = B \cap f(X)$.

Example 32 (Quotients of compact spaces are compact)

If (X, τ) is any compact topological space and $\sim$ is an equivalence relation on X, the quotient space $X/_\sim$ is also compact, being the image of X under the continuous mapping $\pi\colon X \to X/_\sim$.

Proposition 19

Closed subsets of compact spaces are compact.

Proof: Let (X, τ) be a compact topological space, and $C \subseteq X$ be closed. Let $\mathcal{U}$ be an open cover of C, and note that $\mathcal{U} \cup \{X \setminus C\}$ is an open cover of X. Compactness yields a finite subset $\mathcal{F} \subseteq \mathcal{U} \cup \{X \setminus C\}$ such that $\bigcup \mathcal{F} = X$, cf. Figure 21.

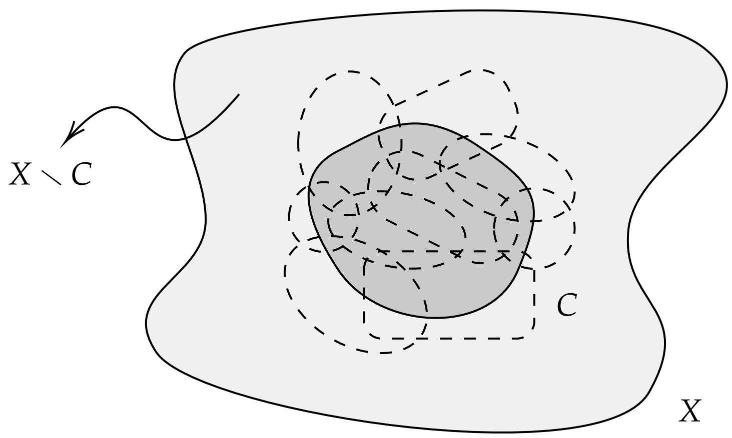


Figure 21: Producing an open cover of X from one of C, by adding $X \setminus C$.

As $(X \setminus C) \cap C = \varnothing$, we have that $C \subseteq \bigcup(\mathcal{F} \setminus \{X \setminus C\})$, making $\mathcal{F} \setminus \{X \setminus C\}$ the desired finite subcover of $\mathcal{U}$. $\square$

Proposition 20

Compact subspaces of Hausdorff spaces are closed.

Proof: Let (X, τ) be a Hausdorff space and $K \subseteq X$ be compact. We argue that the complement $X \setminus K$ is open as follows: let $z \in X \setminus K$ and, for every $x \in K$, take disjoint open sets $U_x, V_x \subseteq X$ such that $x \in U_x$ and $z \in V_x$; then $\{U_x : x \in K\}$ is an open cover of K, so that compactness of K yields a finite subset $F \subseteq K$ for which $\{U_x : x \in F\}$ covers K. See Figure 22.

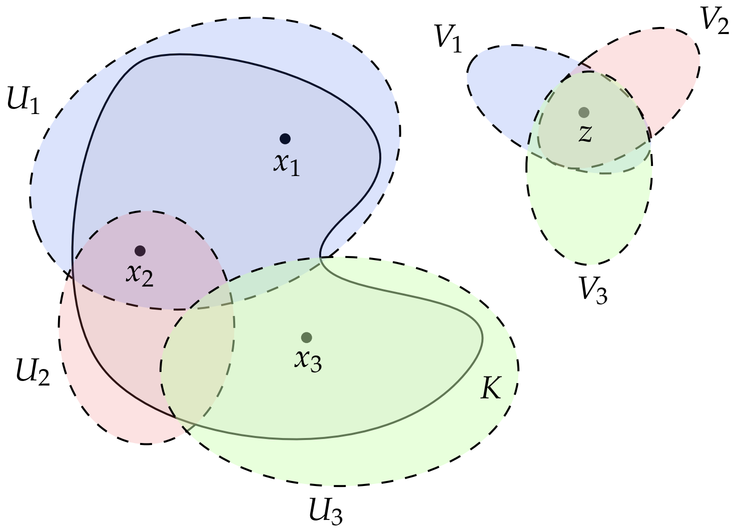


Figure 22: Compact subsets of Hausdorff spaces are closed.

As F is finite, $V = \bigcap_{x\in F} V_x$ is an open neighborhood of z, and by construction we have that $V \subseteq X \smallsetminus K$—if there is $y \in V \cap K$, then $y \in U_x$ for some $x \in F$, while at the same time $y \in V_x$ for this same x (as $V \subseteq V_x$), contradicting that $U_x \cap V_x = \varnothing$. This shows that z is an interior point of $X \smallsetminus K$, making $X \smallsetminus K$ open and therefore K closed. □

Proposition 21

Compact subspaces of metric spaces are bounded.

Proof: Let (X, d) be a metric space, and $K \subseteq X$ be compact. Choose $a \in X$ at will and note that $\{B_r(a) : r > 0\}$ is an open cover of K. Extract a finite subcover and denote by $r_1, \ldots, r_k > 0$ the radii of the remaining balls in the subcover. Now let $R = \max\{r_1, \ldots, r_k\} > 0$ and note that $K \subseteq B_R(a)$, making K bounded as required. □

Corollary 2 (The Extreme-Value Theorem)

If (X, τ) is a compact topological space and $f \colon X \to \mathbb{R}$ is a continuous function, there are $x_{\min}, x_{\max} \in X$ such that $f(x_{\min}) \leq f(x) \leq f(x_{\max})$ for every $x \in X$. In words,

Proof: The image $f(X) \subseteq \mathbb{R}$ is bounded, so $b = \sup f < +\infty$ exists. At the same time, $f(X)$ is closed, so $b = \max f \in f(X)$ and there is $x_{\max} \in X$ such that $f(x_{\max}) = b$, so that $f(x) \leq f(x_{\max})$ for every $x \in X$. As $\sup(-f) = -\inf f$, we may apply this argument to $-f$ as well to obtain the minimizer $x_{\min} \in X$. □

While Example 32 above regards compactness of quotient spaces, the next result is about compactness of product spaces, essentially giving us the best possible outcome:

Proposition 22 (Compact × compact = compact)

The product of two compact topological spaces is compact.

Remark. It turns out that *arbitrary* products (finite or not) of compact spaces are compact—this is called **Tychonoff's Theorem**, and it is a somewhat advanced result in general topology (two of its standard proofs use **nets** and **filters**, respectively; you will certainly come across these concepts when taking an actual topology class). For the fanatics, by the way, this result is equivalent to the axiom of choice.

Proof: Let (X,τ) and (Y,τ') be compact spaces, and $\{W_\alpha\}_{\alpha\in A}$ be an open cover of the product $X\times Y$. Without loss of generality, we may assume that $W_\alpha = U_\alpha \times V_\alpha$ for some open subsets $U_\alpha \subseteq X$ and $V_\alpha \subseteq Y$, for every $\alpha \in A$—having the same index set A for both resulting open covers $\{U_\alpha\}_{\alpha\in A}$ and $\{V_\alpha\}_{\alpha\in A}$ of X and Y will be particularly convenient for the following argument. For each $x\in X$, the set $\{x\}\times Y$ equipped with its subspace topology from $X\times Y$ is homeomorphic to Y, via $(x,y)\mapsto y$, and hence is compact. Then $\{\{x\}\times V_\alpha\}_{\alpha\in A}$ is an open cover of $\{x\}\times Y$, and so we may fix a finite subset $F_x\subseteq A$ for which $\{\{x\}\times V_\alpha\}_{\alpha\in F_x}$ covers $\{x\}\times Y$. This implies that each intersection $U_x = \bigcap_{\alpha\in F_x} U_\alpha$ is open in X, making $\{U_x\}_{x\in X}$ an open cover of X, cf. Figure 23.

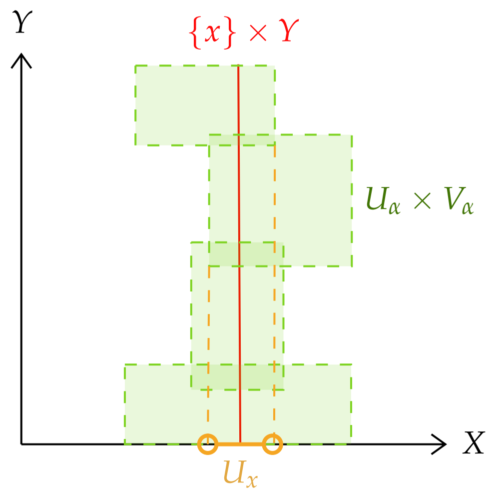


Figure 23: The "tube lemma".

Compactness of X now gives us points $x_1,\ldots,x_k\in X$ for which $\{U_{x_i}\}_{i=1}^k$ already covers X. Then $F=\bigcup_{i=1}^k F_{x_i}$ is a finite subset of A, and we finally claim that $\{W_\alpha\}_{\alpha\in F}$ covers $X\times Y$. Indeed: whenever $(x,y)\in X\times Y$, there is $i=1,\ldots,k$ such that $x\in U_{x_i}$, and hence there is $\alpha\in F_{x_i}$ (in particular, $\alpha\in F$) such that $y\in V_\alpha$; as $U_{x_i}\subseteq U_\alpha$, we conclude that $(x,y)\in U_\alpha\times V_\alpha = W_\alpha$, as required. □

With Proposition 22 in place, we may finally recover one of the main theorems from real analysis:

Corollary 3 (The Heine-Borel Theorem)

A subset $K \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof: If K is compact, then K is closed and bounded as $\mathbb{R}^n$ is a metric space, by Propositions 20 and 21. Conversely, boundedness of K allows us to place it in a box, i.e., write $K \subseteq [a_1, b_1] \times \cdots \times [a_n, b_n]$ for suitable closed intervals $[a_i, b_i] \subseteq \mathbb{R}$; each such interval is compact by Lemma 4, so their product is compact by Proposition 22 (and induction) and, finally, closedness of K allows us to invoke Proposition 21 to conclude that K is compact, as desired. $\square$

Example 33 (Spheres and real projective spaces are compact)

The unit sphere $\mathbb{S}^n \subseteq \mathbb{R}^{n+1}$ is compact, being closed and bounded. It now follows that the real projective space $\mathbb{RP}^n$ (Example 15) is also compact: the restriction of $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{RP}^n$ to $\mathbb{S}^n$ is continuous and still surjective onto $\mathbb{RP}^n$ (since $\pi(p) = \pi(p/\|p\|)$ for every $p \in \mathbb{R}^{n+1} \smallsetminus \{0\}$).

A topological space is called **Lindelöf** if every open cover has a countable subcover. This is a weakening of compactness. Every second-countable space is Lindelöf (can you prove it?), but not conversely, in general (e.g., Sorgenfrey line). For metric spaces (and in particular, manifolds), the converse does hold.

Exercise 30 (Compact × Lindelöf = Lindelöf)

Show that the product of a compact space with a Lindelöf space is Lindelöf.

Hint: adapt the proof of Proposition 22, but be careful—the product of two Lindelöf spaces need not be Lindelöf (e.g., Sorgenfrey line with itself).

We conclude this section with one application of the results obtained so far. When trying to prove that some function between topological spaces is a homeomorphism, the most challenging step is usually proving (once we know that the function is bijective) that the inverse function is also continuous. The result below provides us with a very useful shortcut:

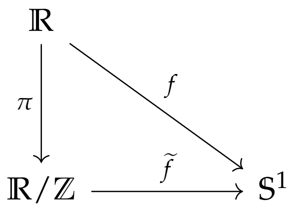
Theorem 2

Let (X, τ) and (Y, τ') be topological spaces, and $f\colon X \to Y$ be a continuous bijection. If (X, τ) is compact and (Y, τ') is Hausdorff, then f is a homeomorphism.

Proof: We just need to establish continuity of the inverse function $f^{-1}\colon Y \to X$. We use the following characterization of continuity: a function is continuous if and only if inverse images of closed sets are closed. So, let $C \subseteq X$ be closed. As (X, τ) is compact, Proposition 21 gives us that C is compact. By Proposition 18, $(f^{-1})^{-1}(C) = f(C)$ is compact. Finally, as (Y, τ') is Hausdorff, we conclude from Proposition 20 that $f(C)$ is closed, as required. $\square$

Example 34 ($\mathbb{R}/\mathbb{Z} \cong \mathbb{S}^1$)

Consider again the quotient space $\mathbb{R}/\mathbb{Z}$, from Example 14, with $x \sim y$ if and only if $x - y \in \mathbb{Z}$. The function $f\colon \mathbb{R} \to \mathbb{S}^1$ given by $f(x) = (\cos(2\pi x), \sin(2\pi x))$ is continuous and surjective, and it holds that $f(x) = f(y)$ if and only if $x - y \in \mathbb{Z}$ (by properties of trigonometric functions). This means that there is a function $\widetilde{f}$ for which the diagram

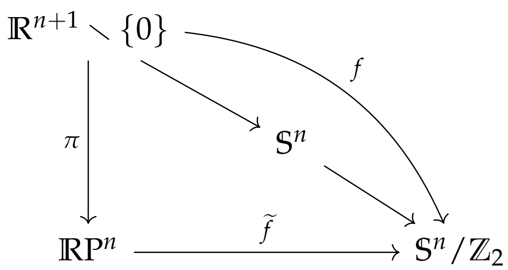


commutes, and by the above combined with the characteristic property of the quotient topology of $\mathbb{R}/Z$, it follows that $\widetilde{f}$ is a continuous bijection.

As the restriction of $\pi\colon \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ to the interval $[0,1]$ is still surjective onto $\mathbb{R}/\mathbb{Z}$ (since every equivalence class has a representative in $[0,1)$), compactness of $[0,1]$ implies the one of $\mathbb{R}/\mathbb{Z}$. Thus, as $\mathbb{S}^1$ is Hausdorff, Theorem 2 says that $\widetilde{f}$ is a homeomorphism. A similar argument says that $\mathbb{R}^n/\mathbb{Z}^n \cong \mathbb{T}^n$.

Example 35 ($\mathbb{S}^n/\mathbb{Z}_2 \cong \mathbb{RP}^n$)

Consider again the quotient space $\mathbb{S}^n/\mathbb{Z}_2$, as well as the real projective space $\mathbb{RP}^n$, and the function $f\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{S}^n/\mathbb{Z}_2$ given by $f(p) = \{p/\|p\|, -p/\|p\|\}$. Note that f is continuous, being a composition $\mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{S}^n \to \mathbb{S}^n/\mathbb{Z}_2$ of continuous functions—the first arrow, of course, is given by $p \mapsto p/\|p\|$. As $f(\lambda p) = f(p)$ for every $p \in \mathbb{R}^{n+1} \smallsetminus \{0\}$ and $\lambda \in \mathbb{R}$, the function f survives in the quotient: that is, there is a function $\widetilde{f}\colon \mathbb{RP}^n \to \mathbb{S}^n/\mathbb{Z}_2$ such that $\widetilde{f} \circ \pi$, where $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{RP}^n$ is the natural projection (taking a point p to the line $\mathbb{R}p$ it spans):



Continuity of $\widetilde{f}$ follows from the one of f, via the characteristic property of the quotient topology (of $\mathbb{RP}^n$, not of $\mathbb{S}^n/\mathbb{Z}_2$). The image of $\widetilde{f}$ equals the one of f, as π is surjective, and so surjectivity of $\widetilde{f}$ follows from the one of f. But we claim that $\widetilde{f}$ is injective as well. Indeed, assume that $\widetilde{f}(\mathbb{R}p) = \widetilde{f}(\mathbb{R}q)$. Then $\{p/\|p\|, -p/\|p\|\} = \{q/\|q\|, -q/\|q\|\}$, meaning that $p/\|p\| = \pm q/\|q\|$ for some choice of sign $\pm$. In any case, $q = \lambda p$ for $\lambda = \pm\|q\|/\|p\|$, so that $\mathbb{R}p = \mathbb{R}q$.

Thus, $\widetilde{f}$ is a continuous bijection. But $\mathbb{RP}^n$ is compact by Example 33 and $\mathbb{S}^n/\mathbb{Z}_2$ is Hausdorff by Example 28, so Theorem 2 allows us to conclude that $\widetilde{f}$ is a homeomorphism. In summary, we can see $\mathbb{RP}^n$ geometrically as a sphere with pairs of antipodal points identified.

1.9 Connectedness

One last important topological property there is still left to discuss is connectedness. It is the way we make precise the notion of a topological space consisting of one single piece.

Definition 17 (Connected spaces)

A topological space (X, τ) is **disconnected** if it can be written as the disjoint union of two proper open subsets. We then say that X is **connected** if it is not disconnected.

In other words, if (X, τ) is connected, the only subsets of X which are simultaneously open and closed are $\varnothing$ and X itself. Otherwise, if A is some such set, $X = A \cup (X \smallsetminus A)$ would show that (X, τ) is disconnected.

Example 36

Consider the real line $\mathbb{R}$, equipped with its Euclidean topology. You might have seen in real analysis that a subset of $\mathbb{R}$, with its subspace topology, is connected if and only if it is an interval. In particular, $\mathbb{R}$ is connected and, more generally, $\mathbb{R}^n$ is connected.

Example 37

Let X be any set. Then $(X, \tau_{\text{disc.}})$ is connected if and only if X is either empty or a singleton. At the other extreme, $(X, \tau_{\text{ch.}})$ is always connected.

Example 38

The Sorgenfrey line $(\mathbb{R}, \tau_S)$ is disconnected, since $\mathbb{R} = (-\infty, 0) \cup [0, \infty)$, and both intervals $(-\infty, 0)$ and $[0, \infty)$ are disjoint and Sorgenfrey-open.

Exercise 31

Is the "vertical" topology from Example 12 connected?

Exercise 32 (Connectedness is absolute)

Like compactness, connectedness is also an "absolute" notion, in the sense similar to the one given in Exercise 28. Turn this into a precise statement (and prove it).

Usually, when presenting standard results regarding connectedness, the arguments all follow set-theoretic calculations with unions and intersections. Here, we take a more "categorical" approach, focusing on functions rather than sets. The next result is what allows us to get away with it:

Theorem 3 (Connectedness via functions)

Let (X,τ) be a topological space. Then X is connected if and only if every continuous function $f\colon X \to \{0,1\}$ is constant, where $\{0,1\}$ is equipped with its discrete topology.

Proof: If X is disconnected and written as $X = U \cup V$, for $U, V \subseteq X$ open and disjoint, both characteristic functions $\chi_U, \chi_V\colon X \to \{0,1\}$ are surjective and continuous (as both $\chi_U^{-1}(1) = \chi_V^{-1}(0) = U$ and $\chi_U^{-1}(0) = \chi_V^{-1}(1) = V$ are open). Conversely, if there is a surjective continuous function $f\colon X \to \{0,1\}$, then $U = f^{-1}(0)$ and $V = f^{-1}(1)$ are open, disjoint, nonempty, and such that $X = U \cup V$. ☐

We assume without further comment that $\{0,1\}$ is always equipped with its discrete topology.

Proposition 23 (Intermediate Value Theorem)

Let X and Y be topological spaces, and $\varphi\colon X \to Y$ be a continuous function. If X is connected, then the image $\varphi(X) \subseteq Y$ is also connected.

Proof: Let $f\colon \varphi(X) \to \{0,1\}$ be a continuous function, and note that the composition $f \circ \varphi\colon X \to \{0,1\}$ is also continuous. By connectedness of X, it follows that $f \circ \varphi$ is constant, and thus f is constant (on the image of φ) as required. Hence, $\varphi(X)$ is connected. ☐

Remark. In the above proof, Exercise 32 would allow us to assume, without loss of generality, that φ is surjective from the start.

Example 39 (Quotients of connected spaces are connected)

If (X,τ) is any connected topological space and $\sim$ is an equivalence relation on X, the quotient space $X/_\sim$ is also connected, being the image of X under the continuous mapping $\pi\colon X \to X/_\sim$. In particular, the real projective space $\mathbb{RP}^n$ is connected, being the image of the connected space $\mathbb{R}^{n+1} \smallsetminus \{0\}$ under the continuous projection $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{RP}^n$.

Exercise 33 (Connected × connected = connected)

Prove that the product of two connected topological spaces is also connected.

Proposition 24

Let X be a topological space and $\{U_\alpha\}_{\alpha \in A}$ be a family of connected subspaces of X such that $\bigcap_{\alpha \in A} U_\alpha \neq \varnothing$. Then, the union $\bigcup_{\alpha \in A} U_\alpha$ is connected.

Proof: Let $f : \bigcup_{\alpha \in A} U_\alpha \to \{0, 1\}$ be a continuous function. For each $\alpha \in A$, as U_α is connected, the restriction $f|_{U_\alpha} : U_\alpha \to \{0, 1\}$ is continuous, and hence constant, say $c_\alpha \in \{0, 1\}$. Now it is easy to see that c_α is in fact the same constant for all $\alpha \in A$: let $z \in \bigcap_{\alpha \in A} U_\alpha$ and note that $c_\alpha = f(z)$. Thus f is constant and $\bigcup_{\alpha \in A} U_\alpha$ is connected. □

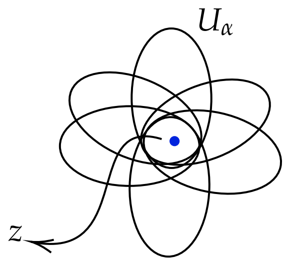


Figure 24: Picking $z \in \bigcap_{\alpha \in A} U_\alpha$.

Here's another concrete application of Theorem 3:

Example 40 (Connectedness of $[0, 1]$ via functions)

The interval $[0, 1] \subseteq \mathbb{R}$, with its Euclidean topology, is connected.

Let $f : [0, 1] \to \{0, 1\}$ be continuous, and consider the set

$$I = \{t \in [0, 1] : f(s) = f(0) \text{ for all } s \in [0, t]\}.$$

It is clear I is an interval with $0 \in I$ (if $t \in I$ and $t' \in [0, t]$, then $t' \in I$), allowing us to consider $b = \sup I$. We first claim that $b \in I$. Indeed, let $(t_n)_{n \geq 1}$ be a sequence in I with $t_n \nearrow b$, and fix $s \in [0, b]$. If $s = b$, then making $n \to \infty$ in $f(t_n) = f(0)$ leads to $f(b) = f(0)$, while if $s \in [0, b)$ there is $n \geq 1$ sufficiently large with $t_n \in I$ and $s < t_n$, so that $s \in I$ as well. Finally, we claim that $b = 1$. Continuity of f and openness of $f^{-1}(f(0))$ yields $\delta > 0$ such that, if $x \in (b - \delta, b + \delta) \cap [0, 1]$, then $f(x) \in f^{-1}(f(0))$. It it were $b < 1$ instead, we could choose $\varepsilon > 0$ such that $b + \varepsilon \in (b - \delta, b + \delta) \cap [0, 1)$, showing that $b + \varepsilon \in I$, contradicting the definition of b as the supremum of I. Thus $I = [0, 1]$ and f is constant, showing that $[0, 1]$ is connected.

Similar arguments show that $(0, 1)$ and $[0, 1)$ are also connected.

To obtain more explicit examples of connected spaces from known ones, we may directly apply Proposition 24—there is a lot of room for creativity when applying it. When combined with Example 39, it shows that spaces obtained by "gluing" connected spaces are again connected.

Example 41 (Balls in $\mathbb{R}^n$ are connected)

Given any $p \in \mathbb{R}^n$ and $r > 0$, the open ball $B_r(p)$ is the union of radial half-open line segments (all of which are connected) of length r starting at p and containing it. By Proposition 24, $B_r(p)$ is connected. A similar argument shows that closed balls are also connected (alternatively, see Exercise 35 below). More generally, any convex subset of $\mathbb{R}^n$ is connected.

Example 42 (Spheres are connected)

For each integer $n \geq 1$, the unit sphere $\mathbb{S}^n \subseteq \mathbb{R}^{n+1}$ is connected. Indeed, both hemispheres $U_\pm = \{(x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} : x_0 = \pm(1 - x_1^2 - \cdots - x_n^2)^{1/2}\}$ are connected, being homeomorphic to the closed unit ball in $\mathbb{R}^n$.

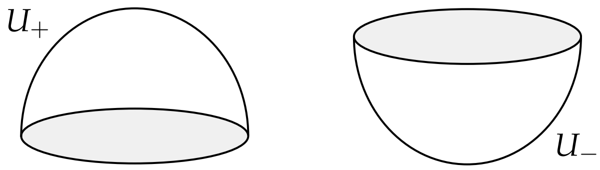


Figure 25: The north and south hemispheres of $\mathbb{S}^n$.

As $U_+ \cap U_- \neq \varnothing$ and $U_+ \cup U_- = \mathbb{S}^n$, we may again apply Proposition 24.

You can practice further in the next three exercises how to use Theorem 3 to prove results on connectedness without excessively relying on set-algebra.

Exercise 34

Let X be a topological space, and $\{U_n\}_{n\geq 1}$ be a sequence of connected subspaces of X such that, for every $n \geq 1$, we have $U_n \cap U_{n+1} \neq \varnothing$.

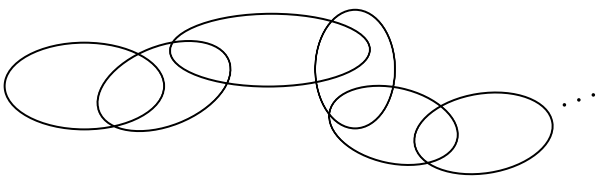


Figure 26: A chained sequence of connected sets.

Show that the union $\bigcup_{n\geq 1} U_n$ is connected.

In general, interiors and boundaries of connected sets are not themselves connected. For interior, consider the union of two closed balls in $\mathbb{R}^2$ which are tangent to each other; for boundaries, just take a bounded interval in $\mathbb{R}$. The story is different with closures—you can explore that in the next exercise.

Exercise 35

Let (X, τ) be a topological space, $U \subseteq X$ be a connected subspace, and let $V \subseteq X$ be a subspace such that $U \subseteq V \subseteq \overline{U}$. Show that V is connected as well. (In particular, closures of connected sets are connected.)

Exercise 36 (Path-connectedness implies connectedness)

A topological space (X, τ) is called **path-connected** if whenever two points $x, y \in X$ are given, there is a continuous function $\gamma\colon [0,1] \to X$ (that is, a "path") such that $\gamma(0) = x$ and $\gamma(1) = y$. Prove that every path-connected space is connected.

Hint: There are multiple ways to skin this cat. If you don't want to use Theorem 3 you can, for example, mimic the idea from Example 41.

We conclude this section illustrating a general technique for dealing with proofs involving connected spaces. Often, one knows that a certain property holds locally (that is, on sufficiently small open subsets around each point), and wants to prove that it holds globally. A very effective strategy is to consider the set consisting of the points near which the property in question holds, and show that it is non-empty, open, and closed. Connectedness then implies that this set must be the entire space. Here is an example:

Proposition 25

Let (X, τ) be a topological space, and assume that it is **locally path-connected**, that is, for every $x \in X$ there is $U \in \tau$ such that U is path-connected (when equipped with its subspace topology). If in addition (X, τ) is connected, then (X, τ) is itself path-connected.

Proof: Fix a point $z \in X$, and consider the set A consisting of all $x \in X$ for which there is a continuous path $\gamma\colon [0,1] \to X$ such that $\gamma(0) = z$ and $\gamma(1) = x$. Considering the constant curve $c_z\colon [0,1] \to X$, given by $c_z(t) = z$ for all $t \in [0,1]$, shows that $z \in A$, so that $A \neq \varnothing$.

We now claim that A is open. If $x \in A$ and $U \subseteq X$ is an open and path-connected subset of X with $x \in U$, we in fact have that $U \subseteq A$. Indeed, if $y \in U$ and $\gamma\colon [0,1] \to X$ goes from z to x, while $\alpha\colon [0,1] \to U$ goes from x to y, the curve $\gamma * \alpha\colon [0,1] \to X$ defined by

$$(\gamma * \alpha)(t) = \begin{cases} \gamma(2t), & \text{if } 0 \leq t \leq 1/2, \\ \alpha(2t-1), & \text{if } 1/2 \leq t \leq 1, \end{cases} \tag{1.14}$$

is continuous[3] and has $(\gamma * \alpha)(0) = z$ and $(\gamma * \alpha)(1) = y$, showing that $y \in A$.

Finally, we argue that A is closed, by showing that $X \smallsetminus A$ is also open. If $x \in X \smallsetminus A$ and $U \subseteq X$ is an open and path-connected subset of X with $x \in U$, perhaps not

[3] This is not entirely trivial, and follows from the famous "pasting lemma".

surprisingly, we have that $U \subseteq X \smallsetminus A$. If there was a point $y \in U \cap A$, then we could join y to x with $\beta\colon [0,1] \to U$, and then $\gamma * \beta\colon [0,1] \to X$ defined as above would joint z to x, contradicting that $x \in X \smallsetminus A$.

As (X, τ) is connected, it follows that $A = X$, as required. □

Try experimenting with that technique yourself:

Exercise 37 (Locally constant functions on connected spaces are constant)

Let (X, τ) be a topological space, and $f\colon X \to \mathbb{R}$ be a continuous function. Assume that f is **locally constant**: for every $x \in X$ there is $U \in \tau$ such that $x \in U$ and $f|_U\colon U \to \mathbb{R}$ is constant. Show that if (X, τ) is connected, then f is constant.

Hint: Fix $z \in X$ and let $A = \{x \in X : f(x) = f(z)\}$.

From here onwards, we will no longer denote a topological space by (X, τ), writing just X, with the topology τ being implicitly understood—in the same way that we simply say that V is a vector space, instead of writing the full tuple $(V, \mathbb{K}, +, \cdot)$ (or a group is G instead of $(G, \cdot)$, etc.).

2 Multivariable calculus: derivatives as linear transformations and the IFTs

2.1 Derivatives as linear transformations

We already have a good grasp of what continuity means for arbitrary topological spaces. Now, restricting ourselves to Euclidean spaces, we turn our attention to differentiability. We assume that the reader has had some exposure to multivariable calculus beyond $\mathbb{R}^2$ and $\mathbb{R}^3$, but not any familiarity with the notion of derivative as a linear transformation. In any case the core idea, as always, is the one of "best linear approximation":

Definition 18 (Total derivative)

Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be a function, where $U \subseteq \mathbb{R}^n$ is open, and $p \in U$. We say that f is **differentiable at** p if there is a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^k$ such that

$$\lim_{h\to 0} \frac{\|f(p+h) - f(p) - Th\|}{\|h\|} = 0. \tag{2.1}$$

Such T, if it exists, is unique and denoted by $\mathrm{D}f(p)\colon \mathbb{R}^n \to \mathbb{R}^k$. It is called the **total derivative** of f at p. We say that f is **differentiable** if it is differentiable at every point $p \in U$.

To validate the above definition, we need to prove that T is indeed unique.

Proof: Let T_1 and T_2 be two total derivatives of f at p, that is, assume that both of them satisfy (2.1). By the triangle inequality, we have that

$$\begin{aligned} 0 &\le \lim_{h\to 0} \frac{\|T_1(h) - T_2(h)\|}{\|h\|} \\ &\le \lim_{h\to 0} \frac{\|f(p+h) - f(p) - T_1(h)\|}{\|h\|} + \lim_{h\to 0} \frac{\|f(p+h) - f(p) - T_2(h)\|}{\|h\|} \\ &= 0 + 0 = 0. \end{aligned}$$

Hence

$$\lim_{h\to 0} \frac{\|T_1(h) - T_2(h)\|}{\|h\|} = 0. \tag{2.2}$$

Now, we fix an arbitrary vector $v \in \mathbb{R}^n \smallsetminus \{0\}$, and argue that $T_1(v) = T_2(v)$ as follows: for any $t \in \mathbb{R} \smallsetminus \{0\}$, linearity of T_1 and T_2 implies that

$$\frac{\|T_1(v) - T_2(v)\|}{\|v\|} = \frac{\|T_1(tv) - T_2(tv)\|}{\|tv\|}, \tag{2.3}$$

allowing us to set $h = tv$ and pass to the limit $t \to 0$ in (2.3), so that (2.2) finally yields $\|T_1(v) - T_2(v)\|/\|v\| = 0$. This obviously implies that $T_1(v) = T_2(v)$. □

Exercise 38

With the notation from Definition 18, show that if f is differentiable at p, then f is continuous at p.

Once we know that such a function f is differentiable at p, we may compute $\mathrm{D}f(p)$ like a directional derivative:

$$\mathrm{D}f(p)v = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} f(p+tv). \tag{2.4}$$

The notation $\mathrm{D}f(p)v$, in my experience teaching this topic, can be a huge source of confusion. The original function is f, but $\mathrm{D}f$ is not a linear transformation; instead, it is the function which assigns to p the linear transformation $\mathrm{D}f(p)\colon \mathbb{R}^n \to \mathbb{R}^k$. Then $\mathrm{D}f(p)v \in \mathbb{R}^k$ is the vector obtained by applying the linear transformation $\mathrm{D}f(p)$ to the vector v. To justify (2.4) (its right side is usually denoted by $(\partial f/\partial v)(p)$), assume for simplicity that $\|v\| = 1$, and compute the limit (2.1) along the line spanned by v (this is allowed because we assume that f is differentiable at p, so (2.1) may be computed along any direction):

$$\begin{aligned}\lim_{t\to 0} \frac{\|f(p+tv) - f(p) - t(\partial f/\partial v)(p)\|}{\|tv\|} &= \lim_{t\to 0}\left\| \frac{f(p+tv)-f(p)}{t} - \frac{\partial f}{\partial v}(p)\right\| \\ &= \left\| \lim_{t\to 0} \frac{f(p+tv)-f(p)}{t} - \frac{\partial f}{\partial v}(p)\right\| \\ &= 0.\end{aligned} \tag{2.5}$$

In particular, we have that $\mathrm{D}f(p)e_j = (\partial f/\partial x_j)(p)$ for every $j = 1,\ldots,n$, where $e_1,\ldots,e_n$ is the standard basis of $\mathbb{R}^n$ and $(\partial f/\partial x_j) \in \mathbb{R}^k$ is the vector containing the partial derivatives relative to x_j of all k components of f. This means that $\mathrm{D}f(p)$ does contain information about all partial derivatives of all components of f at p and, in particular, it tells us how to compute the matrix representation of $\mathrm{D}f(p)$ relative to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^k$. It is given by

$$Jf(p) = \left.\begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots & \partial f_2/\partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_k/\partial x_1 & \partial f_k/\partial x_2 & \cdots & \partial f_k/\partial x_n \end{bmatrix}\right|_p, \tag{2.6}$$

where $f = (f_1,\ldots,f_k)$. The matrix $Jf(p) \in \mathbb{R}^{k\times n}$ (it is not in $\mathbb{R}^{n\times k}$!) is called the **Jacobian matrix of f at** p. We will often identify it with $\mathrm{D}f(p)$ itself, but what we care about here is that there are two possible algorithms to compute it:

(1) Find the gradients $\nabla f_1(p),\ldots,\nabla f_k(p) \in \mathbb{R}^n$ and place them as the rows of $Jf(p)$, or;

(2) Compute the vector partial derivatives $(\partial f/\partial x_1)(p), \ldots, (\partial f/\partial x_k)(p) \in \mathbb{R}^k$ and put them into the columns of $Jf(p)$.

The other reason why (2.4) is very relevant is that, even though we might not know a priori whether f is differentiable at p, it gives us the unique candidate to the quantity Th to be tested when computing the limit (2.1); of course, we replace v with h in (2.4). More precisely, we first compute the right side of (2.4) with v replaced with h, then plug the result into (2.1): if the limit is zero, then f is differentiable and our guess for the definition of $\mathrm{D}f(p)$ is correct, while if the limit does not exist, then f is not differentiable at p (because the only viable candidate failed). We illustrate this idea with the examples below:

Example 43 (Differentiable curves)

Let $I \subseteq \mathbb{R}$ be an open interval, and consider a curve $\alpha\colon I \to \mathbb{R}^n$, explicitly written as $\alpha(t) = (x_1(t), \ldots, x_n(t))$, for suitable functions $x_i\colon I \to \mathbb{R}$. If each x_i is differentiable (in the sense of single-variable calculus) at a point $t_0 \in I$, then α is differentiable at t_0 in the sense of Definition (18), with the corresponding linear transformation $\mathrm{D}\alpha(t_0)\colon \mathbb{R} \to \mathbb{R}^n$ given by $\mathrm{D}\alpha(t_0)h = h\alpha'(t_0)$, where $\alpha'(t_0) = (x_1'(t), \ldots, x_n'(t_0))$ is the classical velocity vector of α. In particular, we see that $\mathrm{D}\alpha(t_0)$ is the zero transformation if and only if $\alpha'(t_0)$ is the zero vector; outside of this degenerate case, $\mathrm{D}\alpha(t_0)$ is essentially a parametrization for the line spanned by $\alpha'(t_0)$ (that is, the tangent line to α at t_0, shifted to pass through the origin). To prove that $\mathrm{D}\alpha(t_0)$ is indeed what we are claiming, we compute:

$$
\begin{aligned}
\lim_{h\to 0} \frac{\|\alpha(t_0+h) - \alpha(t_0) - h\alpha'(t_0)\|}{|h|} &= \lim_{h\to 0} \left\| \frac{\alpha(t_0+h) - \alpha(t_0)}{h} - \alpha'(t_0) \right\| \\
&= \lim_{h\to 0} \left(\sum_{i=1}^n \left(\frac{x_i(t_0+h) - x_i(t_0)}{h} - x_i'(t_0) \right)^2 \right)^{1/2} \\
&= \left(\sum_{i=1}^n (x_i'(t_0) - x_i'(t_0))^2 \right)^{1/2} \\
&= 0.
\end{aligned}
$$

In particular, if $n = 1$ and we consider a single-variable real function $f\colon I \to \mathbb{R}$, differentiable at a point $x_0 \in I$, its total derivative $\mathrm{D}f(x_0)\colon \mathbb{R} \to \mathbb{R}$ must be given by multiplication by a scalar (being a linear operator in a one-dimensional vector space): this scalar is the classical derivative $f'(x_0)$. The Jacobian matrix is a 1×1 matrix, containing the number $f'(x_0)$ and nothing else.

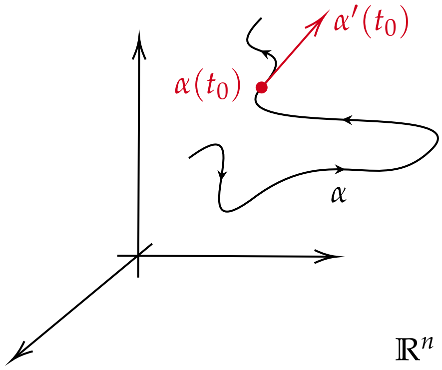


Figure 27: The velocity vector of α at t_0.

Example 44 (Differentiable scalar fields)

Let $U \subseteq \mathbb{R}^n$ be open, and consider a function $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$, differentiable at a point $p \in U$. The total derivative $\mathrm{D}f(p)\colon \mathbb{R}^n \to \mathbb{R}$ is a linear functional, and therefore it may be represented as a vector: the gradient. Namely, we have that $\mathrm{D}f(p)v = \langle \nabla f(p), v\rangle$, where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^n$.

Exercise 39

Verify in the above example that the expression given for $\mathrm{D}f(p)v$ is indeed correct.

Example 45 (Linear transformations)

Let $T\colon \mathbb{R}^n \to \mathbb{R}^k$ be a linear transformation. What is the best linear approximation to something which is already linear? Obviously, itself. So, we claim that T is differentiable everywhere, with total derivative given by $\mathrm{D}T(p) = T$ for each $p \in \mathbb{R}^n$ (that is, $\mathrm{D}T(p)v = T(v)$ for every $v \in \mathbb{R}^n$). Indeed, by linearity of T we have that

$$\frac{\|T(p+h) - T(p) - T(h)\|}{\|h\|} = 0$$

even before passing to the limit.

Example 46 (Bilinear forms)

Let $B\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ be a bilinear form. As

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} B(x+tv, y+tw) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} B(x,y) + t(B(x,w) + B(v,y)) + t^2 B(v,w) \\ &= B(x,w) + B(v,y),\end{aligned}$$

we claim that B is differentiable, with total derivative $\mathrm{D}B(x,y)\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ given by $\mathrm{D}B(x,y)(v,w) = B(x,w) + B(v,y)$ for all $(v,w) \in \mathbb{R}^n \times \mathbb{R}^m$. Indeed, setting $\|B\| = \sup_{\|v\|=\|w\|=1} \|B(v,w)\| < \infty$, we have that

$$\begin{aligned}0 \leq \frac{\|B(x+v, y+w) - B(x,y) - (B(x,w) + B(v,y))\|}{\sqrt{\|v\|^2 + \|w\|^2}} &= \frac{\|B(v,w)\|}{\sqrt{\|v\|^2 + \|w\|^2}} \\ &\leq \frac{\|B\|\|v\|\|w\|}{\sqrt{\|v\|^2 + \|w\|^2}}\end{aligned}$$

Making $(v,w) \to (0,0)$ in the above and using that $\|w\|/\sqrt{\|v\|^2 + \|w\|^2} \to 0$, it follows that the proposed expression for $\mathrm{D}B(x,y)$ does verify (2.1), which means that it is correct.

Now, let's see a couple of more concrete examples:

Example 47

Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x,y) = (x^2 - y^2, 2xy)$. We will soon have a quick way of checking whether things are differentiable without resorting to the limit-definition all the time. Taking this for granted here, we have that

$$\mathrm{D}f(x,y) = \begin{bmatrix} 2x & -2y \\ 2y & 2x \end{bmatrix}.$$

Identifying elements of $\mathbb{R}^2$ with column vectors we have, for instance, that

$$\mathrm{D}f(2,1)(3,4) = \begin{bmatrix} 4 & -2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 18 \end{bmatrix}.$$

Example 48

Let $f\colon \mathbb{R}^3 \to \mathbb{R}^4$ be given by $f(x,y,z) = (xz, yz, xy, xyz)$. We compute its total derivative at an arbitrary point $(x,y,z) \in \mathbb{R}^3$ as

$$\mathrm{D}f(x,y,z) = \begin{bmatrix} z & 0 & x \\ 0 & z & y \\ y & x & 0 \\ yz & xz & xy \end{bmatrix}.$$

To compute, for instance, $\mathrm{D}f(1,1,0)(3,2,2) \in \mathbb{R}^4$, we evaluate

$$\mathrm{D}f(1,1,0)(3,2,2) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 5 \\ 2 \end{bmatrix}.$$

In other words, $\mathrm{D}f(1,1,0)(3,2,2) = (2,2,5,2)$.

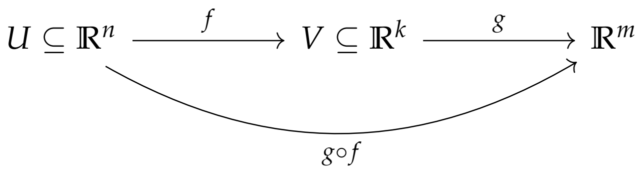
Exercise 40

Compute $\mathrm{D}f(2,3,4,4)(0,1,1,0)$, for the function $f\colon \mathbb{R}^4 \to \mathbb{R}^2$ explicitly defined by $f(x,y,z,w) = (xy^2 + zw, ye^w + xz^2)$.

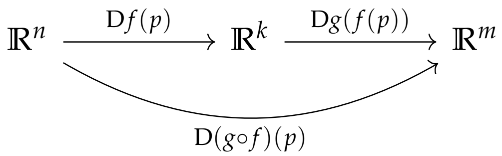
Theorem 4 (Chain rule)

Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^k$ be open sets, and $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ and $g\colon V \subseteq \mathbb{R}^k \to \mathbb{R}^m$ be functions such that $f(U) \subseteq V$, with f differentiable at some point $p \in U$ and g differentiable at the point $f(p) \in V$. Then $g \circ f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at p, and the relation $\mathrm{D}(g \circ f)(p) = \mathrm{D}g(f(p)) \circ \mathrm{D}f(p)$.

Remark. The slogan is "the derivative of a composition is the composition of the derivatives", once the points of evaluation are correct (for example, $\mathrm{D}g(p)$ does not make sense since p is in the domain of f, not of g). Consider the diagram



and now "apply D to it":



Recall that in the case where $n - k = 1$, the big idea was that one wants to write

$$\frac{g(f(x+h)) - g(f(x))}{h} = \frac{g(f(x+h)) - g(f(x))}{f(x+h) - f(x)} \cdot \frac{f(x+h) - f(x)}{h} \tag{2.7}$$

and take the limit as $h \to 0$ along with continuity of f at the point x to conclude that $(g \circ f)'(x) = g'(f(x))f'(x)$, but this argument fails if f is constant near x (causing a division by zero in (2.7)). This same subtlety must be dealt with in the general case; see [24, Theorem 2.2, p. 19] or [19, Theorem 7.1, p. 56] for more details.

Corollary 4 (Other operational rules)

Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ and $g\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be functions, where $U \subseteq \mathbb{R}^n$ is open, and $p \in U$. If f and g are differentiable at p, the following functions are also differentiable at p and have the indicated derivatives at p:

(i) $f + g$ (if $k = m$), with $\mathrm{D}(f+g)(p) = \mathrm{D}f(p) + \mathrm{D}g(p)$;

(ii) $\langle f, g\rangle$ (if $k = m$), with $\mathrm{D}(\langle f, g\rangle)(p)v = \langle \mathrm{D}f(p)v, g(p)\rangle + \langle f(p), \mathrm{D}g(p)\rangle$;

(iii) fg (if $k = 1$), with $\mathrm{D}(fg)(p) = g(p)\,\mathrm{D}f(p) + f(p)\,\mathrm{D}g(p)$;

(iv) f/g (if $m = 1$), with $\mathrm{D}(f/g)(p) = (g(p)\mathrm{D}f(p) - f(p)\,\mathrm{D}g(p))/g(p)^2$, provided that $g(p) \neq 0$.

Proof: The argument is similar to the one used in the proof of Proposition 4 (p. 18). First, the addition function $a\colon \mathbb{R}^{2k} = \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ is linear, and $f+g = a\circ(f,g)$, so the chain rule combined with Example 45 yields (i). Considering the bilinear form $\langle\cdot,\cdot\rangle\colon \mathbb{R}^k\times\mathbb{R}^k \to \mathbb{R}$, with $\langle f,g\rangle = \langle\cdot,\cdot\rangle\circ(f,g)$, and combining Example 46 with the chain rule, gives (ii). Noting that $m\colon \mathbb{R}\times\mathbb{R}^k \to \mathbb{R}^k$ given by $m(\lambda,x) = \lambda x$ is bilinear and that the derivative of $d\colon \mathbb{R}^k \times (\mathbb{R}\smallsetminus\{0\}) \to \mathbb{R}^k$ defined by $d(x,\lambda) = m(1/\lambda, x)$ is given by $\mathrm{D}d(x,\lambda)(\mu,y) = (\lambda y - \mu x)/\lambda^2$, (iii) and (iv) follow from the chain rule. □

Exercise 41

Let $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ satisfy the conditions $f(0,0,0) = (1,2)$ and

$$\mathrm{D}f(0,0,0) = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

If $g\colon \mathbb{R}^2\to\mathbb{R}^2$ is defined by $g(x,y) = (x+2y+1, 3xy)$, compute $\mathrm{D}(g\circ f)(0,0,0)$.

Exercise 42

Let $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ and $g\colon \mathbb{R}^3\to\mathbb{R}^2$ be given by

$$\begin{aligned} f(x_1,x_2) &= (\mathrm{e}^{2x_1+x_2}, 3x_2 - \cos x_1, x_1^2 + x_2 + 2), \\ g(y_1,y_2,y_3) &= (3y_1 + 2y_2 + y_3^2, y_1^2 - y_3 + 1). \end{aligned}$$

Compute $\mathrm{D}(g\circ f)(0,0)$ and $\mathrm{D}(f\circ g)(0,0,0)$.

Hint: It is not a good idea to evaluate $g\circ f$ and $f\circ g$ explicitly.

Exercise 43

Let $U\subseteq\mathbb{R}^n$ be open and $f\colon U\subseteq\mathbb{R}^n\to\mathbb{R}^k$ be differentiable, and consider the functions $\varphi\colon U\to\mathbb{R}^n\times\mathbb{R}^k$ and $F\colon U\times\mathbb{R}^k\to\mathbb{R}^k$ given by

$$\varphi(x) = (x, f(x)) \quad\text{and}\quad F(x,y) = y - f(x),$$

respectively. Show that both φ and F are differentiable, and that the relation $\ker \mathrm{D}F(x_0,y_0) = \operatorname{Im}\mathrm{D}\varphi(x_0)$ holds for all $x_0\in U$ and $y_0\in\mathbb{R}^k$.

For the next exercise, we use the following definition: differentiable analogues of homeomorphisms.

Definition 19 (Diffeomorphism)

Let $U, V \subseteq \mathbb{R}^n$ be open subsets, and $f\colon U \to V$ be a function. We say that f is a **diffeomorphism** if f is a differentiable bijection with differentiable inverse.

Exercise 44

Let $f, g, h\colon \mathbb{R}^n \to \mathbb{R}^n$ be differentiable functions, with h a diffeomorphism, and such that $f = h^{-1} \circ g \circ h$. Show that $h(p_0)$ is a fixed point of g whenever p_0 is a fixed point of f, and that $\mathrm{D}f(p_0), \mathrm{D}g(h(p_0))\colon \mathbb{R}^n \to \mathbb{R}^n$ have the same eigenvalues.

Exercise 45

Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a differentiable function, and consider its norm-squared function, $F\colon \mathbb{R}^n \to \mathbb{R}$ given by $F(x) = \|f(x)\|^2$. Note that F is also differentiable.

(a) Compute $\mathrm{D}F(x)h$ explicitly.

(b) Show that if $\|f(x)\| = 1$ for all $x \in \mathbb{R}^n$, then $\mathrm{D}f(x)\colon \mathbb{R}^n \to \mathbb{R}^n$ is not invertible. **Hint:** argue that $\mathrm{D}f(x)$ cannot be surjective.

2.2 Regularity and Hadamard's lemma

Let $U \subseteq \mathbb{R}^n$ be an open set, and $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be a function. We already know what it means for f to be differentiable on U and, if this is the case, we may now consider $\mathrm{D}f\colon U \to \mathrm{Lin}(\mathbb{R}^n, \mathbb{R}^k)$, where $\mathrm{Lin}(\mathbb{R}^n, \mathbb{R}^k) \cong \mathbb{R}^{k\times n}$ is the vector space of all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^k$. Namely, $\mathrm{D}f$ assigns to each point $p \in U$ the linear transformation $\mathrm{D}f(p)$. As $\mathrm{D}f$ itself is now a function from an open subset of a Euclidean space to another Euclidean space, it makes sense for us to ask ourselves whether $\mathrm{D}f$ is continuous or differentiable.

Definition 20 (C^r-regularity)

Let $U \subseteq \mathbb{R}^n$ be an open set, and $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be a differentiable function. We say that:

- f is of **class** C^1 if $\mathrm{D}f\colon U \to \mathrm{Lin}(\mathbb{R}^n, \mathbb{R}^k)$ is continuous.
- f is of **class** C^r, for $r \geq 2$, if $\mathrm{D}f\colon U \to \mathrm{Lin}(\mathbb{R}^n, \mathbb{R}^k)$ is of class C^{r-1}.
- f is of **class** C^∞ if it is of class C^r for every $r \geq 1$.

Functions of class C^∞ are usually just called **smooth**.

At this point, it should be rather clear that sums, products, and compositions of functions of class C^r are again of class C^r.

Later, we will restrict our attention to objects which are of class C^∞, leaving the cases of regularity C^r for finite r aside. The cheap reason is that we don't want to bother with keep tracking of how many derivatives each proof to come requires. The more honest reason is that, to some extent that will be made precise later, there is "no loss of generality" in assuming that things are C^∞ when developing the theory of smooth manifolds.

Exercise 46 ($C^\infty \subsetneq C^\omega$)

Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

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

(a) Show by induction that for $x > 0$ and $k \geq 0$, the k-th derivative $f^{(k)}(x)$ is of the form $p_{2k}(1/x)e^{-1/x}$ for some polynomial $p_{2k}(y)$ of degree $2k$ in y.

(b) Prove that f is of class C^∞ on $\mathbb{R}$ and that $f^{(k)}(0) = 0$ for all $k \geq 0$.

This means that the Taylor series of f at $x = 0$, which is the zero series (and hence has infinite radius of convergence), does not converge to f on any neighborhood of $x = 0$. In other words, real-analiticity is strictly stronger than smoothness.

The rest of this section regards a technical result we will need when discussing tangent vectors to smooth manifolds later.

Definition 21 (Starshaped sets)

A subset $U \subseteq \mathbb{R}^n$ is said to be **starshaped** around a point $p \in U$ if whenever $x \in U$ and $t \in [0,1]$ are given, we have that

$$p + t(x - p) \in U.$$

In other words, the set U is starshaped if all line segments with endpoints in U are entirely contained in U.

The point p is sometimes referred to as "the center of the star"; see Figure 28 for context.

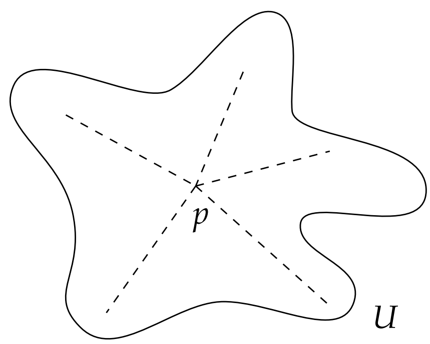


Figure 28: A starshaped subset of $\mathbb{R}^n$.

Example 49

Open and closed balls in $\mathbb{R}^n$ are starshaped around their centers. More generally, convex sets are starshaped around *all* of its points. The argument in Example 41 (p. 46) also shows that any starshaped set is connected.

Theorem 5 (Hadamard's Lemma)

Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a smooth function, where U is open and starshaped around a point $p \in U$. Then, there are smooth functions $g_1, \ldots, g_n\colon U \to \mathbb{R}$ such that

$$f(x) = f(p) + \sum_{i=1}^n (x_i - p_i) g_i(x), \tag{2.8}$$

where $x = (x_1, \ldots, x_n)$ and $p = (p_1, \ldots, p_n)$. In particular, $g_i(p) = (\partial f/\partial x_i)(p)$ for each $i = 1, \ldots, n$.

Remark. This is essentially a first-order Taylor expansion[4], but the would-be error term gets absorbed into the functions g_i. Their value at p equals the partial derivatives of f because, there, the error is zero.

Proof: As U is starshaped, we have that $p + t(x-p) \in U$ for every $t \in [0,1]$, and so we may evaluate f along this line segment. Now, note that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f(p + t(x-p)) &= \mathrm{D}f(p + t(x-p))(x-p) \\ &= \langle \nabla f(p + t(x-p)), x - p\rangle \\ &= \sum_{i=1}^n (x_i - p_i) \frac{\partial f}{\partial x_i}(p + t(x-p)). \end{aligned}$$

Then, integrate everything to obtain

$$f(x) - f(p) = \int_0^1 \sum_{i=1}^n (x_i - p_i) \frac{\partial f}{\partial x_i}(p + t(x-p))\,\mathrm{d}t. \tag{2.9}$$

If we set

$$g_i(x) = \int_0^1 \frac{\partial f}{\partial x_i}(p + t(x-p))\,\mathrm{d}t,$$

then (2.9) becomes (2.8). It is clear that $g_i(p) = (\partial f/\partial x_i)(p)$. □

Exercise 47

Establish a second-order version of Hadamard's Lemma: if $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ is smooth, with $U \subseteq \mathbb{R}^n$ open and starshaped, and $p \in U$ is given, show that there are smooth functions $g_{ij}\colon U \to \mathbb{R}$ such that

$$f(x) = f(p) + \sum_{i=1}^n (x_i - p_i)\frac{\partial f}{\partial x_i}(p) + \sum_{i,j=1}^n (x_i - p_i)(x_j - p_j) g_{ij}(x),$$

and $g_{ij}(p) + g_{ji}(p) = (\partial^2 f/\partial x_i \partial x_j)(p)$.

[4]Think of $f(x) = f(p) + \sum_{i=1}^n (x_i - p_i)\dfrac{\partial f}{\partial x_i}(p) + R(x)$, where $\displaystyle\lim_{x\to p} \frac{R(x)}{\|x-p\|} = 0$.

2.3 The Inverse and Implicit Function Theorems

The next definition—later generalized to the setting of smooth manifolds—plays a central role in the theory.

Definition 22 (Immersions and submersions between Euclidean spaces)

Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be differentiable at a point $p \in U$, where $U \subseteq \mathbb{R}^n$ is open. We say that:

(i) f is an **immersion at** p if $\mathrm{D}f(p)\colon \mathbb{R}^n \to \mathbb{R}^k$ is injective.

(ii) f is a **submersion at** p if $\mathrm{D}f(p)\colon \mathbb{R}^n \to \mathbb{R}^k$ is surjective.

(iii) the **rank of** f **at** p is the dimension of the image of $\mathrm{D}f(p)$.

If we say that f is an immersion or submersion without specifying the point p, we mean that f is so at *all* points $p \in U$.

With the above notation, note that f being an immersion implies that $n \leq k$, while if it is a submersion it must be the case that $n \geq k$. Here are some examples:

Example 50 (Regular curves)

A differentiable curve $\alpha\colon I \to \mathbb{R}^n$, where $I \subseteq \mathbb{R}$ is an open interval, is an immersion at $t_0 \in I$ if and only if $\alpha'(t_0) \neq 0$. This is because the only way the derivative $\mathrm{D}\alpha(t_0)\colon \mathbb{R} \to \mathbb{R}^n$ may fail to be injective is if $\mathrm{D}\alpha(t_0) = 0$, and this is equivalent to vanishing of the velocity vector $\alpha'(t_0)$, cf. Example 43. If $\alpha'(t) \neq 0$ for all $t \in I$, we say that α is a **regular** curve.

Example 51 (Real-valued submersions)

A differentiable function $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$, where $U \subseteq \mathbb{R}^n$ is open, is a submersion at a point $p \in U$ if and only if $\mathrm{D}f(p) \neq 0$ (or, equivalently, $\nabla f(p) \neq 0$, cf. Example 44). This is because $\mathrm{D}f(p)\colon \mathbb{R}^n \to \mathbb{R}$, being a linear functional, is either zero or surjective.

Example 52 (Regular parametrized surfaces)

A differentiable function $f\colon U \subseteq \mathbb{R}^2 \to \mathbb{R}^n$, where $U \subseteq \mathbb{R}^2$ is open, is an immersion at $(u_0, v_0) \in U$ if and only if and only if the partial derivatives $\partial f/\partial u$ and $\partial f/\partial v$ are linearly independent at (u_0, v_0). This condition essentially means that the image $f(U) \subseteq \mathbb{R}^2$ "looks like a surface" near the point $f(u_0, v_0)$, with a well-defined tangent plane at that point (spanned by the partial derivatives of f at (u_0, v_0)).

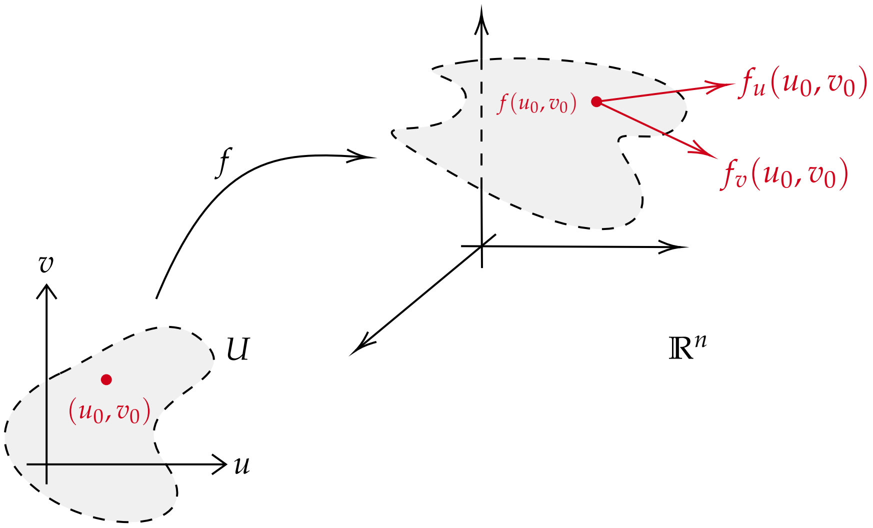


Figure 29: Linear independence of f_u and f_v at (u_0, v_0).

If this holds for every point $(u_0, v_0) \in U$, we say that f is a **regular parametrized surface**.

Example 53 (Inclusions and projections)

Let $n, m \geq 1$ be integers.

(i) The inclusion $f\colon \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m$ defined by $f(x) = (x, 0)$ is an immersion, as $Df(x)\colon \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m$ given by $Df(x)v = (v, 0)$ is injective.

(ii) The projection $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ defined by $f(x, y) = x$ is a submersion, as $Df(x, y)\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ given by $Df(x, y)(v, w) = v$ is surjective.

In both cases, the rank of f is maximal (and equal to n).

These last examples are *locally universal*: "up to a change of coordinates", every immersion and every submersion "locally look like" the above. To make precise sense of this, we need the **Inverse Function Theorem** and the **Implicit Function Theorem**.

Theorem 6 (The Inverse Function Theorem)

Let $U \subseteq \mathbb{R}^n$ be an open subset, $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a function of class C^k, with $k \geq 1$, and $p \in U$ be such that $Df(p)\colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Then, there are open subsets $V, W \subseteq \mathbb{R}^n$, with $p \in V$ and $f(p) \in W$, such that $f|_V\colon V \to W$ is invertible, with inverse $(f|_V)^{-1}\colon W \to V$ of the same class C^k. Finally, the relation $D(f^{-1})(f(p)) = Df(p)^{-1}$ holds (in fact, for any point in V).

For a proof, see [24, Theorem 2.11]. The core idea here is that good properties that $Df(p)$ has (e.g., invertibility) translate into good properties of f itself, near p. The formula for the derivative of the local inverse of f is a direct consequence of the chain rule: from $f^{-1}(f(p)) = p$ it follows that $D(f^{-1})(f(p)) \circ Df(p) = \mathrm{Id}_{\mathbb{R}^n}$.

Example 54

Consider again the function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x,y) = (x^2 - y^2, 2xy)$, from Example 47. This function is not injective (and hence not invertible), due to the relation $f(-x,-y) = f(x,y)$, valid for all $(x,y) \in \mathbb{R}^2$. However, we have that

$$\det \mathrm{D}f(x,y) = \det \begin{bmatrix} 2x & -2y \\ 2y & 2x \end{bmatrix} = 4(x^2+y^2) > 0$$

for every $(x,y) \neq (0,0)$. The Inverse Function Theorem says that f is locally invertible with smooth inverse near every point $(x,y) \neq (0,0)$. On the other hand, f is not invertible on *any* neighborhood of $(0,0)$!

Exercise 48

Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x,y) = (\mathrm{e}^x \cos y, \mathrm{e}^x \sin y)$.

(a) Show that f is injective on $U = \mathbb{R} \times (0, 2\pi)$.

(b) Determine the image $f(U)$.

(c) If $g = f^{-1}$, where defined, compute $\mathrm{D}g(0,1)$.

To state the Implicit Function Theorem, we need to understand how "fat" partial derivatives work. The idea is the same as for classical partial derivatives: we fix one of the variables and differentiate the function with respect to the other variable. The only difference is that, here, these variables are vectors, and differentiating means taking a total derivative. We have the following definition:

Definition 23 (Partial derivatives redux)

Let $f\colon U \subseteq \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ be a differentiable function, where $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ is an open subset. Given $(x_0, y_0) \in U$, the partial derivative $\mathrm{D}_x f(x_0,y_0)\colon \mathbb{R}^n \to \mathbb{R}^m$ is defined to be the total derivative at x_0 of the function $x \mapsto f(x, y_0)$. Similarly, the partial derivative $\mathrm{D}_y f(x_0,y_0)\colon \mathbb{R}^k \to \mathbb{R}^m$ is the total derivative at y_0 of the function $y \mapsto f(x_0, y)$.

By design, we have that $\mathrm{D}f(x_0,y_0)(v,w) = \mathrm{D}_x f(x_0,y_0)v + \mathrm{D}_y f(x_0,y_0)w$ for all $(v,w) \in \mathbb{R}^n \times \mathbb{R}^k$.

At this point, examples involving only two or three variables in the domain are "too small" to be instructive. So, we need to consider something slightly more complicated to properly illustrate Definition 23.

Example 55

Let $f\colon \mathbb{R}^5 \to \mathbb{R}^3$ be given by

$$f(x,y,z,r,s) = \big(x^2y + z\sin(r+s), xy + zr + s^2, xe^z + ye^{rs}\big).$$

At any point $(x,y,z,r,s) \in \mathbb{R}^5$ we have that

$$\mathrm{D}f(x,y,z,r,s) = \begin{bmatrix} 2xy & x^2 & \sin(r+s) & z\cos(r+s) & z\cos(r+s) \\ y & x & r & z & 2s \\ e^z & e^{rs} & xe^z & yse^{rs} & yre^{rs} \end{bmatrix},$$

and it breaks into two partial-derivative blocks

$$\mathrm{D}_{(x,y,z)}f(x,y,z,r,s) = \begin{bmatrix} 2xy & x^2 & \sin(r+s) \\ y & x & r \\ e^z & e^{rs} & xe^z \end{bmatrix}$$

and

$$\mathrm{D}_{(r,s)}f(x,y,z,r,s) = \begin{bmatrix} z\cos(r+s) & z\cos(r+s) \\ z & 2s \\ yse^{rs} & yre^{rs} \end{bmatrix}.$$

Of course, there are other possible block-partitions for $\mathrm{D}f(x,y,z,r,s)$, such as the one involving $\mathrm{D}_{(x,y)}f(x,y,z,r,s)$ and $\mathrm{D}_{(z,r,s)}f(x,y,z,r,s)$, etc.

Understanding this new mechanism with partial derivatives in higher generality can be instructive: let $f\colon \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_r} \to \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_s}$ be a differentiable function, and write it in components as $f = (f_1, \ldots, f_s)$, where for each $j = 1, \ldots, s$ we have that $f_j\colon \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_r} \to \mathbb{R}^{m_j}$ is differentiable. Writing $x = (x_1, \ldots, x_r)$, where $x_i \in \mathbb{R}^{n_i}$ for each $i = 1, \ldots, r$, we then have that

$$\mathrm{D}f(x) = \begin{bmatrix} \mathrm{D}_{x_1}f_1(x) & \cdots & \mathrm{D}_{x_r}f_1(x) \\ \vdots & \ddots & \vdots \\ \mathrm{D}_{x_1}f_s(x) & \cdots & \mathrm{D}_{x_r}f_s(x) \end{bmatrix} \tag{2.10}$$

in block form, where $\mathrm{D}_{x_i}f_j(x) \in \mathbb{R}^{m_j \times n_i}$ for all $i = 1, \ldots, r$ and $j = 1, \ldots, s$. When $n_1 = \cdots = n_r = m_1 = \cdots = m_s = 1$, each block becomes a classical partial derivative $\partial f_j / \partial x_i$, and (2.10) reduces to (2.6). You can think of this as the first algorithm described to compute Jacobian matrices, with the difference that each entry of the "gradient" of f_j is now a block matrix itself.

In any case, we proceed:

Theorem 7 (The Implicit Function Theorem)

Let $f\colon U \subseteq \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k$ be a function of class C^r on an open subset $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$, with $r \geq 1$, and $(x_0, y_0) \in U$ be such that $\mathrm{D}_y f(x_0, y_0)\colon \mathbb{R}^k \to \mathbb{R}^k$ is invertible. Then, if $c = f(x_0, y_0)$, there are open subsets $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^k$ such that $x_0 \in V$ and $y_0 \in W$, with $V \times W \subseteq U$, and a function $\varphi\colon V \to W$ of the same class C^r such that

$$\text{whenever } (x, y) \in V \times W, \text{ we have } f(x, y) = c \iff y = \varphi(x). \tag{2.11}$$

Finally, for any $x \in V$ we have that $\mathrm{D}\varphi(x) = -\mathrm{D}_y f(x, \varphi(x))^{-1} \circ \mathrm{D}_x f(x, \varphi(x))$.

Remark. The real content of the theorem is that $f^{-1}(c) \cap (V \times W) = \mathrm{Gr}(\varphi)$, that is, the inverse image $f^{-1}(c)$ locally looks like the graph of a function of the same regularity as f, near the starting point (x_0, y_0), cf. Figure 30.

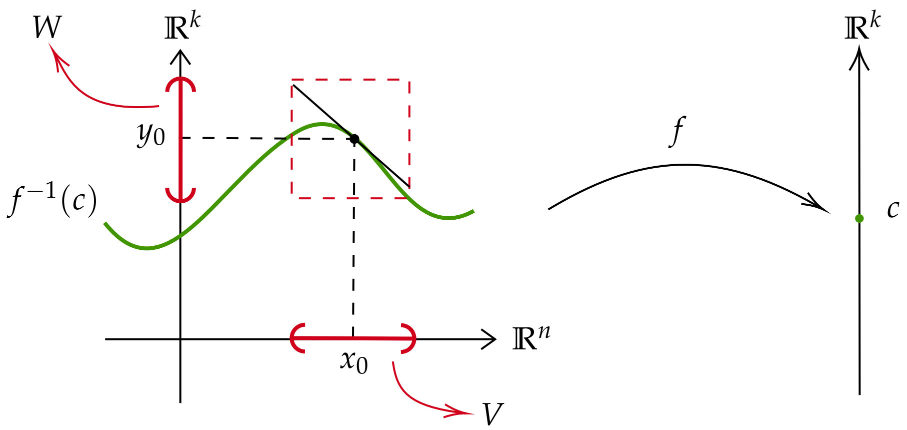


Figure 30: The relation $f^{-1}(c) \cap (V \times W) = \mathrm{Gr}(\varphi)$: the part of $f^{-1}(c)$ inside the box $V \times W$ is the graph of a C^r function $\varphi\colon V \to W$. Invertibility of $\mathrm{D}_y f(x_0, y_0)\colon \mathbb{R}^k \to \mathbb{R}^k$ means that the tangent line-segment to $f^{-1}(c)$, indicated in black, is not horizontal.

Proof: We start by considering the auxilliary function $F\colon U \to \mathbb{R}^n \times \mathbb{R}^k$ given by $F(x, y) = (x, f(x, y))$, which is of class C^r since f is, and compute its total derivative as

$$\mathrm{D}F(x_0, y_0) = \begin{bmatrix} \mathrm{Id}_n & 0 \\ \mathrm{D}_x f(x_0, y_0) & \mathrm{D}_y f(x_0, y_0) \end{bmatrix}.$$

It is invertible since $\mathrm{D}_y f(x_0, y_0)$ is. Thus we may apply the Inverse Function Theorem to F, obtaining a local inverse $F^{-1}\colon Z \subseteq \mathbb{R}^n \times \mathbb{R}^k \to V_0 \times W \subseteq \mathbb{R}^n \times \mathbb{R}^k$ of class C^r; here, $Z \subseteq \mathbb{R}^n \times \mathbb{R}^k$, as well as $V_0 \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^k$, are all open. Such inverse, however, is necessarily of the form $F^{-1}(u, v) = (u, h(u, v))$ for some function $h\colon Z \to W$ of class C^r. With this in place, we define $\varphi\colon V \to W$ by $\varphi(x) = h(x, c)$, where $V = \{x \in \mathbb{R}^n : x \in V_0 \text{ and } (x, c) \in Z\}$ (it is open in $\mathbb{R}^n$). To establish (2.11), note that for every $x \in V$ we have that

$$(x, c) = FF^{-1}(x, c) = F(x, h(x, c)) = F(x, \varphi(x)) = (x, f(x, \varphi(x))),$$

leading to $c = \varphi(x)$; conversely, if $(x,y) \in V \times W$ has $f(x,y) = c$, then $(x,c) = F(x,y)$ means that $F(x,y) = F(x,\varphi(x))$, and injectivity of F in this region yields $y = \varphi(x)$.

It remains to establish the formula for the derivative of φ. However, it suffices to differentiate the relation $f(x,\varphi(x)) = c$ using the chain rule, to obtain the relation $\mathrm{D}_x f(x,\varphi(x)) + \mathrm{D}_y f(x,\varphi(x)) \circ \mathrm{D}\varphi(x) = 0$. Invertibility of $\mathrm{D}_y f(x,\varphi(x))$ allows us to solve for $\mathrm{D}\varphi(x)$, as required. □

Let's see one concrete example:

Example 56

Let $f\colon \mathbb{R}^4 \to \mathbb{R}^2$ be the (clearly smooth) function given by the formula

$$f(x,y,z,w) = \big(\mathrm{e}^{z+w} + x(z^2+w^2)\,,\ \sin(z+2w) + x + y\big).$$

Question: *Can we describe the set in $\mathbb{R}^4$ given by the equation $f(x,y,z,w) = (1,0)$ as the graph of a function $(z,w) = \varphi(x,y)$, near the origin $(x,y,z,w) = (0,0,0,0)$? Or equivalently, can we solve the system*

$$\begin{cases} \mathrm{e}^{z+w} + x(z^2+w^2) = 1 \\ \sin(z+2w) + x + y = 0 \end{cases}$$

for z and w as functions of x and y near the origin?

We compute the partial derivative of f relative to (z,w) as

$$\mathrm{D}_{(z,w)} f(x,y,z,w) = \begin{bmatrix} \mathrm{e}^{z+w} + 2xz & \mathrm{e}^{z+w} + 2xw \\ \cos(z+2w) & 2\cos(z+2w) \end{bmatrix},$$

and evaluate it at the origin:

$$\mathrm{D}_{(z,w)} f(0,0,0,0) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Answer: *As the latter matrix is invertible, yes!*

Since $f(0,0,0,0) = (1,0)$, there are open subsets $V, W \subseteq \mathbb{R}^2$, with $(1,0) \in V$ and $(0,0) \in W$, and a smooth function $\varphi\colon V \to W$ such that whenever $(x,y) \in V$ and $(z,w) \in W$, we have that

$$f(x,y,z,w) = (1,0) \iff (z,w) = \varphi(x,y).$$

The common abuse of notation, in a situation like this, would be to write the function φ as $\varphi(x,y) = (z(x,y), w(x,y))$, so that $\mathrm{D}\varphi$ contains all four partial derivatives $\partial z/\partial x$, $\partial z/\partial y$, $\partial w/\partial x$, and $\partial w/\partial y$. In any case, we may also find $\mathrm{D}\varphi(0,0)$

with the chain rule: from $f(x,y,\varphi(x,y)) = (1,0)$ it follows that

$$\mathrm{D}_{(x,y)}f(0,0,0,0) + \mathrm{D}_{(z,w)}f(0,0,0,0) \circ \mathrm{D}\varphi(0,0) = 0,$$

which reads as

$$\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \mathrm{D}\varphi(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

and leads to

$$\mathrm{D}\varphi(0,0) = -\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

(For example, the value of $\partial z/\partial x$ at $(x,y) = (0,0)$ is 1, while $\partial w/\partial x$ at the same point is -1.)

Exercise 49

Let $F\colon \mathbb{R}^2 \to \mathbb{R}$ be a function of class C^2, with $F(0,0) = 0$ and total derivative given by $\mathrm{D}F(0,0) = [2 \quad 3]$.

(a) Show that the "surface" $F(x+2y+3z-1, x^3+y^2-z^2) = 0$ can be expressed, near the point $(-2,3,1)$, as the graph of a C^2 function $z = z(x,y)$.

(b) Compute $(\partial z/\partial y)(-2,3)$.

As it turns out, the Implicit and Inverse Function Theorems are logically equivalent to each other. The next exercise gives you a guide to derive the Inverse Function Theorem from the Implicit Function Theorem:

Exercise 50 (From Implicit to Inverse)

Here, we assume that the Implicit Function Theorem is true. Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a function of class C^r, with $r \geq 1$, and let $x_0 \in U$ be such that $\mathrm{D}f(x_0)\colon \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism.

(a) Consider the function $F\colon U \times \mathbb{R}^n \subseteq \mathbb{R}^{2n} \to \mathbb{R}^n$ given by $F(x,y) = f(x) - y$. It has $F(x_0, f(x_0)) = 0$. Show that $\mathrm{D}_x F(x_0, f(x_0))\colon \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism.

(b) Explain how this implies the existence of neighborhoods V and W of x_0 and $f(x_0)$, respectively, together with a function $\varphi\colon W \to V$ having the property that if $(x,y) \in V \times W$, then $F(x,y) = 0$ if and only if $x = \varphi(y)$.

(c) Reducing V and W if needed, we may assume that $V \subseteq f^{-1}(W)$. Then conclude that $f|_V$ is a bijection onto its image, with $(f|_V)^{-1} = \varphi$, by showing that both $\varphi(f(x)) = x$ and $f(\varphi(y)) = y$ for all $(x,y) \in V \times W$.

We are now ready to establish the local classifications of immersions and submersions:

Proposition 26 (Local form of immersions)

Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^{n+k}$ be a function of class C^r, with $r \geq 1$, where $U \subseteq \mathbb{R}^n$ is open, and assume that f is an immersion at $x_0 \in U$. Then, there are open subsets $V \times W$ and Z of $\mathbb{R}^{n+k} \cong \mathbb{R}^n \times \mathbb{R}^k$ containing $(x_0, 0)$ and $f(x_0)$, respectively, and a C^r diffeomorphism $\psi\colon Z \to V \times W$ such that $(\psi \circ f)(x) = (x, 0)$ for every $x \in V$.

Proof: By applying a permutation of the axes in $\mathbb{R}^{n+k}$ if needed, we may assume that $f(x) = (u(x), v(x))$, with $u(x) \in \mathbb{R}^n$, $v(x) \in \mathbb{R}^k$, and $\mathrm{D}u(x_0)\colon \mathbb{R}^n \to \mathbb{R}^n$ invertible. Now, we consider the auxiliary function (of the same class C^r) $F\colon U \times \mathbb{R}^k \to \mathbb{R}^{n+k}$ given by $F(x,y) = (u(x), v(x) + y)$, and compute its total derivative as

$$\mathrm{D}F(x_0, 0) = \begin{bmatrix} \mathrm{D}u(x_0) & 0 \\ \mathrm{D}v(x_0) & \mathrm{Id}_k \end{bmatrix}.$$

It is invertible since $\mathrm{D}u(x_0)$ is. The Inverse Function Theorem now yields the required neighborhoods $V \times W$ and Z in $\mathbb{R}^{n+k} \cong \mathbb{R}^n \times \mathbb{R}^k$ of $(x_0, 0)$ and $f(x_0)$, on which $F|_{V \times W}\colon V \times W \to Z$ is invertible with C^r inverse. Setting $\psi = (F|_{V\times W})^{-1}$, we simply compute $(\psi \circ f)(x) = (\psi \circ F)(x, 0) = (x, 0)$, for every $x \in V$. ☐

Corollary 5 (Immersions are locally injective)

Every immersion of class C^r, with $r \geq 1$, is locally injective.

Proof: Inclusions are injective. ☐

Proposition 27 (Local form of submersions)

Let $f\colon U \subseteq \mathbb{R}^{n+k} \to \mathbb{R}^n$ be a function of class C^r, with $r \geq 1$, where $U \subseteq \mathbb{R}^{n+k}$ is open, and assume that f is a submersion at $(x_0, y_0) \in U$. Then, there are open subsets $V \times W$ and Z in $\mathbb{R}^{n+k}$ containing $(f(x_0, y_0), y_0)$ and (x_0, y_0), respectively, and a C^r diffeomorphism $\psi\colon V \times W \to Z$ such that $(f \circ \psi)(x, y) = x$ for every $(x, y) \in V \times W$.

Proof: By applying a permutation of the axes in $\mathbb{R}^{n+k}$ if needed, we may assume that $\mathrm{D}_x f(x_0, y_0)\colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Now, we consider the auxilliary function (of the same class C^r) $F\colon U \to \mathbb{R}^{n+k}$ given by $F(x, y) = (f(x, y), y)$, and compute its total derivative as

$$\mathrm{D}F(x_0, y_0) = \begin{bmatrix} \mathrm{D}_x f(x_0, y_0) & \mathrm{D}_y f(x_0, y_0) \\ 0 & \mathrm{Id}_k \end{bmatrix}.$$

It is invertible since $\mathrm{D}_x f(x_0, y_0)$ is. By the Inverse Function Theorem, there are neighborhoods Z and $V \times W$ in $\mathbb{R}^{n+k}$ of (x_0, y_0) and $(f(x_0, y_0), y_0)$, respectively, on which

$F|_Z : Z \to V \times W$ is invertible with C^r inverse. This inverse is necessarily of the form $F^{-1}(u,v) = (h(u,v), v)$ for some function h of class C^r, which has $f(h(u,v), v) = u$ for all $(u,v) \in V \times W$ (as a consequence of $(F \circ F^{-1})(u,v) = (u,v)$). It follows that if we set $\psi = (F|_Z)^{-1}$, we have that $(f \circ \psi)(u,v) = u$ for all $(u,v) \in V \times W$, as required (we now rename (u,v) back to (x,y)). □

Corollary 6 (Submersions are open)

Every submersion of class C^r, with $r \geq 1$, is an open mapping.

Proof: Euclidean projections are open mappings (cf. Example 17, p. 22). □

Exercise 51

Give proofs of Propositions 26 and 27 directly using the Implicit Function Theorem instead of the Inverse Function Theorem.

We will revisit and expand on the ideas discussed in this section once we have the language of smooth manifolds available to us.

3 From topological manifolds to smooth manifolds

3.1 Locally Euclidean spaces and examples

Armed with the tools from multivariable calculus seen in Section 2, we return to topology and move towards the definition of a smooth manifold.

Definition 24 (Locally Euclidean spaces)

Let X be a topological space, and $n \geq 0$ be an integer.

- An **n-dimensional chart** for X is a pair (U, φ), where $U \subseteq X$ is open and $\varphi\colon U \to \varphi(U) \subseteq \mathbb{R}^n$ is a homeomorphism, with the image $\varphi(U)$ being an open subset of $\mathbb{R}^n$.
- We say that X is **locally Euclidean of dimension** n (or, **locally** $\mathbb{R}^n$) if for every $p \in X$ there is an n-dimensional chart (U, φ) with $p \in U$. In this case we say that (U, φ) is a chart for X **around** p and, if $\varphi(p) = 0$, that (U, φ) is **centered at** p. The **dimension** of X is defined to be $\dim X = n$.

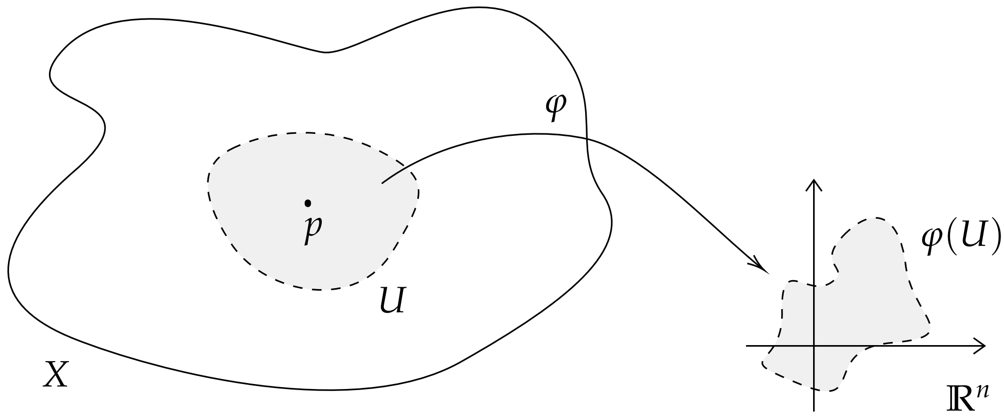


Figure 31: The definition of a locally Euclidean space.

Remark. If an open subset of $\mathbb{R}^n$ is homeomorphic to an open subset of $\mathbb{R}^m$, then we necessarily have that $n = m$. This is a very advanced theorem, called **Invariance of Domain**. It in particular implies that if (U, φ) is an n-dimensional chart and (V, ψ) is an m-dimensional chart, and $U \cap V \neq \varnothing$, then $n = m$: just consider the homeomorphism $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \subseteq \mathbb{R}^n \to \psi(U \cap V) \subseteq \mathbb{R}^m$. It now follows that for each $n \geq 0$, the set

$$\{p \in X : \text{there is an } n\text{-dimensional chart } (U, \varphi) \text{ around } p\}$$

is both open and closed in X. This means that if X is a topological space with the property that around every point there is a chart, the resulting "dimension" of X is in fact well-defined in each connected component of X. Some books such as [15] actually allow for that in the definition of a locally Euclidean space, and consequently allow for manifolds having connected components of different dimensions.

We may now put together some of the main topological properties which $\mathbb{R}^n$ has:

Definition 25 (Topological manifold)

A **topological manifold** is a topological space M which is Hausdorff, second-countable, and locally Euclidean. A collection $\mathfrak{A} = \{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ of charts with $\bigcup_{\alpha \in A} U_\alpha = M$ is called an **atlas** for M.

Example 57 (Vector spaces are topological manifolds)

Euclidean space $\mathbb{R}^n$ admits a global chart $(\mathbb{R}^n, \mathrm{Id}_{\mathbb{R}^n})$, and so $\mathbb{R}^n$ is locally Euclidean (as it should be). Hence $\mathbb{R}^n$ is a topological manifold. More generally, if V is any abstract finite-dimensional real vector space and $T\colon V \to \mathbb{R}^n$ is an isomorphism, then (V, T) is a global chart for V. Recall that T is a homeomorphism by the definition of the Euclidean topology on V, cf. Example 5 (p. 7).

Example 58 (Open subsets of topological manifolds)

Open subsets W of topological manifolds M are topological manifolds on their own right, when equipped the subspace topology. The Hausdorff and second-countability conditions are automatically inherited. Finally, whenever (U, φ) is a chart for M with $U \cap W \neq \varnothing$, we have that $(U \cap W, \varphi|_{U \cap W})$ is a chart for W. In particular, $\dim W = \dim M$.

Example 59 (A non-example: the cross)

Let $X = (\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R})$ be the union of the coordinate axes in $\mathbb{R}^2$, equipped with its subspace topology. It is clearly Hausdorff and second-countable, but we claim that it is *not* locally Euclidean—namely, there is no chart for X around the point $(0,0) \in X$. Indeed, since X is connected, we know that if X were locally Euclidean, it would be a topological manifold of dimension 1.

However, we claim that there is no open neighborhood of $(0,0)$ in X which is homeomorphic to an open interval in $\mathbb{R}$. Indeed, if U were such a neighborhood and $\varphi\colon U \subseteq X \to I \subseteq \mathbb{R}$ were a chart, φ would restrict to a homeomorphism between $U \setminus \{(0,0)\}$ and $I \setminus \{\varphi(0,0)\}$, and this is impossible: the former space has four connected components, while the latter has only two.

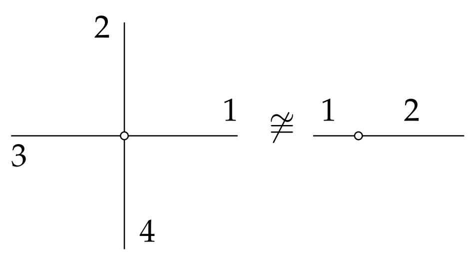


Figure 32: The cross is not locally Euclidean at the origin.

The condition that a topological space M is locally Euclidean alone implies that M is **locally connected** (every point has a connected open neighborhood), **locally**

compact (every point has an open neighborhood with compact closure), **locally path-connected** (every point has a path-connected open neighborhood), and first-countable. It does not imply, however, that M is Hausdorff or second-countable—which is the reason these conditions had to be separately included in Definition 25.

Example 60 (The line with two origins redux)

Consider the line with two origins, from Exercise 12 (p. 15). In items (b) and (c), you were asked to prove that it is not Hausdorff (the two origins z_1 and z_2 cannot be separated by disjoint open subsets), and that it is second-countable. Here, we also claim that it is locally Euclidean. If $a \in (0,\infty)$ (or, $a \in (-\infty,0)$) we may simply take the identity chart on $(0,\infty)$ (or, on $(-\infty,0)$), while given $i \in \{1,2\}$ we may set $U_i = (\mathbb{R} \smallsetminus \{0\}) \cup \{z_i\}$ and define $\varphi_i \colon U_i \to \mathbb{R}$ by

$$\varphi_i(x) = \begin{cases} x, & \text{if } x \neq z_i, \\ 0, & \text{if } x = z_i. \end{cases}$$

Then each (U_i, φ_i) is a chart around z_i. In other words, an atlas for the line with two origins is $\mathfrak{A} = \{((-\infty,0), \mathrm{Id}_{(-\infty,0)}), ((0,\infty), \mathrm{Id}_{(0,\infty)}), (U_1,\varphi_1), (U_2,\varphi_2)\}$.

Example 61 (The vertical topology redux)

Consider our vertical topology $\tau_{\text{vert.}}$ on $\mathbb{R}^2$, from Example 12 (p. 14).

It is Hausdorff, since if $(x_1,y_1),(x_2,y_2) \in \mathbb{R}^2$ are distinct, then either $x_1 \neq x_2$ allows us to take $U_1 = \{x_1\} \times \mathbb{R}$ and $U_2 = \{x_2\} \times \mathbb{R}$, while if $x_1 = x_2$ and $y_1 \neq y_2$, we may take $U_1 = \{x_1\} \times (y_1 - \varepsilon, y_1 + \varepsilon)$ and $U_2 = \{x_1\} \times (y_2 - \varepsilon, y_2 + \varepsilon)$ with $\varepsilon = |y_1 - y_2|/2 > 0$; in either case, $U_1, U_2 \in \tau_{\text{vert.}}$ are disjoint with $(x_1,y_1) \in U_1$ and $(x_2,y_2) \in U_2$.

Next, you should have already verified in Exercise 11 (p. 15) that $(\mathbb{R}^2, \tau_{\text{vert.}})$ is not second-countable.

However, we claim that $(\mathbb{R}^2, \tau_{\text{vert.}})$ is locally Euclidean and, in this case, its dimension as a topological manifold equals 1 (and not 2). Indeed, for any $(x_0,y_0) \in \mathbb{R}^2$, we have that $\varphi_{x_0} \colon \{x_0\} \times \mathbb{R} \to \mathbb{R}$ given by $\varphi_{x_0}(x_0,y) = y$ is a one-dimensional chart (the codomain is equipped with its Euclidean topology) around (x_0,y_0). In other words, the collection $\mathfrak{A} = \{(\{x\} \times \mathbb{R}, \varphi_x)\}_{x \in \mathbb{R}}$ is an atlas for $(\mathbb{R}^2, \tau_{\text{vert.}})$.

Examples 59, 60, and 61 show that any of the three conditions—Hausdorffness, second-countability, being locally Euclidean—in Definition 25 alone is not implied by the remaining two together.

In the exercise below, you can further explore how these conditions play with each other.

Exercise 52 (Locally Euclidean topological groups)

Let G be a **topological group**, that is, a group equipped with a topology for which the group operation $G \times G \to G$ and the inversion mapping $G \to G$ are both continuous. Assume that a chart around the identity element $e \in G$ exists.

(a) Show that the singleton $\{e\}$ is closed. (In fact, all singletons are closed in locally Euclidean spaces.)

(b) Show that G is Hausdorff.
Hint: prove that the diagonal $\Delta \subseteq G \times G$ is closed by expressing it as the inverse image of a closed set under a suitable continuous function $G \times G \to G$, and use Proposition 17 (p. 33).

(c) Show that for each $g \in G$, the mapping $\mathrm{L}_g \colon G \to G$ given by $\mathrm{L}_g(h) = gh$ is a homeomorphism. Use this to show that G is locally Euclidean.

Unfortunately, the humble assumption that G is locally Euclidean just at the identity does not seem to imply that G is a topological manifold. But we still have the conclusion that if a *second-countable* topological group is locally Euclidean at the identity, then it is a topological manifold.

Finally, it will be useful to know that arbitrary charts can be modified to have "better images", such as balls, cubes, or the entire Euclidean space. More precisely:

Exercise 53

Let X be a locally Euclidean topological space, and let $p \in X$ be fixed. In each item below, show that there is a chart (U, φ) around p (that is, with $p \in U$), satisfying the given condition:

(a) $\varphi(p) = 0$.

(b) $\varphi(p) = 0$ and $\varphi(U) = B_r(0)$, where $r > 0$ is prescribed.

(c) $\varphi(p) = 0$ and $\varphi(U) = (-r, r) \times \cdots \times (-r, r)$, where $r > 0$ is prescribed.

(d) $\varphi(p) = 0$ and $\varphi(U) = \mathbb{R}^n$.

3.2 C^k-compatibility between charts and atlases

The definition of topological enough is still broad enough to allow for things like the graph $M = \{(x, y) \in \mathbb{R}^2 : y = |x|\}$ of the absolute-value function, with its subspace topology induced from $\mathbb{R}^2$, to be topological manifolds (namely, we have a global chart $M \ni (x, y) \mapsto x \in \mathbb{R}$). But the function $x \mapsto |x|$ is not differentiable at $x = 0$, there is no "tangent line" to M at $(0, 0)$, and we cannot reasonably perform Calculus on M.

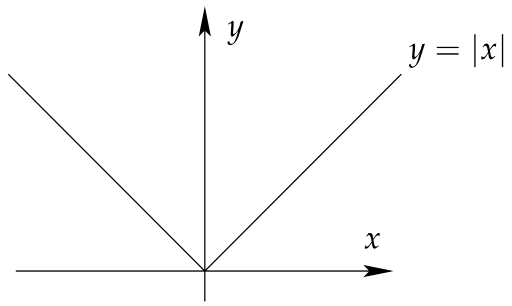


Figure 33: The graph of $y = |x|$ has a cusp at the origin.

If M is a topological manifold and $n = \dim M$, two charts (U, φ) and (V, ψ) for M with $U \cap V \neq \varnothing$ always satisfy that $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ is a homeomorphism between open subsets of $\mathbb{R}^n$; its inverse is $\varphi \circ \psi^{-1}\colon \psi(U \cap V) \to \varphi(U \cap V)$. See Figure 34.

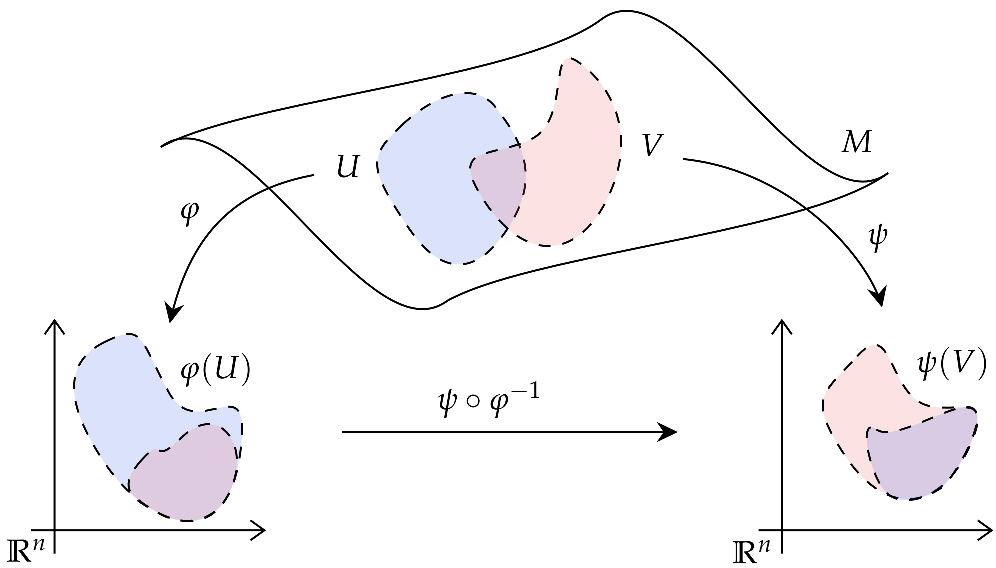


Figure 34: The chart transition $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$.

We need to demand more of $\psi \circ \varphi^{-1}$: being a homeomorphism is not strong enough of a condition to enable us to do Calculus.

Definition 26 (C^k-compatibility of charts)

Let M be a topological manifold, and $k \in \{1, 2 \ldots, \infty\}$. Two charts (U, φ) and (V, ψ) are called C^k**-compatible** if either $U \cap V = \varnothing$, or $U \cap V \neq \varnothing$ and both transitions $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ and $\varphi \circ \psi^{-1}\colon \psi(U \cap V) \to \varphi(U \cap V)$ are of class C^k. An atlas $\mathfrak{A}$ for M is called a C^k**-atlas** (or, an **atlas of class** C^k) if all of its charts are pairwise C^k-compatible.

Example 62 (A C^∞-atlas for the circle)

Consider the circle $\mathbb{S}^1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, and let

$$U_1 = \{(x,y) \in \mathbb{S}^1 : x > 0\}, \quad U_2 = \{(x,y) \in \mathbb{S}^1 : y > 0\}$$
$$U_3 = \{(x,y) \in \mathbb{S}^1 : x < 0\}, \quad U_4 = \{(x,y) \in \mathbb{S}^1 : y < 0\}.$$

They are all open in $\mathbb{S}^1$—for example, $U_1 = ((0,\infty) \times \mathbb{R}) \cap \mathbb{S}^1$. See Figure 35.

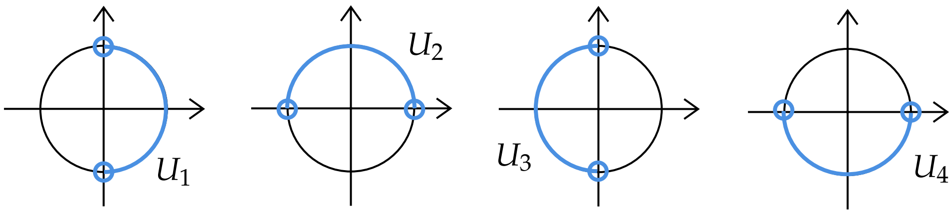


Figure 35: Defining a smooth atlas for $\mathbb{S}^1$.

We define four charts $\varphi_i \colon U_i \to (-1,1)$ by

$$\varphi_1(x,y) = \varphi_3(x,y) = y \quad \text{and} \quad \varphi_2(x,y) = \varphi_4(x,y) = x.$$

They are all continuous, being restrictions of continuous functions defined on all of $\mathbb{R}^2$. To argue that they are indeed homeomorphisms, we exhibit their inverses, thinking about how each open half of $\mathbb{S}^1$ (upper, lower, left, and right) can be expressed as graphs of continuous functions:

$$\varphi_1^{-1}(t) = \big(\sqrt{1-t^2}, t\big), \quad \varphi_2^{-1}(t) = \big(t, \sqrt{1-t^2}\big),$$
$$\varphi_3^{-1}(t) = \big(-\sqrt{1-t^2}, t\big), \quad \varphi_4^{-1}(t) = \big(t, -\sqrt{1-t^2}\big).$$

They are continuous as functions valued in each U_i because they are continuous when seen as functions valued in $\mathbb{R}^2$, cf. Proposition 2 (p. 8).

Let's now show that the charts (U_1, φ_1) and (U_2, φ_2) are C^∞-compatible. We have that $U_1 \cap U_2 = \{(x,y) \in \mathbb{S}^1 : x, y > 0\}$, so that

$$\varphi_2 \circ \varphi_1^{-1} \colon \underbrace{\varphi_1(U_1 \cap U_2)}_{=(0,1)} \to \underbrace{\varphi_2(U_1 \cap U_2)}_{=(0,1)}$$

given by $(\varphi_2 \circ \varphi_1^{-1})(t) = \sqrt{1-t^2}$ is of class C^∞ on $(0,1)$ (but not on $(0,1]$!); the inverse $\varphi_1 \circ \varphi_2^{-1} \colon (0,1) \to (0,1)$ is also of class C^∞, being given by the same formula. Hence (U_1, φ_1) and (U_2, φ_2) are C^∞-compatible.

A similar argument establishes C^∞-compatibility of all other possible pairs of charts, making $\mathfrak{A} = \{(U_i, \varphi_i)\}_{i=1}^4$ a C^∞-atlas for $\mathbb{S}^1$.

Exercise 54 (A C^∞-atlas for the sphere)

Generalize to $\mathbb{S}^2 = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1\}$, with its standard Euclidean topology, the previous example. Namely, consider the following domains and charts:

$$\begin{aligned}
U_1 &= \{(x,y,z) \in \mathbb{S}^2 : x > 0\}, & \varphi_1(x,y,z) &= (y,z)\\
U_2 &= \{(x,y,z) \in \mathbb{S}^2 : x < 0\}, & \varphi_2(x,y,z) &= (y,z)\\
U_3 &= \{(x,y,z) \in \mathbb{S}^2 : y > 0\}, & \varphi_3(x,y,z) &= (x,z)\\
U_4 &= \{(x,y,z) \in \mathbb{S}^2 : y < 0\}, & \varphi_4(x,y,z) &= (x,z)\\
U_5 &= \{(x,y,z) \in \mathbb{S}^2 : z > 0\}, & \varphi_5(x,y,z) &= (x,y)\\
U_6 &= \{(x,y,z) \in \mathbb{S}^2 : z < 0\}, & \varphi_1(x,y,z) &= (x,y)
\end{aligned}$$

Exhibit each inverse function φ_i^{-1} and its domain, describe the images $\varphi_3(U_3 \cap U_6)$ and $\varphi_2(U_2 \cap U_5)$, and show that the transitions $\varphi_6 \circ \varphi_3^{-1}$ and $\varphi_5 \circ \varphi_2^{-1}$ are of class C^∞. With a little more patience, you can show that $\mathfrak{A} = \{(U_i, \varphi_i)\}_{i=1}^6$ is a C^∞-atlas for $\mathbb{S}^2$.

How many charts would you need to cover the n-dimensional sphere $\mathbb{S}^n$ in this manner? Can you write down the resulting atlas? (We will see later in Exercise 60 that it is actually possible to cover $\mathbb{S}^n$ with only two charts.)

Example 63 (Singleton atlases are C^∞)

Any atlas for a topological manifold consisting of a single global chart (U, φ) is automatically a C^∞-atlas: there is only one transition mapping to consider, which is the identity $\varphi \circ \varphi^{-1} = \mathrm{Id}_{\varphi(U)}$.

In particular, for a finite-dimensional real vector space V, the global linear chart (V, T) corresponding to an isomorphism $T\colon V \to \mathbb{R}^n$ (cf. Example 57) is a C^∞-atlas. But we can say more: if (V, S) is another such global chart, induced by a second isomorphism $S\colon V \to \mathbb{R}^n$, then (V,T) and (V,S) are C^∞-compatible. Indeed, the transitions $T \circ S^{-1}, S \circ T^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$ are linear, and linear operators on $\mathbb{R}^n$ are always of class C^∞.

Example 64 (A C^∞-atlas for the real projective plane)

Recall that the real projective plane is defined as the quotient $\mathbb{RP}^2 = (\mathbb{R}^3 \setminus \{0\})/{\sim}$, where $p \sim q$ if $q = \lambda p$ for some $\lambda \in \mathbb{R} \setminus \{0\}$. We denote the equivalence class of (x,y,z) by $[x:y:z]$, so that $\pi(x,y,z) = [x:y:z]$ for the quotient projection $\pi\colon \mathbb{R}^3 \setminus \{0\} \to \mathbb{RP}^2$. We call $[x:y:z]$ the **homogeneous coordinates** of the line spanned by (x,y,z) since, by construction, the relation $[\lambda x : \lambda y : \lambda z] = [x:y:z]$ holds for every $\lambda \neq 0$; note however that here the word "coordinates" does not refer to a chart or coordinate system. Let

$$U_0 = \{[x:y:z] \in \mathbb{R}P^2 : x \neq 0\}, \quad U_1 = \{[x:y:z] \in \mathbb{R}P^2 : y \neq 0\}, \quad \text{and} \quad U_2 = \{[x:y:z] \in \mathbb{R}P^2 : z \neq 0\}. \tag{3.1}$$

Note that they are well-defined open subsets of $\mathbb{R}P^2$—consider U_0 for instance: given $\lambda \neq 0$ and a point (x,y,z), we have that $\lambda x \neq 0$ if and only if $x \neq 0$, while $\pi^{-1}(U_0) = \{(x,y,z) \in \mathbb{R}^3 \smallsetminus \{0\} : x \neq 0\}$ is clearly open in $\mathbb{R}^3 \smallsetminus \{0\}$. In other words, taking an inverse image under π of such sets amounts to replacing $[x:y:z]$ with (x,y,z). Note that whenever $[x:y:z] \in U_0$, we may "normalize" the x-component and write $[x:y:z] = [1:y/x:z/x]$. This tells us how to build charts $\varphi_i\colon U_i \to \mathbb{R}^2$ for $\mathbb{R}P^2$: just let

$$\varphi_0([x:y:z]) = \left(\frac{y}{x}, \frac{z}{x}\right), \ \varphi_1([x:y:z]) = \left(\frac{x}{y}, \frac{z}{y}\right), \ \text{and} \ \varphi_2([x:y:z]) = \left(\frac{x}{z}, \frac{y}{z}\right).$$

They are all continuous due to the characteristic property of the quotient topology, as $\pi^{-1}(U_0) \ni (x,y,z) \mapsto (y/x, z/x) \in \mathbb{R}^2$ is continuous (here, we use Proposition 9, p. 24), and similarly for the other two maps. Their inverses $\varphi_i^{-1}\colon \mathbb{R}^2 \to U_i$, given by

$$\varphi_0^{-1}(u,v) = [1:u:v], \quad \varphi_1^{-1}(u,v) = [u:1:v], \quad \text{and} \quad \varphi_2^{-1}(u,v) = [1:u:v],$$

are also continuous, being the composition of each of the continuous functions $(u,v) \mapsto (1,u,v)$, $(u,v) \mapsto (u,1,v)$, and $(u,v) \mapsto (u,v,1)$, with π. This shows that $\mathbb{R}P^2$ is locally Euclidean of dimension 2. Being homeomorphic to $\mathbb{S}^2/\mathbb{Z}_2$ (Example 35, p. 42), $\mathbb{R}P^2$ is also Hausdorff (Example 28, p. 33) and second-countable (by Proposition 7 in p. 23, in view of Example 18 in p. 24), and hence a topological manifold.

We now claim that $\mathfrak{A} = \{(U_i, \varphi_i)\}_{i=0}^2$ is a C^∞-atlas, and show in detail that the charts (U_0, φ_0) and (U_1, φ_1) are C^∞-compatible: their domains are explicitly given by $\varphi_0(U_0 \cap U_1) = \varphi_1(U_0 \cap U_1) = (\mathbb{R} \smallsetminus \{0\}) \times \mathbb{R}$ and both transitions

$$\varphi_1 \circ \varphi_0^{-1}, \varphi_0 \circ \varphi_1^{-1}\colon (\mathbb{R} \smallsetminus \{0\}) \times \mathbb{R} \to (\mathbb{R} \smallsetminus \{0\}) \times \mathbb{R}$$

computed as

$$(\varphi_1 \circ \varphi_0^{-1})(u,v) = \varphi_1([1:u:v]) = \left(\frac{1}{u}, \frac{v}{u}\right),$$
$$(\varphi_0 \circ \varphi_1^{-1})(u,v) = \varphi_0([u:1:v]) = \left(\frac{1}{u}, \frac{v}{u}\right),$$

are of class C^∞ as their components are rational functions. Similarly, on their appropriate domains, all other possible transitions

$$\varphi_0 \circ \varphi_2^{-1}, \quad \varphi_2 \circ \varphi_0^{-1}, \quad \varphi_1 \circ \varphi_2^{-1}, \quad \text{and} \quad \varphi_2 \circ \varphi_1^{-1}$$

are also smooth.

Exercise 55 (A canonical C^∞-atlas for $\mathbb{R}P^n$)

Generalize Example 64 and exhibit a C^∞-atlas $\{(U_i, \varphi_i)\}_{i=0}^n$ for the real projective space $\mathbb{R}P^n$, verifying in detail its compatibility conditions.

The notion of C^k-compatibility of charts, however, has a crucial flaw: it is not an equivalence relation. It is obviously reflexive (revisit Example 63), and symmetry is built into its definition since *both* $\psi \circ \varphi^{-1}$ and $\varphi \circ \psi^{-1}$ are required to be of class C^k, but it in general fails to be transitive: if (U, φ) and (V, ψ) are C^k-compatible, as well as (V, ψ) and (W, σ), we may only guarantee that $\sigma \circ \varphi^{-1} \colon \varphi(U \cap W) \to \sigma(U \cap W)$ is of class C^k on the open subset $\varphi(U \cap V \cap W) \subseteq \varphi(U \cap W)$ (the "triple intersection"), which is where expressing $\sigma \circ \varphi^{-1} = (\sigma \circ \psi^{-1}) \circ (\psi \circ \varphi^{-1})$ as a composition of C^k-functions is allowed; see Figure 36.

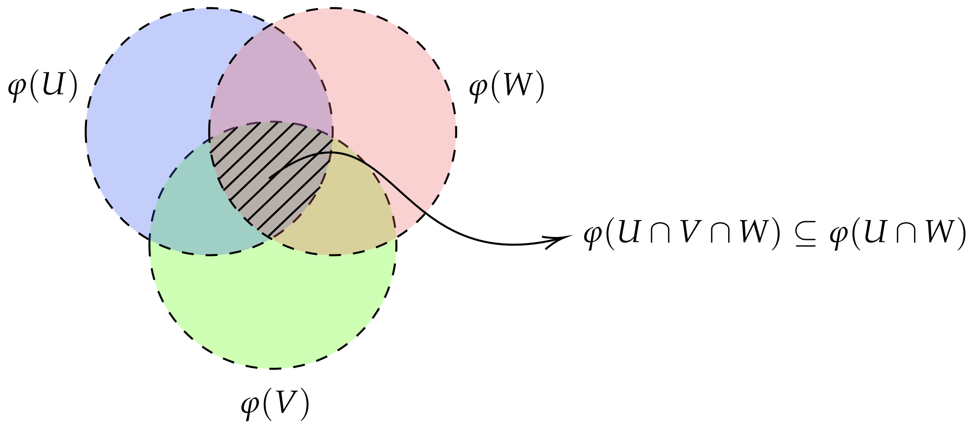


Figure 36: The failure of the transitivity property for C^k-compatibility of charts.

The solution to this problem is to extend the notion of C^k-compatibility:

Definition 27 (C^k-compatibility of chart and atlas)

Let M be a topological manifold and $k \in \{1, 2 \ldots, \infty\}$; let (U, φ) be a chart for M, and $\mathfrak{A}$ be a C^k-atlas for M. We say that (U, φ) and $\mathfrak{A}$ are C^k**-compatible** if (U, φ) is C^k-compatible with every chart in $\mathfrak{A}$ or, equivalently, if the union $\mathfrak{A} \cup \{(U, \varphi)\}$ is also a C^k-atlas for M.

This gives us the missing transitivity property: we replace the chart (V, ψ) in the discussion preceding Figure 36 with an entire atlas. Instead of having the triple intersection be fixed, we now have the freedom to say that any point in $\varphi(U \cap W)$ lies inside the image of *some* such triple intersection.

Lemma 5

Let M be a topological manifold and $k \in \{1, 2 \ldots, \infty\}$, and $\mathfrak{A}$ be a C^k-atlas for M. If (U, φ) and (W, σ) are both C^k-compatible with $\mathfrak{A}$, then they are C^k-compatible.

Proof: Consider the transition $\sigma \circ \varphi^{-1} \colon \varphi(U \cap W) \to \sigma(U \cap W)$. Since "$C^k$-ness" is a local notion, it suffices to show that $\sigma \circ \varphi^{-1}$ is of class C^k in an open neighborhood of each point $x \in \varphi(U \cap W)$. This neighborhood will be the image $\varphi(U \cap V \cap W)$ of a triple intersection, where $(V, \psi) \in \mathfrak{A}$ is a chart for M around $\varphi^{-1}(x)$. Namely, in $\varphi(U \cap V \cap W)$, we may write $\sigma \circ \varphi^{-1} = (\sigma \circ \psi^{-1}) \circ (\psi \circ \varphi^{-1})$ as a composition of C^k-functions. Therefore, $\sigma \circ \varphi^{-1}$ is of class C^k. We may prove with the same argument that $\varphi \circ \sigma^{-1}$ is of class C^k, and so (U, φ) and (W, σ) are C^k-compatible, as required. □

With Lemma 5 in place, there is one last issue to address. Even if M is a topological manifold admitting a C^k-atlas $\mathfrak{A}$, what is to say that $\mathfrak{A}$ is the best possible such atlas? If $(U, \varphi) \in \mathfrak{A}$, and we have that $U' \subseteq U$ is an open subset and $F \colon \mathbb{R}^n \to \mathbb{R}^n$ is a C^k-diffeomorphism, we may very well consider the charts $(U', \varphi|_{U'})$ and $(F \circ \varphi, U)$, which are C^k-compatible with $\mathfrak{A}$, but they are *not guaranteed to be in* $\mathfrak{A}$. One way to ensure that everything that could conceivably be a chart is indeed a valid chart in $\mathfrak{A}$, is to impose a maximality condition on $\mathfrak{A}$.

Definition 28 (C^k-manifolds)

Let $k \in \{1, 2 \ldots, \infty\}$.

- A C^k**-structure** on a topological manifold is a C^k-atlas $\mathfrak{A}$ which is **maximal**, that is, whenever (U, φ) is a chart for M which is C^k-compatible with $\mathfrak{A}$, we have $(U, \varphi) \in \mathfrak{A}$.
- A C^k**-manifold** is a topological manifold equipped with a C^k-structure.
- A C^∞-structure and a C^∞-manifold are simply called a **smooth structure** and a **smooth manifold**.

Remark. There are other conditions one can impose on homeomorphisms between open subsets of Euclidean spaces, e.g. being piecewise-linear or real-analytic. Therefore, it makes sense to talk about piecewise-linear manifolds, or real-analytic manifolds. But we will not pursue these directions here.

Definition 28 would mean that, in principle, to make a topological manifold M into a C^k-manifold, we would have to find not just a C^k-atlas for M, but instead a C^k-structure on M. The atlases found on all the examples presented so far are certainly not maximal. However, here is the saving grace of it all: every C^k-atlas $\mathfrak{A}$ is contained in a unique C^k-structure $\mathfrak{A}_{\max}$—namely, we let $\mathfrak{A}_{\max}$ be the collection of all possible charts for M which are C^k-compatible with $\mathfrak{A}$. Then $\mathfrak{A}_{\max}$ is maximal by design, and hence **finite-dimensional real vector spaces, spheres, and real projective spaces, can all be made into smooth manifolds**: just consider the smooth structures determined by the C^∞-atlases presented in Example 57 and Exercises 54 and 55. **Open**

subsets of smooth manifolds also become smooth manifolds, with the smooth structure induced by the atlas built with the charts described in Example 58. In the exercise that follows, be careful with the distinction between a "maximal element" (relative to an ordering) and a maximal atlas (in the sense of Definition 28).

The description of $\mathfrak{A}_{\max}$ given above might seem a bit too vague and perhaps not formal enough. To make it precise, recall **Zorn's Lemma**: if $(\mathscr{P}, \leq)$ is a non-empty partially ordered set with the property that every totally ordered subset of $\mathscr{P}$ has an upper bound, then $(\mathscr{P}, \leq)$ has a maximal element.

Exercise 56 (Existence and uniqueness of C^k-structures)

Let M be a topological manifold, $k \in \{1, 2, \ldots, \infty\}$, and $\mathfrak{A}$ be a C^k atlas for M.

(a) Consider $\mathscr{P} = \{\mathfrak{B} : \mathfrak{B} \text{ is a } C^k \text{ atlas for } M \text{ and } \mathfrak{A} \subseteq \mathfrak{B}\}$, ordered by inclusion. Show that $\mathscr{P} \neq \varnothing$ and that $\bigcup_{\mathfrak{B} \in \mathscr{C}} \mathfrak{B} \in \mathscr{P}$ whenever $\mathscr{C} \subseteq \mathscr{P}$ is totally ordered.

As the union $\bigcup_{\mathfrak{B} \in \mathscr{C}} \mathfrak{B}$ is clearly an upper bound for $\mathscr{C}$, Zorn's Lemma kicks in and gives us a maximal element $\mathfrak{A}_{\max} \in \mathscr{P}$. Show that:

(b) $\mathfrak{A}_{\max}$ is a C^k-structure for M.

(c) if $\mathfrak{A}'$ is another C^k-structure with $\mathfrak{A} \subseteq \mathfrak{A}'$, then $\mathfrak{A}_{\max} = \mathfrak{A}'$.
Hint: the argument is "synthetic" and uses no topology whatsoever.

Whether a topological space is a topological manifold or not has a clear-cut "yes" or "no" answer: we just need to check whether it is Hausdorff, second-countable, and locally Euclidean. As for how to turn a topological manifold into a C^k-manifold, things are more subtle. Although we now know that every C^k-atlas is contained in a unique C^k-structure, this does not mean that a given topological manifold cannot admit more than one possible C^k-structure, or even that it admits one at all. Case in point:

Exercise 57 (Uncountably many smooth structures on $\mathbb{R}$)

Consider $\mathbb{R}$ with its usual Euclidean topology and, for every $r > 0$, let $\varphi_r \colon \mathbb{R} \to \mathbb{R}$ be given by

$$\varphi_r(t) = \begin{cases} t, & \text{if } t \leq 0 \\ rt, & \text{if } t > 0 \end{cases}$$

Show that:

(a) for each $r > 0$, the pair $(\mathbb{R}, \varphi_r)$ is a chart for $\mathbb{R}$.

(b) if $r, s > 0$ are distinct, then φ_r and φ_s are not C^k-compatible for any $1 \leq k \leq \infty$.

This means that the smooth structures $\mathfrak{A}_r$ on $\mathbb{R}$, each determined by a chart $(\mathbb{R}, \varphi_r)$, are all pairwise distinct; note also that $\mathfrak{A}_1$ is just the standard Euclidean smooth structure on $\mathbb{R}$. We will see later in Exercise 66 that this is not too bad: all such $\mathfrak{A}_r$ are diffeomorphic to each other.

We will mention some more advanced results in these directions later, but it clearly becomes necessary to understand when two given C^k-atlases on the same topological manifold determine the same C^k-structure. The next definition should not be too surprising:

Definition 29 (C^k-compatibility of atlases)

Let M be a topological manifold, and $k \in \{1, 2, \ldots, \infty\}$. Two C^k-atlases $\mathfrak{A}_1$ and $\mathfrak{A}_2$ for M are called **C^k-compatible** if every chart in $\mathfrak{A}_1$ is C^k-compatible with every chart in $\mathfrak{A}_2$ or, equivalently, if the union $\mathfrak{A}_1 \cup \mathfrak{A}_2$ is also a C^k-atlas for M.

Exercise 58

Using Lemma 5, show that C^k-compatibility of atlases is an equivalence relation.

Theorem 8

Let M be a topological manifold, $k \in \{1, 2, \ldots, \infty\}$, and $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be two C^k-atlases for M. Then $\mathfrak{A}_1$ and $\mathfrak{A}_2$ determine the same C^k-structure if and only if $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are C^k-compatible.

Proof: First of all, note that two comparable C^k-atlases (that is, one of them is contained in the other) are necessarily C^k-compatible. Now, let $\mathfrak{A}_1'$ and $\mathfrak{A}_2'$ be the C^k-structures containing $\mathfrak{A}_1$ and $\mathfrak{A}_2$, respectively. We will repeatedly use transitivity of the relation of C^k-compatibility between atlases.

Assume that $\mathfrak{A}_1' = \mathfrak{A}_2'$, and let us denote this C^k-structure simply by $\mathfrak{A}'$. Then both $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are C^k-compatible with $\mathfrak{A}'$, and hence C^k-compatible with each other.

Conversely, assume that $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are C^k-compatible. Since $\mathfrak{A}_1$ is C^k-compatible with $\mathfrak{A}_1'$ and $\mathfrak{A}_2$ is C^k-compatible with $\mathfrak{A}_2'$, it follows that $\mathfrak{A}_1'$ and $\mathfrak{A}_2'$ are C^k-compatible. Now $\mathfrak{A}_1' \cup \mathfrak{A}_2'$ is a C^k-atlas (in fact, a C^k-structure) containing $\mathfrak{A}_1$, and so maximality of $\mathfrak{A}_1'$ implies that $\mathfrak{A}_1' \cup \mathfrak{A}_2' \subseteq \mathfrak{A}_1'$, leading to $\mathfrak{A}_1' \cup \mathfrak{A}_2' = \mathfrak{A}_1'$. Similarly, $\mathfrak{A}_1' \cup \mathfrak{A}_2' = \mathfrak{A}_2'$. We conclude that $\mathfrak{A}_1' = \mathfrak{A}_2'$, as required. □

Some texts define a C^k-structure to be an equivalence class for C^k-compatibility of C^k-atlases. Theorem 8 above says that this approach is equivalent to the one we chose to present here. More concretely, both notions are reconciled by noting that such an equivalence class has a canonical representative: the maximal one.

Example 65 (Standard smooth structures on vector spaces)

If V is a finite-dimensional real vector space, the smooth structure on V determined by the atlas containing the single global linear chart induced by some basis of V in fact does not depend on the choice of basis, as a direct consequence of the second half of Example 63 combined with Theorem 8. We call it the **standard smooth structure of V**.

Exercise 59 (Product manifolds)

Let M and N be topological manifolds, with $n = \dim M$ and $m = \dim N$, and C^k-atlases $\mathfrak{A} = \{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ and $\mathfrak{B} = \{(V_\lambda, \psi_\lambda)\}_{\lambda \in \Lambda}$. Show that

$$\mathfrak{C} = \{(U_\alpha \times V_\lambda, \varphi_\alpha \times \psi_\lambda)\}_{(\alpha,\lambda) \in A \times \Lambda}$$

is a C^k-atlas for the product space $M \times N$. Conclude that $M \times N$ is a C^k-manifold as well, with $\dim(M \times N) = n + m$.

Exercise 60 (Stereographic projections)

Write $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$, and consider the unit sphere $\mathbb{S}^n = \{p \in \mathbb{R}^{n+1} : \|p\| = 1\}$. Denote the north and south poles of $\mathbb{S}^n$ by $(0, \pm 1)$, and define two mappings $\mathrm{St}_\pm \colon \mathbb{S}^n \smallsetminus \{(0, \pm 1)\} \to \mathbb{R}^n$ as follows: for $p \in \mathbb{S}^n$, $(\mathrm{St}_\pm(p), 0)$ is the intersection between $\mathbb{R}^n \times \{0\}$ and the line joining $(0, \pm 1)$ and p. See Figure 37.

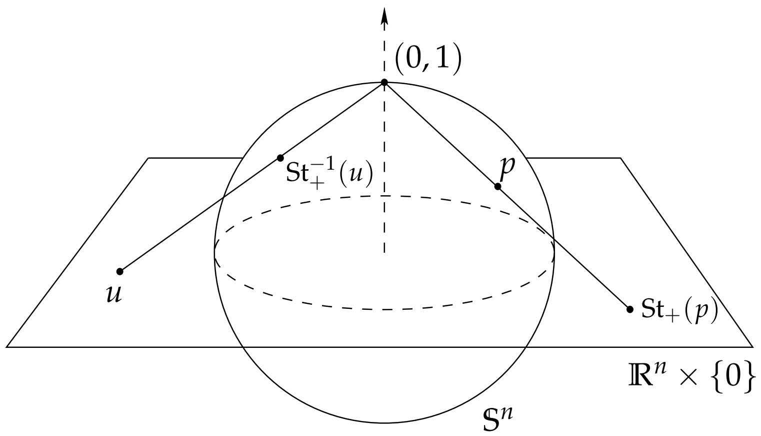


Figure 37: The stereographic projection St_+.

(a) Show that

$$\mathrm{St}_\pm(p) = \frac{p_0}{1 \mp p_{n+1}} \quad \text{and} \quad \mathrm{St}_\pm^{-1}(u) = \left(\frac{2u}{\|u\|^2 + 1}, \pm \frac{\|u\|^2 - 1}{\|u\|^2 + 1} \right)$$

for all $p = (p_0, p_{n+1}) \in \mathbb{S}^n \smallsetminus \{(0, \pm 1)\}$ and $u \in \mathbb{R}^n$.
Hint: Don't try to invert the formula for $\mathrm{St}_\pm(p)$ directly, think of Figure 37.

(b) Check that the transition $\mathrm{St}_+ \circ \mathrm{St}_-^{-1} \colon \mathbb{R}^n \smallsetminus \{0\} \to \mathbb{R}^n \smallsetminus \{0\}$ is of class C^∞.
Hint: The formula for the transition will come out cleaner than you expect.

A similar calculation shows that $\mathrm{St}_- \circ \mathrm{St}_+^{-1} \colon \mathbb{R}^n \smallsetminus \{0\} \to \mathbb{R}^n \smallsetminus \{0\}$ is also of class C^∞. Hence $\mathfrak{A}_{\mathrm{St}} = \{(\mathbb{S}^n \smallsetminus \{(0,1)\}, \mathrm{St}_+), (\mathbb{S}^n \smallsetminus \{(0,-1)\}, \mathrm{St}_-)\}$ defines a C^∞-atlas for $\mathbb{S}^n$. We call it the **stereographic atlas** of $\mathbb{S}^n$.

(c) Show that $\mathfrak{A}_{\mathrm{St}}$ is C^∞-compatible with the atlas suggested in Exercise 54.

Moving forward, we will focus solely on smooth manifolds instead of more general C^k-manifolds. The cheap reason is that we don't want to bother with keeping track of regularity classes of functions and manifolds, or counting how many derivatives each proof to come requires. The more legitimate reason is the following advanced result, originally due to Whitney [25]: whenever $1 \leq k < \ell \leq \infty$, every C^k-structure contains a unique C^ℓ-structure (you may also find a modern proof in [10, p. 51], but it requires some analysis). In other words, starting with a C^k-structure, we may simply throw away charts until what remains is a smooth structure (of course, arguably the problem with that is that fewer functions defined on the manifold would be smooth, in the sense to be discussed in Section 4.1 ahead).

For topological manifolds of dimensions 1, 2, and 3, there always exists a smooth structure, which is essentially unique (that is, unique up to diffeomorphism): namely, every 1-dimensional smooth manifold is diffeomorphic to either $\mathbb{R}$ or $\mathbb{S}^1$—see e.g. [16, Appendix]—while for $n = 2$ it is due to Radó [21], and for $n = 3$ to Moise [18]. When the dimension is $n \geq 4$, Stranger Things™ may happen: there are topological manifolds having many non-equivalent smooth structures (the ones not equivalent to the "standard" ones are often called "exotic"), and there are topological manifolds which admit no smooth structure whatsoever. In more detail:

- Euclidean space $\mathbb{R}^n$ has a smooth structure that is unique up to diffeomorphism whenever $n \neq 4$: the case $n \geq 3$ is covered by the results just mentioned, and for $n \geq 5$ the first proof seems to be due to Stallings [22]. But $\mathbb{R}^4$, on the other hand, has uncountably many non-diffeomorphic smooth structures, cf. Gompf [7], Donaldson [3, 4], and Freedman and Taylor [5], etc.

- As for spheres $\mathbb{S}^n$: Kervaire and Milnor have shown that there are 28 distinct diffeomorphism classes of $\mathbb{S}^7$'s [14], and 16 of them (including the standard $\mathbb{S}^7$) are realized by Milnor's classical construction as $\mathbb{S}^3$-bundles over $\mathbb{S}^4$ [17]. The four-dimensional case, of course, is the outlier—either $\mathbb{S}^4$ has infinitely many non-diffeomorphic exotic structures, or none at all [6].

- The first example of a topological manifold which does not admit any smooth structure at all has dimension 10 and is due to Kervaire [13].

Next, we will see how to do Calculus on smooth manifolds.

4 Differential calculus on smooth manifolds

4.1 Smooth functions

Now that the definition of a smooth manifold is in place, we may start redeveloping in this new setting some of the well-known concepts from multivariable calculus. Whenever a smooth manifold M is given, and we refer to a chart (U, φ) for M, it will be always understood $(U, \varphi) \in \mathfrak{A}$, where $\mathfrak{A}$ is the smooth structure which turns M into a smooth manifold; no explicit mention to $\mathfrak{A}$ will be made.

The first concept we will revisit is the one of smoothness. Since there is already a concept of smoothness for functions between open subsets of Euclidean spaces, we will temporarily refer to it as "Euclidean-smoothness". Then, "smoothness" of a function between smooth manifolds will be defined in terms of "Euclidean-smoothness", but it is important to keep in mind that these concepts are technically different. If this distinction is not clear to you, many arguments involving smoothness may sound like "... and this function is smooth because it is smooth," but there is something actually happening there.

We begin with real-valued functions.

Definition 30 (Smoothness of real-valued functions on manifolds)

Let M be a smooth manifold of dimension n. A function $f \colon M \to \mathbb{R}$ is said to be **smooth** if for every chart (U, φ) for M, the composition $f \circ \varphi^{-1} \colon \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ (also called a **local expression** for f) is Euclidean-smooth.

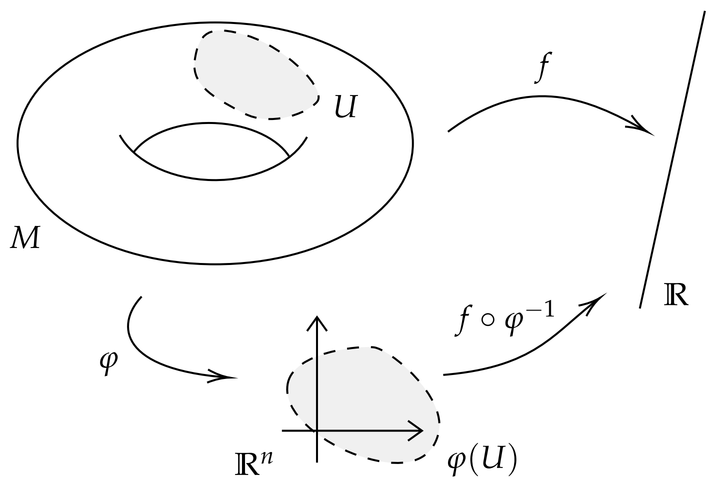


Figure 38: Definition of smoothness for a real-valued function on a manifold.

With pointwise operations, $C^\infty(M) = \{f \colon M \to \mathbb{R} : f \text{ is smooth}\}$ becomes an algebra over $\mathbb{R}$.

The above definition would require us to check that $f \circ \varphi^{-1} \colon \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ is Euclidean-smooth for *every single chart* in the smooth structure of M—and at this point we know that there simply might be too many of them. Here is where maximality of smooth structures comes to the rescue: it in fact suffices to establish Euclidean-smoothness of $f \circ \varphi^{-1}$ for enough charts to cover M. The reason for this, at the end of the day, is that Euclidean-smoothness (and hence also smoothness!) is a local notion. If (U, φ) and (V, ψ) are two charts for M with $U \cap V \neq \varnothing$, the transition mapping $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ is a diffeomorphism, and so whenever $p \in U \cap V$ we have that $f \circ \varphi^{-1}$ is Euclidean-smooth near $\varphi(p)$ if and only if $f \circ \psi^{-1}$ is Euclidean-smooth near $\psi(p)$, as a consequence of the relation $f \circ \varphi^{-1} = (f \circ \psi^{-1}) \circ (\psi \circ \varphi^{-1})$ on $\varphi(U \cap V)$; "near" meaning "in some neighborhood of". See Figure 39.

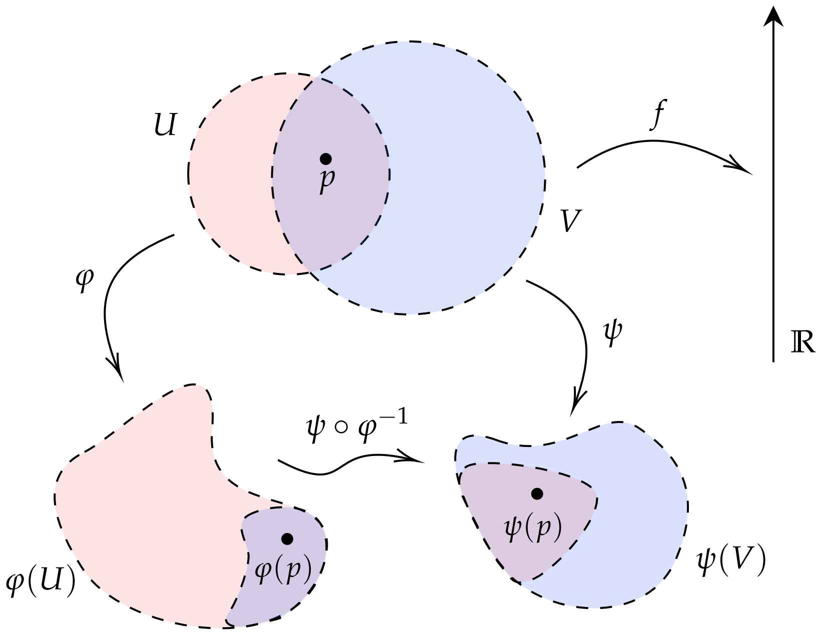


Figure 39: Smoothness of $f \circ \varphi^{-1}$ at $\varphi(p)$ is equivalent to the one of $f \circ \psi^{-1}$ at $\psi(p)$.

We proceed to the more general case, where the codomain $\mathbb{R}$ is replaced with a second smooth manifold N:

Definition 31 (Smoothness of mappings between manifolds)

Let M and N be smooth manifolds of dimensions n and m, respectively, and $F\colon M \to N$ be a continuous mapping. Then F is called **smooth** if for all charts (U, φ) and (V, ψ) for M and N, the **local representation**

$$\psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \subseteq \mathbb{R}^n \to \mathbb{R}^m$$

is Euclidean-smooth. Finally, F is called a **diffeomorphism** if it smooth, bijective, and the inverse F^{-1} is also smooth.

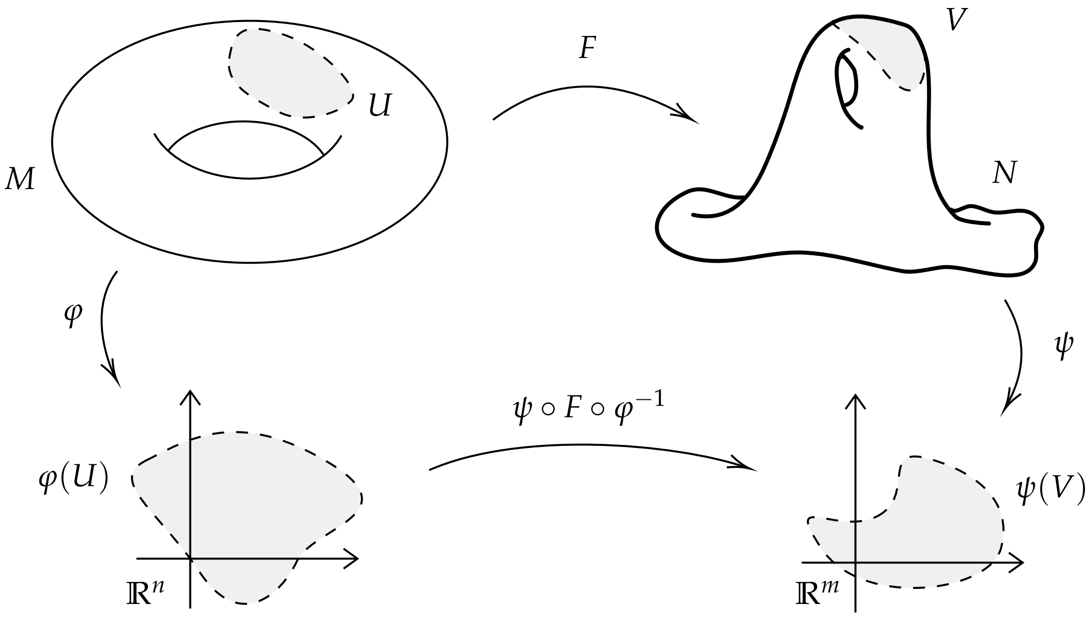


Figure 40: Definition of smoothness for a mapping between manifolds.

Remark. This time, the function F must be assumed to be continuous from the beginning, so that $F^{-1}(V)$ (and hence $\varphi(U \cap F^{-1}(V))$) is open. If this assumption is not made, there are examples of smooth functions which are not continuous, but this is not really something acceptable for us. See [15, Problem 2-1]. And, as before, it is suffices to establish Euclidean-smoothness of local representations just for enough charts to cover the domain M.

When $N = \mathbb{R}$ is equipped with its standard smooth structure, we take the chart $(V, \psi) = (\mathbb{R}, \mathrm{Id}_{\mathbb{R}})$ by default, and Definition 31 reduces to Definition 30. If both M and N are open subsets of Euclidean spaces, then smoothness and Euclidean-smoothness agree.

Proposition 28 (Composition of smooth mappings is smooth)

If M, N, and P are smooth manifolds, and $F\colon M \to N$ and $G\colon N \to P$ are smooth mappings, the composition $G \circ F\colon M \to P$ is also smooth.

Proof: Let $p \in M$, and choose charts (U, φ), (V, ψ), (W, σ) be charts for M, N, and P, around p, $F(p)$, and $G(F(p))$, respectively, for which both $\psi \circ F \circ \varphi^{-1}$ and $\sigma \circ G \circ \psi^{-1}$ are well-defined and Euclidean-smooth on their domains. Then

$$\sigma \circ (G \circ F) \circ \varphi^{-1} = (\sigma \circ G \circ \psi^{-1}) \circ (\psi \circ F \circ \varphi^{-1})$$

is a composition of Euclidean-smooth functions where defined, hence Euclidean-smooth as well. □

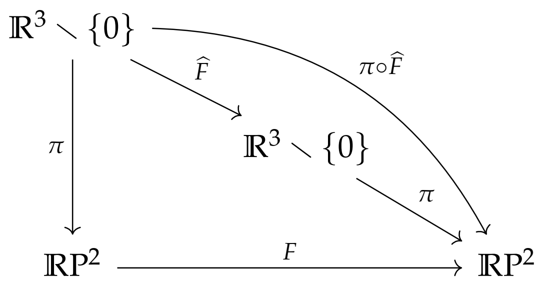
Next, we discuss some examples.

Example 66 (Charts become diffeomorphisms by default)

If M is any smooth manifold (say, of dimension n) and (U, φ) is any chart for M, then the chart $\varphi\colon U \subseteq M \to \varphi(U) \subseteq \mathbb{R}^n$ itself is a diffeomorphism; here, we consider the open subsets U and $\varphi(U)$ with their smooth structures induced from M and $\mathbb{R}^n$, respectively.

Example 67

Let $F\colon \mathbb{RP}^2 \to \mathbb{RP}^2$ be given by $F([x:y:z]) = [yz : xz : xy]$. First of all, we argue that F is continuous. Indeed, the lift $\widehat{F}\colon \mathbb{R}^3 \smallsetminus \{0\} \to \mathbb{R}^3 \smallsetminus \{0\}$ given by $\widehat{F}(x,y,z) = (yz, xz, xy)$ is clearly continuous, making $\pi \circ \widehat{F}\colon \mathbb{R}^3 \smallsetminus \{0\} \to \mathbb{RP}^2$ continuous as well (it is a composition of continuous functions), allowing us to invoke the characteristic property of the quotient topology of $\mathbb{RP}^2$ (see Proposition 5, p. 19):



Continuity of F now follows from continuity of $F \circ \pi = \pi \circ \widehat{F}$. With this in place, we now claim that F is smooth. To check this, we use the atlas $\{(U_i, \varphi_i)\}_{i=0}^2$ from Example 64 (p. 74), and compute

$$(\varphi_0 \circ F \circ \varphi_0^{-1})(u,v) = (\varphi_0 \circ F)([1:u:v]) = \varphi_0([uv : v : u]) = \left(\frac{v}{uv}, \frac{u}{uv}\right) = \left(\frac{1}{u}, \frac{1}{v}\right).$$

Being a rational function, it is Euclidean-smooth where defined. We similarly compute the other local representations, e.g.,

$$(\varphi_0 \circ F \circ \varphi_1^{-1})(u,v) = \left(u, \frac{u}{v}\right) \quad \text{and} \quad (\varphi_0 \circ F \circ \varphi_2^{-1})(u,v) = \left(\frac{u}{v}, u\right),$$

and also $\varphi_1 \circ F \circ \varphi_0^{-1}$, $\varphi_1 \circ F \circ \varphi_1^{-1}$, etc. We will develop very effective shortcuts for avoiding these calculations later.

Exercise 61

Show, in detail, that the function $f\colon \mathbb{RP}^2 \to \mathbb{R}$ given by

$$f([x:y:z]) = \frac{xy + yz + xz}{x^2 + y^2 + z^2}$$

is well-defined and smooth.

Example 68 (Smoothness of the projection onto $\mathbb{RP}^n$)

The quotient projection $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{RP}^n$ is smooth. We take the identity chart on $\mathbb{R}^{n+1} \smallsetminus \{0\}$, and one of the standard charts (U_i, φ_i) for $\mathbb{RP}^n$ (cf. Exercise 55, p. 76), and compute the local representation $\varphi_i \circ \pi\colon \pi^{-1}(U_i) \to \mathbb{R}^n$ as

$$(\varphi_i \circ \pi)(x_0, x_1, \ldots, x_n) = \left(\frac{x_0}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_n}{x_i}\right).$$

As rational functions are Euclidean-smooth and $\{(U_i, \varphi_i)\}_{i=0}^n$ covers $\mathbb{RP}^n$, it follows that π is smooth.

The situation from Example 67 is a particular case of a very general phenomenon:

Exercise 62

Let $P\colon \mathbb{R}^3 \smallsetminus \{0\} \to \mathbb{R}^3 \smallsetminus \{0\}$ be a smooth function, and suppose that, for some $d \in \mathbb{Z}$, the function P is *homogeneous of degree d*: it holds that

$$P(\lambda x, \lambda y, \lambda z) = \lambda^d P(x, y, z), \quad \text{for all } \lambda \in \mathbb{R} \smallsetminus \{0\} \text{ and } (x, y, z) \in \mathbb{R}^3 \smallsetminus \{0\}.$$

Show that the induced function $\widetilde{P}\colon \mathbb{RP}^2 \to \mathbb{RP}^2$, characterized by the relation $\widetilde{P}([x : y : z]) = [P(x, y, z)]$, is well-defined and smooth.

Hint: The conclusion should obviously remain true for homogeneous functions $P\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{R}^{k+1} \smallsetminus \{0\}$, inducing a smooth mapping $\widetilde{P}\colon \mathbb{RP}^n \to \mathbb{RP}^k$.

In Example 33 (p. 41) we have seen that the restricted projection $\pi|_{\mathbb{S}^n}\colon \mathbb{S}^n \to \mathbb{RP}^n$ is continuous and surjective, and in Example 35 (p. 42) that it induced a homeomorphism $\mathbb{S}^n/\mathbb{Z}_2 \cong \mathbb{RP}^n$. Recall here that $\mathbb{S}^n/\mathbb{Z}_2$ denotes the quotient of $\mathbb{S}^n$ under the action of the group $\mathbb{Z}_2$, with $(-1) \cdot p = -p$, cf. Example 18 (p. 24).

Exercise 63

Show that $\pi|_{\mathbb{S}^n}\colon \mathbb{S}^n \to \mathbb{RP}^n$ is smooth.

We will see in the next section how to give the quotient $\mathbb{S}^n/\mathbb{Z}_2$ a smooth structure, and the homeomorphism $\mathbb{S}^n/\mathbb{Z}_2 \cong \mathbb{RP}^n$ will then get upgraded to a diffeomorphism.

Still in the topic of spheres, in the next exercise we have a very nice description of real-valued functions defined on standard spheres.

Exercise 64 (A characterization of smooth functions on the sphere)

(a) Show that the function $\nu\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{S}^n$ given by $\nu(x) = x/\|x\|$ is smooth.
Hint: This is obvious if you consider the target of ν to be $\mathbb{R}^{n+1} \smallsetminus \{0\}$ and think about Euclidean-smoothness, but this is not what the exercise is about. The question is whether ν is smooth as a mapping between manifolds. By the

way, note that the Euclidean topology of $\mathbb{S}^n$ agrees with the quotient topology induced by ν (that is, ν is a quotient map—can you prove this?).

(b) Show that the inclusion $\iota\colon \mathbb{S}^n \to \mathbb{R}^{n+1} \smallsetminus \{0\}$ is smooth.

(c) Conclude from (a) and (b) that a continuous function $f\colon \mathbb{S}^n \to \mathbb{R}$ is smooth if and only if its **radial extension** $\widetilde{f}\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{R}$ given by $\widetilde{f}(x) = f(x/\|x\|)$ is Euclidean-smooth.

It will also turn out that if M is any manifold, a function $f\colon M \to \mathbb{S}^n$ is smooth if and only if $\iota \circ f\colon M \to \mathbb{R}^{n+1} \smallsetminus \{0\}$ is smooth. This phenomenon is not specific to the sphere, but instead relies on the fact that $\mathbb{S}^n$ is an *embedded submanifold* of $\mathbb{R}^{n+1} \smallsetminus \{0\}$. Think of it as a type of characteristic property for embedded submanifolds, analogous to what we had in Proposition 2 (p. 8). More on this later.

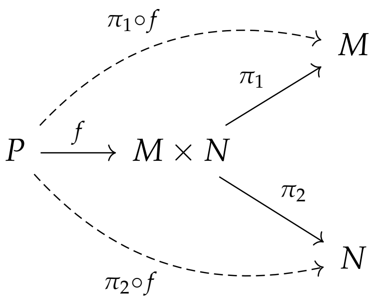
We have seen in Example 59 (p. 80) how products of smooth manifolds are again smooth manifolds. With this in place and the definitions introduced earlier in this section, it now makes sense to consider smoothness of cartesian projections.

Exercise 65 (Cartesian projections are smooth)

Let M and N be smooth manifolds, and consider their cartesian product $M \times N$, along with the canonical projections $\pi_1\colon M \times N \to M$ and $\pi_2\colon M \times N \to N$.

(a) Show that π_1 and π_2 are smooth.

(b) If P is a third smooth manifold, show that a function $f\colon P \to M \times N$ is smooth if and only if both compositions $\pi_1 \circ f\colon P \to M$ and $\pi_2 \circ f\colon P \to N$ (that is, the "components" of f) are smooth.



(c) Conclude that $(M \times N) \times P$ is diffeomorphic to $M \times (N \times P)$.

Moving on, recall from the end of Section 3 that all smooth structures on the real line are diffeomorphic. One of the usual proofs for this relies on the concept of a *partition of unity*, which we will not encounter in these notes until much later. In some nice cases, the diffeomorphism in question can be computed very explicitly. The next exercise presents a rather instructive example of this.

Exercise 66 (Uncountably many smooth structures on $\mathbb{R}$ redux)

In Exercise 57 (p. 78), we introduced a family $\{\mathfrak{A}_r : r > 0\}$ of smooth structures on the real line $\mathbb{R}$, where each $\mathfrak{A}_r$ is the smooth structure determined by $\{(\mathbb{R}, \varphi_r)\}$, for $\varphi_r\colon \mathbb{R} \to \mathbb{R}$ defined by $\varphi_r(t) = t$ if $t \leq 0$, and $\varphi_r(t) = rt$ if $t > 0$.

(a) Given any continuous function $f\colon \mathbb{R} \to \mathbb{R}$, compute the local representation of f relative to the charts $(\mathbb{R}, \varphi_r)$ and $(\mathbb{R}, \varphi_s)$. In other words, fill in the blanks:

$$(\varphi_s \circ f \circ \varphi_r^{-1})(x) = \begin{cases} \dots, & \text{if } x \leq 0 \text{ and } f(x) \leq 0, \\ \dots, & \text{if } x \leq 0 \text{ and } f(x) > 0, \\ \dots, & \text{if } x > 0 \text{ and } f(x) \leq 0, \\ \dots, & \text{if } x > 0 \text{ and } f(x) > 0. \end{cases}$$

(b) Find a function f for which $\varphi_s \circ f \circ \varphi_r^{-1} = \mathrm{Id}_{\mathbb{R}}$. This implies (very strongly) that f is a diffeomorphism $(\mathbb{R}, \mathfrak{A}_r) \to (\mathbb{R}, \mathfrak{A}_s)$.
Hint: Try f of the form $f = \varphi_a$ for some $a > 0$. Is there a natural guess for a?

This means that while all the smooth structures $\mathfrak{A}_r$ on $\mathbb{R}$ are pairwise distinct, they are still all diffeomorphic to each other.

Finally, the last exercise for this section in some sense justifies the central role that $\mathbb{R}^n$ plays in the definition of a smooth manifold. Namely, fix a smooth manifold E of dimension n. Let's say that a topological space M is "*locally* E" if for every $p \in M$ there is an "E-*valued chart*" around p, that is, an open subset $U \subseteq M$ with $p \in M$ and a homeomorphism $\varphi \colon U \to \varphi(U) \subseteq E$, where $\varphi(U)$ is open in E. An E-atlas $\mathfrak{A}$ would then be a collection of E-valued charts whose domains cover M, and we'll say that $\mathfrak{A}$ is of class C^k if for every $(U, \varphi), (V, \psi) \in \mathfrak{A}$, either $U \cap V = \varnothing$, or $U \cap V \neq \varnothing$ and both transitions $\psi \circ \varphi^{-1} \colon \varphi(U \cap V) \to \psi(U \cap V)$ and $\varphi \circ \psi^{-1} \colon \psi(U \cap V) \to \varphi(U \cap V)$ are functions of class C^k between open subsets of E (this makes sense since E is a smooth manifold).

Exercise 67

With the above setup and notation, show that if M is Hausdorff and second-countable, and admits a C^k-atlas of E-valued charts, then M is a C^k-manifold.

In other words, if you tried to come up with a new notion of manifold replacing $\mathbb{R}^n$ with, say $\mathbb{S}^n$ (and called it "locally spherical" instead of "locally Euclidean"), you wouldn't obtain anything new. To obtain something new, we need to replace $\mathbb{R}^n$ with something which is **not** a smooth manifold in the sense already being discussed here. For instance, the study of infinite-dimensional manifolds begins when $\mathbb{R}^n$ gets replaced with a **Banach space**, that is, a complete[5] normed vector space; the majority of the results on multivariable calculus we have encountered in Section 2 holds in the setting of Banach spaces, and this is what allows for a satisfactory theory to be developed.

[5]Completeness here means that every Cauchy sequence converges, just like you have already seen in real analysis.

4.2 A digression: smooth quotients

Here, we need the following definition:

Definition 32 (Local diffeomorphism)

Let M and N be smooth manifolds. A mapping $F\colon M \to N$ is called a **local diffeomorphism** if for every $p \in M$ there are neighborhoods $U \subseteq M$ and $V \subseteq N$ of p and $f(p)$, respectively, such that the restriction $F|_U\colon U \to V$ is a diffeomorphism.

The statement below is essentially taken from [12, p. 15], and we provide more details for its proof.

Theorem 9 (Free $\mathbb{Z}_2$-quotients of smooth manifolds are smooth)

Let M be a smooth manifold, and $\tau\colon M \to M$ be a **fixed-point-free involution**, that is, $\tau \circ \tau = \mathrm{Id}_M$ and $\tau(p) \neq p$ for every $p \in M$. Then the quotient M/τ of M under the equivalence relation $\sim$ given by $p \sim \tau(p)$ is a topological manifold, and it has a unique smooth structure for which the quotient projection $\pi\colon M \to M/\tau$ is a local diffeomorphism.

Remark. Giving τ as above is the same as giving a free action of the group $\mathbb{Z}_2$ on M, via $1 \cdot p = p$ and $(-1) \cdot p = \tau(p)$. In particular, when $M = \mathbb{S}^n$ and $\tau(p) = -p$ is the antipodal mapping, we have that $\mathbb{S}^n/\mathbb{Z}_2 \cong \mathbb{RP}^n$ (cf. Example 35, p. 42), this time as manifolds.

Proof: We start by equipping M/τ with the quotient topology induced by the natural projection[6] $\pi\colon M \to M/\tau$, and noting that π is an open mapping: namely, whenever $U \subseteq M$ is open, we have that $\pi^{-1}(\pi(U)) = U \cup \tau(U)$ is the union of open sets, and hence open (here we have used that τ is a diffeomorphism), making $\pi(U)$ an open subset of M/τ, by definition of quotient topology. See Figure 41.

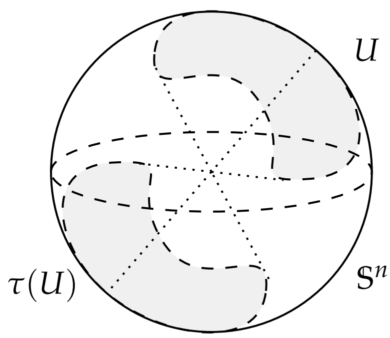


Figure 41: The equality $\pi^{-1}(\pi(U)) = U \cup \tau(U)$ for $\pi\colon \mathbb{S}^n \to \mathbb{RP}^n$.

We organize the rest of the proof in several claims.

[6] It is explicitly given by $\pi(p) = \{p, \tau(p)\}$, for each $p \in M$.

Claim 1: *M/τ is Hausdorff and second-countable.* Let $p, q \in M$ be points such that $\pi(p) \neq \pi(q)$ in M/τ. As M itself is Hausdorff, there are open subsets $U, V \subseteq M$ such that $p \in U$, $q \in V$, and $U \cap V = U \cap \tau(V) = \varnothing$, cf. Figure 42.

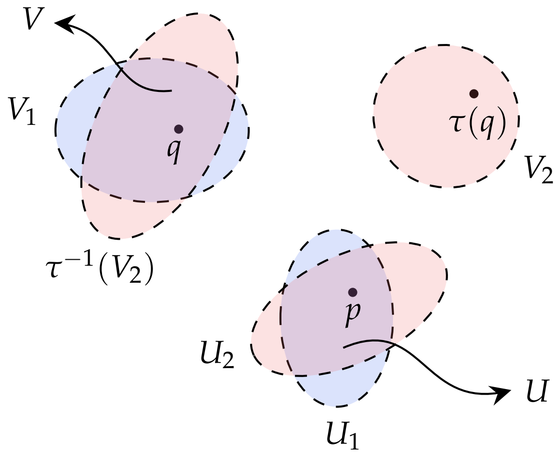


Figure 42: The Hausdorff property of M/τ.

In more detail, we use the Hausdorff property of M to fix disjoint open neighborhoods U_1 and V_1 of p and q, as well as disjoint open neighborhoods of p and $\tau(q)$. Then we let $U = U_1 \cap U_2$ and $V = V_1 \cap \tau^{-1}(V_2)$. In any case, $\pi(U)$ and $\pi(V)$ are now disjoint open neighborhoods of $\pi(p)$ and $\pi(q)$ in M/τ, making M/τ Hausdorff as required. As for second-countability of M/τ, it is immediate from Proposition 7 (p. 23).

Claim 2: *M/τ is locally Euclidean.* Let us say that an open subset $U \subseteq M$ is **τ-small** if $U \cap \tau(U) = \varnothing$ (for example, $U \subseteq \mathbb{S}^n$ in Figure 41 is τ-small). This condition is exactly what we need to say that the restriction $\pi|_U \colon U \to \pi(U)$ is injective, and hence a homeomorphism (continuity of $(\pi|_U)^{-1}$ follows from π being an open mapping). Now, we observe that every $p \in M$ has a τ-small neighborhood: as M is Hausdorff there are disjoint open neighborhoods V and W of p and $\tau(p)$ in M, and we may simply set $U = V \cap \tau^{-1}(W)$, cf. Figure 43.

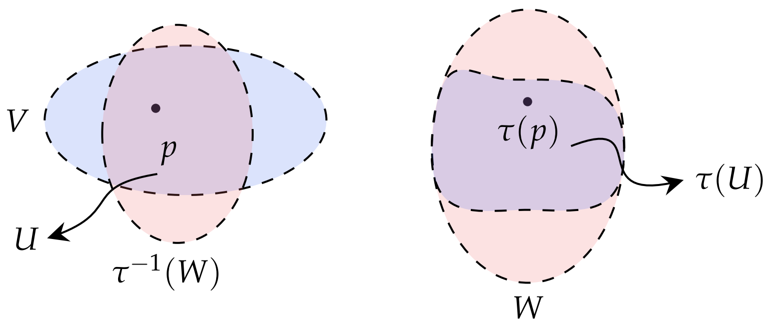


Figure 43: Constructing τ-small neighborhoods of arbitrary points.

Such U is indeed τ-small: if $x \in U \cap \tau(U)$, then $x, \tau^{-1}(x) \in U$, and so $x \in V$ and $\tau^{-1}(x) \in \tau^{-1}(W)$, leading to $x \in V \cap W$—a contradiction. With this in place, if $\mathfrak{A}$ denotes the smooth structure of M, the above shows that the collection

$$\mathfrak{A}_\tau = \{(\pi(U), \varphi \circ (\pi|_U)^{-1}) : (U, \varphi) \in \mathfrak{A} \text{ and } U \text{ is } \tau\text{-small}\} \tag{4.1}$$

is an atlas for M/τ.

Claim 3: *The atlas $\mathfrak{A}_\tau$ in (4.1) is a C^∞-atlas.* Indeed, if $(U,\varphi),(V,\psi) \in \mathfrak{A}$ have τ-small domains, the transitions between $(\pi(U), \varphi \circ (\pi|_U)^{-1})$ and $(\pi(V), \psi \circ (\pi|_V)^{-1})$ in fact agree with the transitions between (U,φ) and (V,ψ), which are smooth by default.

Claim 4: *$\pi\colon M \to M/\tau$ is a local diffeomorphism.* Just note that $\pi|_U\colon U \to \pi(U)$ is a diffeomorphism whenever $U \subseteq M$ is τ-small. Namely, π locally looks like the identity mapping if we choose a chart (U,φ) for M and the induced chart $(\pi(U), \varphi \circ (\pi|_U)^{-1})$, cf. the commutative diagram displayed on the right.

$$\begin{array}{ccc} U & \xrightarrow{\pi|_U} & \pi(U) \\ {\scriptstyle\varphi}\big\downarrow & & \big\downarrow{\scriptstyle\varphi\circ(\pi|_U)^{-1}} \\ \varphi(U) & \xrightarrow{\mathrm{Id}_{\varphi(U)}} & \varphi(U) \end{array}$$

Claim 5: *The smooth structure on M/τ for which the previous claim holds is unique.* Let $\mathfrak{A}'_\tau$ be the unique smooth structure containing $\mathfrak{A}_\tau$, cf. Exercise 56 (p. 78). Note that whenever N is another smooth manifold with smooth structure $\mathfrak{B}$, then a mapping $F\colon (M/\Gamma, \mathfrak{A}'_\tau) \to (N,\mathfrak{B})$ is smooth if and only if $F \circ \pi\colon (M,\mathfrak{A}) \to (N,\mathfrak{B})$ is smooth. And similarly if we replace $\mathfrak{A}'_\tau$ with another smooth structure $\mathfrak{A}'$ for which π becomes a local diffeomorphism:

$$\begin{array}{ccc} (M,\mathfrak{A}) & & \\ {\scriptstyle\pi}\big\downarrow & \searrow^{F\circ\pi} & \\ (M/\Gamma,\mathfrak{A}'_\tau) & \xrightarrow{F} & (N,\mathfrak{B}) \end{array} \qquad \begin{array}{ccc} (M,\mathfrak{A}) & & \\ {\scriptstyle\pi}\big\downarrow & \searrow^{F\circ\pi} & \\ (M/\Gamma,\mathfrak{A}') & \xrightarrow{F} & (N,\mathfrak{B}) \end{array}$$

Taking $(N,\mathfrak{B}) = (M/\Gamma,\mathfrak{A}')$ and $F = \mathrm{Id}_{M/\Gamma}$ in the first diagram and using smoothness of $\pi\colon (M,\mathfrak{A}) \to (M/\Gamma,\mathfrak{A}')$ now implies that $\mathrm{Id}_{M/\Gamma}\colon (M/\Gamma,\mathfrak{A}'_\tau) \to (M/\Gamma,\mathfrak{A}')$ is smooth. On the other hand, if we set $(N,\mathfrak{B}) = (M/\Gamma,\mathfrak{A}'_\tau)$ and $F = \mathrm{Id}_{M/\Gamma}$ in the second diagram, and use smoothness of $\pi\colon (M,\mathfrak{A}) \to (M/\Gamma,\mathfrak{A}'_\tau)$, we obtain that $\mathrm{Id}_{M/\Gamma}\colon (M/\Gamma,\mathfrak{A}') \to (M/\Gamma,\mathfrak{A}'_\tau)$ is smooth. Hence, $\mathrm{Id}_{M/\Gamma}$ is a diffeomorphism between $(M/\Gamma,\mathfrak{A}'_\tau)$ and $(M/\Gamma,\mathfrak{A}')$. This means that $\mathfrak{A}' = \mathfrak{A}'_\tau$. ☐

The generalization from $\mathbb{Z}_2$ to any arbitrary finite group is now straightforward. You can see how it goes in the exercise below:

Exercise 68 (Quotients under free actions of finite groups are smooth)

Let M be a smooth manifold, and Γ be a finite group of diffeomorphisms of M acting freely on M, that is:

- each $\tau \in \Gamma$ is a diffeomorphism $\tau\colon M \to M$;
- Γ is closed under compositions and taking inverses;
- and $\tau(p) \neq p$ for every $p \in M$ and $\tau \in \Gamma \smallsetminus \{\mathrm{Id}_M\}$.

Let M/Γ denote the quotient of M under the equivalence relation $\sim$ defined as: $p \sim q$ if there is $\tau \in \Gamma$ such that $q = \tau(p)$. As above, here we will show that M/Γ is a topological manifold and that it has a unique smooth structure for which the quotient projection $\pi\colon M \to M/\Gamma$ is a local diffeomorphism.

(a) Equip M/Γ with the quotient topology and show that $\pi\colon M \to M/\Gamma$ is an open mapping, i.e., takes open sets to open sets.
Hint: if $U \subseteq M$ is open, write $\pi^{-1}(\pi(U))$ as an union of open sets.

(b) Use (a) to show that M/Γ is Hausdorff and second-countable.
Hint: if $\pi(p) \neq \pi(q)$, then for every $\tau \in \Gamma$ there are disjoint open neighborhoods $U_\tau, V_\tau \subseteq M$ of p and $\tau(q)$; let $U = \bigcap_{\tau\in\Gamma} U_\tau$ and $V = \bigcap_{\tau\in\Gamma} \tau^{-1}(V_\tau)$. What can you say about $\pi(U)$ and $\pi(V)$?

An open subset $U \subseteq M$ is called **Γ-small** if $U \cap \tau(U) = \varnothing$ for each $\tau \in \Gamma \setminus \{\mathrm{Id}_M\}$.

(c) Show that every point in M has a Γ-small open neighborhood, and use this to conclude that M/Γ is locally Euclidean.
Hint: given any $p \in M$, for every $\tau \in \Gamma \setminus \{\mathrm{Id}_M\}$ there are disjoint open neighborhoods $V_\tau, W_\tau \subseteq M$ of p and $\tau(p)$ (as M is Hausdorff); then consider the intersection $U = \bigcap_{\tau\in\Gamma\setminus\{\mathrm{Id}_M\}} \big(V_\tau \cap \tau^{-1}(W_\tau)\big)$.

Hence, M/Γ is a topological manifold. Now, show that:

(d) The atlas $\mathfrak{A}_\Gamma$ constructed in (d) is a C^∞-atlas (and hence induces a smooth structure $\mathfrak{A}'_\Gamma$), and that $\pi\colon M \to M/\Gamma$ is a local diffeomorphism.

(e) If $\mathfrak{A}'$ is another smooth structure on M/Γ with the property that π is a local diffeomorphism, then the identity mapping $(M/\Gamma, \mathfrak{A}'_\Gamma) \to (M/\Gamma, \mathfrak{A}')$ is a diffeomorphism. (And hence $\mathfrak{A}' = \mathfrak{A}'_\Gamma$.)

Exercise 69 (Lens spaces)

Here, we use the result of the previous exercise to explore a concrete situation.

Consider the sphere $\mathbb{S}^3 = \{(z,w) \in \mathbb{C}^2 : |z|^2 + |w|^2 = 1\}$, and let $p, q > 1$ be coprime integers. Define $\tau\colon \mathbb{S}^3 \to \mathbb{S}^3$ by $\tau(z,w) = (\mathrm{e}^{2\pi \mathrm{i}/p} z, \mathrm{e}^{2\pi \mathrm{i} q/p} w)$, and consider $\Gamma = \{\mathrm{Id}_{\mathbb{S}^3}, \tau, \tau^2, \ldots, \tau^{p-1}\}$. Show that τ is a diffeomorphism having $\tau^p = \mathrm{Id}_{\mathbb{S}^3}$, and that the action of $\Gamma \cong \mathbb{Z}_p$ on $\mathbb{S}^3$ is free. Conclude that the quotient space $L(p,q) \doteq \mathbb{S}^3/\mathbb{Z}_p$ is a smooth manifold.

The quotients $L(p,q)$ are called **Lens spaces**. A deep result in geometric topology (the **Seifert-Threlfall theorem**) says that $L(p,q_1)$ is homeomorphic to $L(p,q_2)$ if and only if $q_1 \equiv \pm q_2^{\pm 1} \pmod{p}$. (Recall **Moise's Theorem** [18]: two smooth 3-manifolds are homeomorphic if and only if they are diffeomorphic.)

Of course, all of this can be further generalized for Lie groups instead of finite groups. The difference here is that the action of a Lie group G on the smooth manifold M, in addition to be free, must also be **proper**, that is, the enhanced action-mapping $G \times M \ni (g,x) \mapsto (x, g\cdot x) \in M \times M$ is proper in the general-topology sense: inverse images of compact sets are compact. In this case, the quotient M/G is a topological manifold, and it has a unique smooth structure for which $\pi\colon M \to M/G$ is a surjective submersion (we will revisit immersions and submersions later). The reason we have a submersion as opposed to a local diffeomorphism is that when $\dim G \geq 1$, the dimension of the quotient M/G takes a hit: $\dim(M/G) = \dim M - \dim G$ (every finite group is a discrete 0-dimensional Lie group, explaining why $\dim M = \dim M/\Gamma$ above). See [15, Theorem 21.10] for a proof and more details.

There is one last result—**Godement's Criterion**—which goes beyond quotients of smooth manifolds under "nice" group actions. It considers the situation where M is a smooth manifold and $\sim$ is *any* equivalence relation on M, and gives us necessary and sufficient conditions for the existence of a smooth structure on the quotient $M/_\sim$ making the projection $\pi\colon M \to M/_\sim$ a surjective submersion; see [9, Lemma 3.7.9 and Theorem 3.7.10] for the details.

4.3 Tangent vectors and tangent spaces

Curves have tangent lines, and surfaces have tangent planes. Manifolds, being higher-dimensional generalizations of curves and surfaces, should have tangent *spaces*. Understanding how they work is the goal of this section.

Elements of Euclidean space $\mathbb{R}^n$ are usually called "points" or "vectors", interchangeably, with not enough emphasis given to the difference: "points" are static, but "vectors" can be moved around. Now, the difference is crucial. We would like to think that a tangent vector to $\mathbb{R}^n$ at a point p is a vector "starting at p", cf. Figure 44.

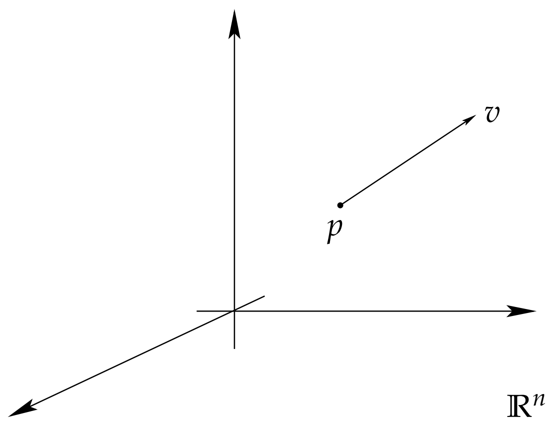


Figure 44: The vector v "touches" $\mathbb{R}^n$ only at the point p, hence "tangent".

If $T_p(\mathbb{R}^n)$ denotes the tangent space to $\mathbb{R}^n$ at p, however we define it, at the end of the day we should have that $T_p(\mathbb{R}^n) \cong \mathbb{R}^n$. But how to make sense of this for an arbitrary smooth manifold M? Vectors are no longer necessarily n-tuples of real numbers, since M is abstract, and not necessarily given as a subset of some $\mathbb{R}^N$.

Remark. Every smooth manifold can be embedded into some $\mathbb{R}^N$. This is called the **Whitney Embedding Theorem**. We will see the definition of embedding later, but the point is that Whitney tells us that there would be no loss of generality in defining smooth manifolds as already being inside some $\mathbb{R}^N$ by default. While it is true that in this case, tangent spaces to M could be directly defined as certain vector subspaces of $\mathbb{R}^N$, the main issue with this approach is that relying too much on this background $\mathbb{R}^N$, like a crutch, makes other subtler points of the theory more obscure.

Back to tangent vectors. A first natural idea is to use charts; see Figure 45:

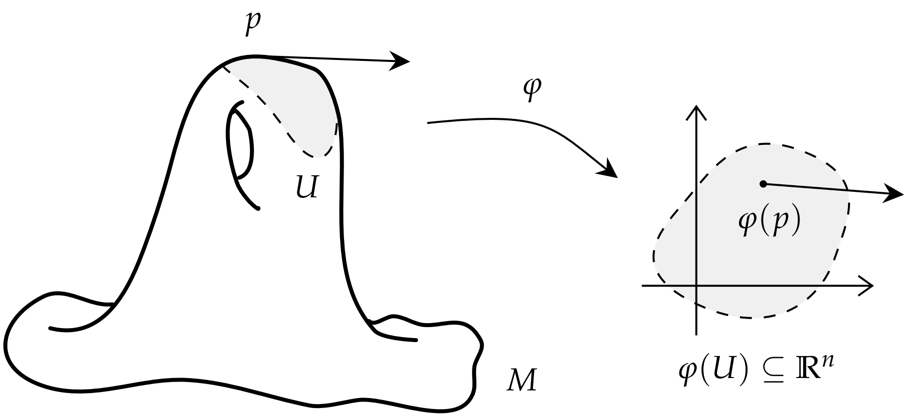


Figure 45: The first attempt to define a tangent vector to a manifold via charts...

One could say that a tangent vector to M at p is the same thing as a tangent vector to $\mathbb{R}^n$ at $\varphi(p)$. The issue with it is that if (V, ψ) is another chart around p, the "corresponding" tangent vector to $\mathbb{R}^n$ at $\psi(p)$ is, in general, different from the one at $\varphi(p)$:

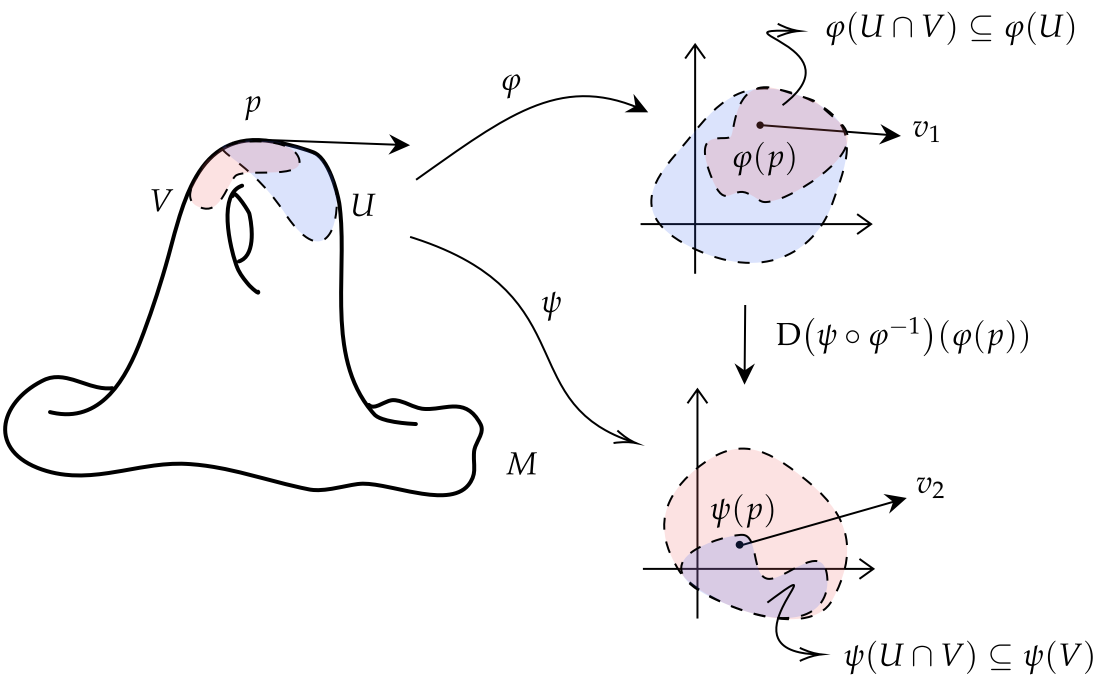


Figure 46: ... and the potential issue with it.

Namely, if "$v_1 \in T_{\varphi(p)}(\mathbb{R}^n)$" and "$v_2 \in T_{\psi(p)}(\mathbb{R}^n)$" correspond to the "same" tangent vector we are trying to define, they should be necessarily related via the transition between the charts, as $\mathrm{D}(\psi \circ \varphi^{-1})(\varphi(p))v_1 = v_2$. This suggests that the definition of tangent space should somehow involve an equivalence relation, and a quotient (identifying $v_1 \sim v_2$). This is in fact an approach often used by physicists, but it relies too heavily on charts. We will instead develop things in a coordinate-free manner, but you can explore the details of this physics-based approach on Exercise 73 ahead.

The result below starts paving our way forward:

Lemma 6 ($\partial/\partial v$ determines v)

Let $p \in \mathbb{R}^n$ and $v, w \in \mathbb{R}^n$, to be thought of as tangent to $\mathbb{R}^n$ at p. If we have that $(\partial f/\partial v)(p) = (\partial f/\partial w)(p)$ for every smooth real-valued function f defined on some neighborhood of p, then $v = w$. In other words, *vectors are characterized by how they act as directional derivatives on smooth functions defined near p.*

Proof: Writing $v = (v_1, \ldots, v_n)$ and $w = (w_1, \ldots, w_n)$, it suffices to take f to be the i-th coordinate function, $(x_1, \ldots, x_n) \mapsto x_i$. Then

$$\frac{\partial f}{\partial v}(p) = v_i \quad \text{and} \quad \frac{\partial f}{\partial w}(p) = w_i \implies v_i = w_i \text{ for all } i = 1, \ldots, n.$$

Hence, $v = w$. □

As the above proof shows, assuming that the equality between directional derivatives of p held for *all* smooth functions near p was overkill: we only really needed the coordinate projections. On manifolds, we have coordinate systems. With the goal of adopting a coordinate-free approach, and at the same time inspired by the above, we turn our attention to the $\mathbb{R}$-algebra (see Definition 34 below) $C^\infty(M)$ of smooth functions on a smooth manifold M. In particular, $f, g \in C^\infty(M)$ implies that $f + g, fg \in C^\infty(M)$. The vector space structure of tangent spaces will ultimately come from $C^\infty(M)$.

Definition 33 (Germs of smooth functions)

Let M be a smooth manifold, and fix a point $p \in M$. On the set consisting of all the pairs (U, f) where $U \subseteq M$ is an open neighborhood of p and $f\colon U \to \mathbb{R}$ is smooth, we define a relation $\sim$ by declaring that $(U, f) \sim (V, g)$ is there is an open neighborhood $W \subseteq U \cap V$ of p such that $f|_W = g|_W$. Then, $\sim$ is an equivalence relation, an equivalence class $[(U, f)]$ is called a **smooth germ at** p, and the quotient set $\mathscr{G}_p^\infty(M)$ is called the **algebra of smooth germs at** p.

Exercise 70

Convince yourself that $\sim$ in the above definition is indeed an equivalence relation.

When U is a subset of $\mathbb{R}^n$, the germ $[(U, f)]$ clearly contains information about all partial derivatives of all orders of f at p, but in fact it has more than that: two functions having the same Taylor series at p do not necessarily have the same germ at p. Consider in the real line the function f given by $f(x) = \mathrm{e}^{-1/x^2}$ if $x > 0$, and $f(x) = 0$ if $x \leq 0$, from Exercise 46 (p. 57). All derivatives of f at $p = 0$ vanish, but $[(\mathbb{R}, f)]$ is not the zero germ (that is, it does not have the same germ as the zero function) since f does not identically vanish on any interval around the origin.

In any case, we called $\mathscr{G}_p^\infty(M)$ an algebra, so we should make its operations clear and verify that they make sense:

$$\begin{aligned} &\text{(i)}\ \ [(U,f)] + [(V,g)] = [(U\cap V, f|_{U\cap V} + g|_{U\cap V})], \\ &\text{(ii)}\ \ [(U,f)]\cdot[(V,g)] = [(U\cap V, f|_{U\cap V}\, g|_{U\cap V})], \\ &\text{(iii)}\ \ \lambda\cdot[(U,f)] = [(U,\lambda f)], \end{aligned} \tag{4.2}$$

for all $[(U,f)], [(V,g)] \in \mathscr{G}_p^\infty(M)$ and $\lambda \in \mathbb{R}$. Let us show that addition in $\mathscr{G}_p^\infty(M)$ is well-defined, i.e., that if we have $(U,f) \sim (U',f')$ and $(V,g) \sim (V',g')$, then necessarily $(U\cap V, f|_{U\cap V} + g|_{U\cap V}) \sim (U'\cap V', f'|_{U\cap V} + g'|_{U\cap V})$. Fix open neighborhoods $U'' \subseteq U \cap U''$ and $V'' \subseteq V\cap V'$ of p on which $f|_{U''} = f'|_{U''}$ and $g|_{V''} = g'|_{V''}$, so that $U''\cap V'' \subseteq (U\cap V)\cap(U'\cap V')$ is also an open neighborhood of p, and

$$\begin{aligned} (f|_{U\cap V} + g|_{U\cap V})|_{U''\cap V''} &= (f|_{U\cap V})|_{U''\cap V''} + (g|_{U\cap V})|_{U''\cap V''} \\ &= f|_{U''\cap V''} + g|_{U''\cap V''} \\ &= f'|_{U''\cap V''} + g'|_{U''\cap V''} \\ &= (f'|_{U'\cap V'})|_{U''\cap V''} + (g'|_{U'\cap V'})|_{U''\cap V''} \\ &= (f'|_{U'\cap V'} + g'|_{U'\cap V'})|_{U''\cap V''}, \end{aligned} \tag{4.3}$$

showing that the definition of $[(U,f)] + [(V,g)]$ does not depend on the choice of representatives for $[(U,f)]$ and $[(V,g)]$.

Exercise 71

Show, similarly to what was done in (4.3), that scalar-multiplication and the product of germs in $\mathscr{G}_p^\infty(M)$ are also well-defined.

As a direct consequence of the pointwise definitions of the operations in $C^\infty(M)$, it follows that $\mathscr{G}_p^\infty(M)$ is an $\mathbb{R}$-algebra. Now, while germs are not functions, they still can be "evaluated" at the point p: we have a homomorphism $\delta_p\colon \mathscr{G}_p^\infty(M) \to \mathbb{R}$ of $\mathbb{R}$-algebras, defined by $\delta_p([(U,f)]) = f(p)$ (it is obviously well-defined).

With this in place, we are ready to outsource the rest of the construction to abstract algebra. Here are some precise definitions:

Definition 34 ($\mathbb{K}$-algebras and derivations)

Let $\mathbb{K}$ be any field.

- A **$\mathbb{K}$-algebra** is a vector space A over the field $\mathbb{K}$, equipped with a $\mathbb{K}$-bilinear operation $\cdot\colon A\times A \to A$.
- A **homomorphism of $\mathbb{K}$-algebras** is a linear transformation $\delta\colon A_1 \to A_2$

with the additional property $\delta(ab) = \delta(a)\delta(b)$, for all $a, b \in A_1$.

- If A is a $\mathbb{K}$-algebra and $\delta\colon A \to A$ is a homomorphism, a **δ-derivation of** A is a linear functional $v \in A^*$ such that $v(ab) = v(a)\delta(b) + \delta(a)v(b)$, for all $a, b \in A$. We set $\mathrm{Der}(A, \delta) = \{v \in A^* : v \text{ is a } \delta\text{-derivation of } A\}$.

The space $\mathrm{Der}(A, \delta)$ of derivations is always a vector space over $\mathbb{K}$, simply because it turns out to be a vector subspace of the dual space A^*. Without further ado:

Definition 35 (Tangent space)

Let M be a smooth manifold, and $p \in M$. The **tangent space to** M **at** p is defined to be $T_pM = \mathrm{Der}(\mathscr{C}^\infty_p(M), \delta_p)$, i.e., it is the space of all $v\colon \mathscr{C}^\infty_p(M) \to \mathbb{R}$ such that

$$v([f] + [g]) = v[f] + v[g] \quad \text{and} \quad v([f][g]) = g(p)v[f] + f(p)v[g], \tag{4.4}$$

for all $[f], [g] \in \mathscr{C}^\infty_p(M)$. Elements of T_pM are then called **tangent vectors**.

Remark. Note here the first instance of a common abuse of notation: we denote a germ $[(U, f)]$ simply by $[f]$. It is justified since whenever $U' \subseteq U$ is an open neighborhood of p, we have the equality $[(U', f|_{U'})] = [(U, f)]$. This also suggests the following: if $U \subseteq M$ is an open subset and $p \in U$, there is a **natural isomorphism** $T_pU \cong T_pM$ (that is, an isomorphism which does not depend on a choice of basis). The reason is that if $f\colon U \to \mathbb{R}$ is a smooth function, the value $v[f] \in \mathbb{R}$ depends only on the values of f on *some* open neighborhood of p, not on all of U. Formally, the isomorphism is the derivative of the inclusion mapping $\iota\colon U \hookrightarrow M$, but we will discuss derivatives of smooth mappings in the next section only.

Note that tangent vectors also act on actual smooth functions defined near p, by first composing the function with the projection onto its germ. More precisely, whenever $U \subseteq M$ is an open neighborhood of p and $f\colon U \to \mathbb{R}$ is a smooth function, we may set $v(f) = v[f]$, where $[f]$ denotes the germ of f.

By the discussed above, each tangent space T_pM is automatically a real vector space. The natural guess is that their dimension, as vector spaces, equals the dimension of M as a manifold.

In the next proof, we (partially) start using **Einstein's summation convention**: it states that if the same index appears in a monomial expression once above and once below, then it is being summed over. The range of summation is then implied, and the summation sign is omitted. Here are some examples:

- If $e_1, \ldots, e_n$ is a basis for a vector space V, and $v \in V$ is expressed as a linear combination of this basis as $v = \sum_{i=1}^n a_i e_i$, we would write it simply as $v = a^i e_i$.
- Linear functionals, getting paired with vectors, should be indexed with upper indices, so that the dual basis would be $\varphi^1, \ldots, \varphi^n$, and $\xi \in V^*$ would be expressed as a linear combination of this dual basis as $\xi = \xi_i \varphi^i$.

- If $\tilde{e}_1, \ldots, \tilde{e}_m$ is a basis for a second vector space $\tilde{V}$ and $T\colon V \to \tilde{V}$ is a linear transformation, the matrix of T relative to the given bases is defined via the relations $Te_j = \sum_{i=1}^m a_{ij}\tilde{e}_i$, for $j = 1, \ldots, n$. Having the index i appear twice below on the right side would violate the summation convention, and so we would rewrite the last expression as $Te_j = a^i_j \tilde{e}_i$.

The real power of Einstein's convention is not avoiding the summation signs, but instead keeping track of the correct "index balance". It gives us a built-in error detector when doing index computations. I lost count of how many times I have seen people (and even textbooks) write things like $u^i v^i$ meaning the sum $\sum_{i=1}^n u^i v^i$, and tried to explain these subtleties to no avail. To try and make my point, we will keep writing summation signs, but will follow the correct index balance in abstract index computations. Doing so will require the following change in notation, which might be upsetting at a first moment: if (U, φ) is a chart for a manifold M, we will write $\varphi(p) = (x^1(p), \ldots, x^n(p))$ instead of $(x_1(p), \ldots, x_n(p))$. So, for instance, x^2 would mean the second component function of a chart, and not "x squared"; then $(x^2)^2$ would mean the second component squared, and not the fourth power of x, etc. No serious confusion will arise from this.

Theorem 10

Let M be a smooth manifold, and $p \in M$ be any point. Then, $\dim T_pM = \dim M$.

Proof: We will do this by exhibiting a basis of T_pM containing n elements. Namely, consider a chart (U, φ) for M around p, and write its components as $\varphi = (x^1, \ldots, x^n)$. That is, denoting the Euclidean coordinate functions by $u^i\colon \mathbb{R}^n \to \mathbb{R}$, we have that $x^i = u^i \circ \varphi\colon U \to \mathbb{R}$. We then define

$$\left.\frac{\partial}{\partial x^i}\right|_p \in T_pM \quad \text{by} \quad \left.\frac{\partial}{\partial x^i}\right|_p [f] = \frac{\partial(f \circ \varphi^{-1})}{\partial u^i}(\varphi(p)), \tag{4.5}$$

where f is a representative of the germ $[f]$, defined on some open neighborhood of p contained in U. The second relation in (4.4) for $\partial/\partial x^i|_p$ is nothing more than the product rule for the Euclidean partial derivatives $\partial/\partial u^i$. Note that

$$\left.\frac{\partial}{\partial x^i}\right|_p [x^j] = \frac{\partial u^j}{\partial u^i}(\varphi(p)) = \delta^j_i = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases} \tag{4.6}$$

A consequence of (4.6) is that

$$\left\{ \left.\frac{\partial}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p \right\} \text{ is linearly independent.} \tag{4.7}$$

If $a^1, \ldots, a^n \in \mathbb{R}$ are such that $\sum_{i=1}^n a^i \partial/\partial x^i|_p = 0$, evaluating both sides as $[x^j]$ leads to $\sum_{i=1}^n a^i \delta^j_i = 0$, that is, $a^j = 0$. And finally, we claim that

$$v = \sum_{i=1}^n v[x^i] \left.\frac{\partial}{\partial x^i}\right|_p. \tag{4.8}$$

To establish (4.8), we must evaluate both sides at an arbitrary germ $[f]$ and check that they produce the same output. First note that $v(1) = 0$, by writing $1 \cdot 1 = 1$ and applying the product rule. Hence $v(c) = 0$ for any $c \in \mathbb{R}$. Now we apply Hadamard's lemma (p. 58) to $f \circ \varphi^{-1}$ (reducing U if needed to make it starshaped around $\varphi(p)$) to write

$$f = f(p) + \sum_{i=1}^{n} (x^i - x^i(p)) g_i, \tag{4.9}$$

where g_i are smooth functions defined on some open neighborhood of p satisfying $g_i(p) = (\partial(f \circ \varphi^{-1})/\partial u^i)(\varphi(p))$. Now:

$$\begin{aligned} v[f] &= v\left[f(p) + \sum_{i=1}^{n} (x^i - x^i(p)) g_i \right] = v[f(p)] + \sum_{i=1}^{n} v[(x^i - x^i(p)) g_i] \\ &= 0 + \sum_{i=1}^{n} \Big((x^i(p) - x^i(p)) g_i(p) + v[(x^i - x^i(p)] g_i(p) \Big) = \sum_{i=1}^{n} v[x^i] g_i(p) \\ &= \sum_{i=1}^{n} v[x^i] \frac{\partial (f \circ \varphi^{-1})}{\partial u^i}(\varphi(p)) = \sum_{i=1}^{n} v[x^i] \frac{\partial}{\partial x^i}\Big|_p [f] \\ &= \left(\sum_{i=1}^{n} v[x^i] \frac{\partial}{\partial x^i}\Big|_p \right) [f], \end{aligned} \tag{4.10}$$

as required. $\square$

The proof of Theorem 10 is more important than its statement, which is intuitive and trivial to remember. It contains:

- (4.5), which is how we locally define partial derivatives induced by a chart, as derivations. In other words, it only makes sense to take partial derivatives of a function defined on a manifold once a chart has been fixed, and the result in general depends on the choice of chart;
- and (4.8), which actually tells us *how* to write a tangent vector as a linear combination of a coordinate basis.

Exercise 72 (Transitions between coordinate bases)

Let M be a smooth manifold, $p \in M$ be any point, and let (U, φ) and (V, ψ) be two charts for M around p. Writing them as $\varphi = (x^1, \ldots, x^n)$ and $\psi = (y^1, \ldots, y^n)$, we obtain two bases

$$\left\{ \frac{\partial}{\partial x^1}\Big|_p, \ldots, \frac{\partial}{\partial x^n}\Big|_p \right\} \quad \text{and} \quad \left\{ \frac{\partial}{\partial y^1}\Big|_p, \ldots, \frac{\partial}{\partial y^n}\Big|_p \right\}$$

of T_pM.

(a) Writing $\varphi \circ \psi^{-1} \colon \psi(U \cap V) \to \varphi(U \cap V)$ as $(\varphi \circ \psi^{-1})(y^1, \ldots, y^n) = (x^1, \ldots, x^n)$, show that

$$\left.\frac{\partial}{\partial y^j}\right|_p = \sum_{i=1}^n \frac{\partial x^i}{\partial y^j}(\psi(p))\left.\frac{\partial}{\partial x^i}\right|_p$$

for all $j = 1, \ldots, n$. Here, $(\partial x^i/\partial y^j)(\psi(p))$ denotes the Euclidean partial derivative of the i-th component x^i of $\varphi \circ \psi^{-1}$ relative to the j-th variable y^j, evaluated at the point $\psi(p)$.

(b) In particular, when $M = \mathbb{R}^2$ and $U = \mathbb{R}^2 \smallsetminus ([0,\infty) \times \{0\})$, we have polar coordinates $(r,\theta)\colon U \to (0,\infty) \times (0,2\pi)$, characterized by $x = r\cos\theta$ and $y = r\sin\theta$. Write $\partial/\partial r$ and $\partial/\partial\theta$ as linear combinations of $\partial/\partial x$ and $\partial/\partial y$, at any point in U.

Our discussion, which started with the idea of defining tangent vectors via charts, now comes full circle:

Exercise 73 (How a physicist usually thinks about T_pM)

Let M be a smooth manifold, and fix $p \in M$. If $\mathfrak{A}$ denotes the smooth structure of M, let $\mathfrak{A}_p = \{(U,\varphi) \in \mathfrak{A} : p \in U\}$. On the set $\mathfrak{A}_p \times \mathbb{R}^n$, we define a relation $\sim$ by saying that $((U,\varphi),\boldsymbol{v}) \sim ((V,\psi),\boldsymbol{w})$ if and only if $\mathrm{D}(\psi\circ\varphi^{-1})(\varphi(p))\boldsymbol{v} = \boldsymbol{w}$.

(a) Show that $\sim$ is an equivalence relation.

Then let $(T_pM)_{\text{PHYS.}}$ denote the quotient $(\mathfrak{A}_p \times \mathbb{R}^n)/_\sim$. This time, the vector space structure is less obvious. Using brackets for equivalence classes, we define

$$[((U,\varphi),\boldsymbol{v})] + [((V,\psi),\boldsymbol{w})] = [((U,\varphi),\boldsymbol{v} + \mathrm{D}(\varphi\circ\psi^{-1})(\psi(p))\boldsymbol{w})],$$

and $\lambda[((U,\varphi),\boldsymbol{v})] = [((U,\varphi),\lambda\boldsymbol{v})]$.

(b) Show that the proposed addition and scalar multiplication in $(T_pM)_{\text{PHYS.}}$ are well-defined.

With these operations, $(T_pM)_{\text{PHYS.}}$ becomes a vector space.

(c) Show that $\Phi\colon (T_pM)_{\text{PHYS.}} \to T_pM$ given by

$$\Phi[((U,\varphi),(v^1,\ldots,v^n))] = \sum_{i=1}^n v^i \left.\frac{\partial}{\partial x^i}\right|_p$$

is a well-defined vector space isomorphism.

Note: Physicists like to think about the relation $\mathrm{D}(\psi\circ\varphi^{-1})(\varphi(p))\boldsymbol{v} = \boldsymbol{w}$ as a *transformation law*. Namely, for every chart $(U,\varphi) \in \mathfrak{A}_p$ they assign a collection $\{v^i\}_{i=1}^n$ of numbers, and they are subject to the following rule: whenever

$(U, \varphi), (V, \psi) \in \mathfrak{A}_p$, the associated $\{v^i\}_{i=1}^n$ and $\{w^i\}_{i=1}^n$ are related through

$$v^i = \sum_{j=1}^n \frac{\partial x^i}{\partial y^j} w^j,$$

where $\varphi = (x^1, \ldots, x^n)$ and similarly for ψ. Tangent vectors to M are treated as elements of $\mathbb{R}^n$, and then every time they want to do something, they have to check that the results are chart-independent by using this transformation law — otherwise their constructions and calculations don't hold at the manifold level. This gets old really fast, which is why we won't pursue it when building up the theory in this class. It's still good to understand their philosophy so you can communicate with them anyway.

We are now in position to justify, very generally, why $T_p(\mathbb{R}^n) \cong \mathbb{R}^n$ in a natural manner:

Exercise 74 (Tangent spaces to vector spaces)

Let V be a finite-dimensional real vector space, equipped with its standard Euclidean smooth structure. Show that for any point $p \in V$, the linear mapping $\Phi_p \colon V \to T_p V$ given by

$$\Phi_p(v)[f] = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(p + tv)$$

is an isomorphism. When $V = \mathbb{R}^n$ and e_i is the i-th vector in the standard basis, what is the derivation $\Phi_p(e_i)$? Does the answer surprise you?

We will see some concrete examples of tangent spaces once we have the language of differentials and total derivatives available on manifolds.

4.4 Derivatives of smooth functions between manifolds

With the definition of smoothness in place, We are now ready to define derivatives (also known as differentials) of smooth functions between manifolds.

Definition 36 (Total derivatives redux)

Let M and N be smooth manifolds, and $p \in M$ be any point.

- If $f \colon M \to \mathbb{R}$ is a smooth function, the **derivative of** f **at** p is the linear functional $\mathrm{d}f_p \colon T_pM \to \mathbb{R}$ given by $\mathrm{d}f_p(v) = v[f]$, where $[f] \in \mathscr{G}_p^\infty(M)$ is the germ of f at p.
- If $F \colon M \to N$ is a smooth mapping, the **derivative of** F **at** p is the linear transformation $\mathrm{d}F_p \colon T_pM \to T_{F(p)}N$ given by $\mathrm{d}F_p(v)[g] = v[g \circ F]$, where $[g] \in \mathscr{G}_{F(p)}^\infty(N)$ and $[g \circ F]$ is the germ of $g \circ F$ at p, where g is any represen-

tative of $[g]$; see Figure 47.

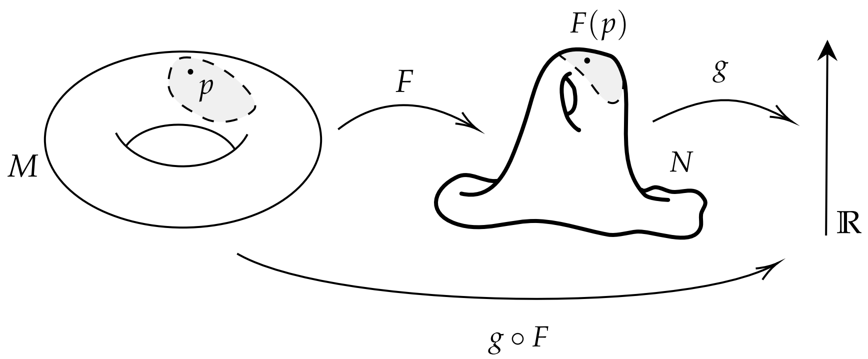


Figure 47: The definition of derivative via germs of smooth functions.

Remark. In the second case, the value $\mathrm{d}F_p(v)$ must be an element of $T_{F(p)}N$, which is a derivation of $G^\infty_{F(p)}(N)$, and this is why we define it by declaring what is the value of $\mathrm{d}F_p(v)[g]$ for each $[g] \in \mathscr{G}^\infty_{F(p)}(N)$. When $N = \mathbb{R}$ has its standard smooth structure and we use the global identity chart to obtain $\partial/\partial t|_t \in T_t\mathbb{R}$, and then identify $T_t\mathbb{R} \cong \mathbb{R}$, both definitions of derivative given above agree.

Example 69 (Differentials in coordinates)

Let M be a smooth manifold, $f\colon M \to \mathbb{R}$ be a smooth function, and $(U; x^1, \ldots, x^n)$ be a chart for M. We will show that

$$\mathrm{d}f_p = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(p)\, \mathrm{d}x^i|_p, \quad \text{for every } p \in U, \tag{4.11}$$

where $(\partial f/\partial x^i)(p) \doteq (\partial/\partial x^i)|_p[f]$ and $\mathrm{d}x^i|_p$ is the derivative at p of the coordinate function $x^i\colon U \to \mathbb{R}$. Indeed, if $v \in T_pM$ is arbitrary, we have that $\mathrm{d}x^i|_p(v) = v[x^i]$ for each $i = 1, \ldots, n$ (by definition of $\mathrm{d}x^i|_p$), and so (4.8) leads to

$$\begin{aligned}
\mathrm{d}f_p(v) = \mathrm{d}f_p\left(\sum_{i=1}^n v[x^i]\frac{\partial}{\partial x^i}\bigg|_p\right) &= \sum_{i=1}^n v[x^i]\,\mathrm{d}f_p\left(\frac{\partial}{\partial x^i}\bigg|_p\right) \\
&= \sum_{i=1}^n \mathrm{d}x^i|_p(v)\,\frac{\partial}{\partial x^i}\bigg|_p[f] = \sum_{i=1}^n \mathrm{d}x^i|_p(v)\,\frac{\partial f}{\partial x^i}(p) \\
&= \left(\sum_{i=1}^n \frac{\partial f}{\partial x^i}(p)\,\mathrm{d}x^i|_p\right)(v),
\end{aligned} \tag{4.12}$$

as required.

Exercise 75 (Cotangent spaces)

Let M be a smooth manifold, and $p \in M$ be any point. The **cotangent space** to M at p is defined to be $T_p^*M \doteq (T_pM)^*$, the dual space to the tangent space at p. In other words, $T_p^*M = \{\xi\colon T_pM \to \mathbb{R} : \xi \text{ is linear}\}$. Let (U, φ) be a chart for M around p, and consider its components $\varphi = (x^1, \ldots, x^n)$.

(a) Show that the differentials $\mathrm{d}x^i|_p\colon T_pM \to \mathbb{R}$ of the coordinate functions $x^i\colon U \to \mathbb{R}$ form a basis for T_p^*M.

(b) How is the basis $\{\mathrm{d}x^1|_p, \ldots, \mathrm{d}x^n|_p\}$ of T_p^*M obtained above related to the coordinate basis $\{\partial/\partial x^1|_p, \ldots, \partial/\partial x^n|_p\}$ of T_pM?

Cotangent spaces, being dual to tangent spaces, play a central role in differential geometry and differential topology. For the more algebraically-inclined reader, here's another way to think about them:

Exercise 76 (The algebraic approach to cotangent spaces)

Let M be a smooth manifold, $p \in M$, and consider the kernel of δ_p:

$$\mathcal{I}_p = \{[f] \in \mathcal{G}_p^\infty(M) : f(p) = 0\}.$$

(a) Show that $\mathcal{I}_p$ is an ideal of the algebra $\mathcal{G}_p^\infty(M)$ (that is, closed under sums and absorbs multiplications)

(b) Find a canonical isomorphism (that is, basis-independent) between the cotangent space T_p^*M and the quotient $\mathcal{I}_p/\mathcal{I}_p^2$.

Hint: Find a surjective linear transformation $\Phi\colon \mathcal{I}_p \to T_p^*M$ with $\ker\Phi = \mathcal{I}_p^2$ (Hadamard's lemma is useful to establish this last relation).

Can you identify what $\mathcal{I}_p^2/\mathcal{I}_p^3$ is? What about $\mathcal{I}_p^k/\mathcal{I}_p^{k+1}$?

In Algebraic Geometry, it is common to first define the cotangent space T_p^*M as the quotient $\mathcal{I}_p/\mathcal{I}_p^2$, and then the tangent space as the dual $T_pM = (T_p^*M)^*$.

Example 70 (Coordinate Jacobians)

Let M and N be smooth manifolds of dimensions n and m, $F\colon M \to N$ be a smooth mapping, and $(U; x^1, \ldots, x^n)$ and $(V; y^1, \ldots, y^m)$ be charts for M and N, respectively, with $U \cap F^{-1}(V) \neq \varnothing$. We will show that, for each $p \in U \cap F^{-1}(V)$, the matrix representation of $\mathrm{d}F_p\colon T_pM \to T_{F(p)}N$ relative to the associated coordinate bases

$$\left\{ \frac{\partial}{\partial x^i}\bigg|_p \right\}_{i=1}^n \quad \text{and} \quad \left\{ \frac{\partial}{\partial y^a}\bigg|_{F(p)} \right\}_{a=1}^m$$

of T_pM and $T_{F(p)}N$ is precisely the Jacobian matrix

$$\left[\frac{\partial F^a}{\partial x^i}(p)\right]_{\substack{a=1,\ldots,m\\ i=1,\ldots,n}} = \left.\begin{bmatrix} \frac{\partial F^1}{\partial x^1} & \cdots & \frac{\partial F^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F^m}{\partial x^1} & \cdots & \frac{\partial F^m}{\partial x^n} \end{bmatrix}\right|_p \in \mathbb{R}^{m\times n}, \tag{4.13}$$

where $F^a = y^a \circ F\colon U\cap F^{-1}(V) \to \mathbb{R}$, and $(\partial F^a/\partial x^i)(p) = (\partial/\partial x^i)|_p[F^a]$ as before.

To find such matrix representation, we must compute $\mathrm{d}F_p(\partial/\partial x^i|_p)$, write the result as a linear combination of the $\partial/\partial y^a|_{F(p)}$, and place the coefficients as the columns of a matrix. By (4.8), we have that

$$\mathrm{d}F_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) = \sum_{a=1}^m \mathrm{d}F_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right)[y^a]\left.\frac{\partial}{\partial y^a}\right|_{F(p)},$$

while at the same time

$$\mathrm{d}F_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right)[y^a] = \left.\frac{\partial}{\partial x^i}\right|_p[y^a\circ F] = \left.\frac{\partial}{\partial x^i}\right|_p[F^a] = \frac{\partial F^a}{\partial x^i}(p),$$

as required.

Examples 69 and 70 tell us how to compute derivatives using charts, but it will also be convenient to have an essentially coordinate-free way to do these calculations.

Definition 37 (Abstract velocity vectors)

Let M be a smooth manifold, $p \in M$ be any point, and $\alpha\colon (-\varepsilon,\varepsilon) \to M$ be any smooth curve with $\alpha(0) = p$. The **velocity vector** $\alpha'(0) \in T_pM$ is defined as the derivation $\alpha'(0)\colon \mathscr{C}^\infty_p(M) \to \mathbb{R}$ given by

$$\alpha'(0)[f] = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\alpha(t)), \tag{4.14}$$

for every $[f] \in \mathscr{C}^\infty_p(M)$.

Remark. Note that $f\circ\alpha\colon (-\varepsilon,\varepsilon) \to \mathbb{R}$ is a real-valued function of a single real variable, so in the right side of (4.14) we have a classical derivative from single-variable Calculus. More generally, it we don't evaluate such derivative at $t = 0$, we may consider $\alpha'(t) \in T_{\alpha(t)}M$. Each $\alpha'(t)$ is indeed a derivative due to the product rule for classical derivatives.

When dealing with functions between open subsets of Euclidean spaces, we could compute total derivatives as directional derivatives, via the formula

$$\mathrm{D}f(p)v = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(p+tv).$$

The problem here is that if $p \in M$ is any point and $v \in T_pM$ is a tangent vector, it is not generally true that $p + tv \in M$. In fact, if M does not live inside any Euclidean space, then the quantity $p + tv$ is not even well-defined. The solution, of course, seems to be replacing $p + tv$ with $\alpha(t)$, once the correct initial conditions $\alpha(0) = p$ and $\alpha'(0) = v$ have been arranged for. The next result says that we can indeed do this and proceed without further issues:

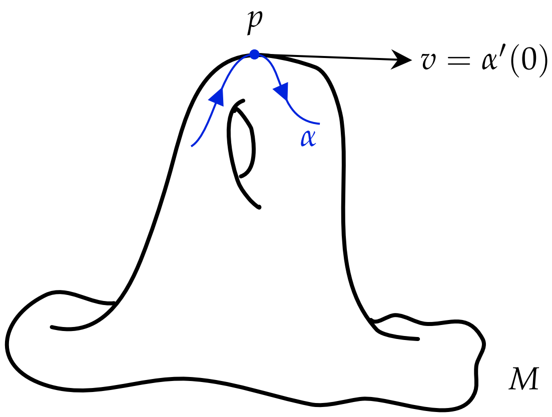


Figure 48: The curve $\alpha\colon (-\varepsilon, \varepsilon) \to M$ with the correct initial conditions.

Lemma 7 (All tangent vectors are velocity vectors)

Let M be a smooth manifold, and $p \in M$. For every $v \in T_pM$, there is a smooth curve $\alpha\colon (-\varepsilon, \varepsilon) \to M$ with $\alpha(0) = p$ and $\alpha'(0) = v$.

Remark. The curve α is far from unique.

Proof: Let $(U, \varphi = (x^1, \ldots, x^n))$ be a chart for M around p, and write the linear combination $v = \sum_{i=1}^n a^i \partial/\partial x^i|_p$, for some coefficients $a^1, \ldots, a^n \in \mathbb{R}$. As $\varphi(U) \subseteq \mathbb{R}^n$ is open, there is $\varepsilon > 0$ small enough so that the curve $\alpha\colon (-\varepsilon, \varepsilon) \to M$ given by $\alpha(t) = \varphi^{-1}(\varphi(p) + t(a^1, \ldots, a^n))$ makes sense. It is clear that $\alpha(0) = \varphi^{-1}(\varphi(p)) = p$. As for $\alpha'(0)$, note that

$$\alpha'(0)[x^i] = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} x^i(\alpha(t)) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (x^i(p) + ta^i) = a^i,$$

so that

$$\alpha'(0) = \sum_{i=1}^n \alpha'(0)[x^i] \left.\frac{\partial}{\partial x^i}\right|_p = \sum_{i=1}^n a^i \left.\frac{\partial}{\partial x^i}\right|_p = v$$

by (4.8), as required. □

Remark. The above proof also shows that, with the slight abuse of notation of identifying $x^i \circ \alpha$ with α, if $\varphi(\alpha(t)) = (x^1(t), \ldots, x^n(t))$, then

$$\alpha'(t) = \sum_{i=1}^n (x^i)'(t) \left.\frac{\partial}{\partial x^i}\right|_{\alpha(t)},$$

where each $(x^i)'(t)$ is a derivative from single-variable calculus.

This description of tangent vectors as velocity vectors also opens the way for us to understand more concrete examples of tangent spaces:

Exercise 77 (Tangent spaces to spheres)

Let $\mathbb{S}^n = \{p \in \mathbb{R}^{n+1} : \|p\| = 1\}$ be the unit n-sphere. Fix $p \in \mathbb{S}^n$. Find a canonical isomorphism between $T_p(\mathbb{S}^n)$ and $\{v \in \mathbb{R}^{n+1} : \langle v, p\rangle = 0\}$.
Hint: One option is to write $v \in T_p(\mathbb{S}^n)$ as a *derivation* $\alpha'(0)$, then map it to the *vector* $\alpha'(0) \in \mathbb{R}^{n+1}$. You will need to show that this is well-defined (i.e., doesn't depend on the choice of curve), that the *vector* $\alpha'(0)$ is indeed orthogonal to p, and that this assignment does define an isomorphism. Or consider the inclusion $\iota\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ and prove that $\mathrm{d}\iota_p\colon T_p(\mathbb{S}^n) \to T_{\iota(p)}(\mathbb{R}^{n+1}) \cong \mathbb{R}^{n+1}$ is injective and has the "correct" image.

We are ready to compute derivatives of functions and mappings on manifolds using velocity vectors. If $f\colon M \to \mathbb{R}$ is smooth, $v \in T_pM$, and a smooth curve $\alpha\colon (-\varepsilon, \varepsilon) \to M$ has $\alpha(0) = p$ and $\alpha'(0) = v$, we have that

$$\mathrm{d}f_p(v) = v[f] = \alpha'(0)[f] = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\alpha(t)). \tag{4.15}$$

If $F\colon M \to N$ is a smooth mapping, then $\mathrm{d}F_p(v) = (F \circ \alpha)'(0) \in T_{F(p)}N$ is the velocity vector at $t = 0$ of the curve $F \circ \alpha\colon (-\varepsilon, \varepsilon) \to N$, as a derivation, cf. Figure 49.

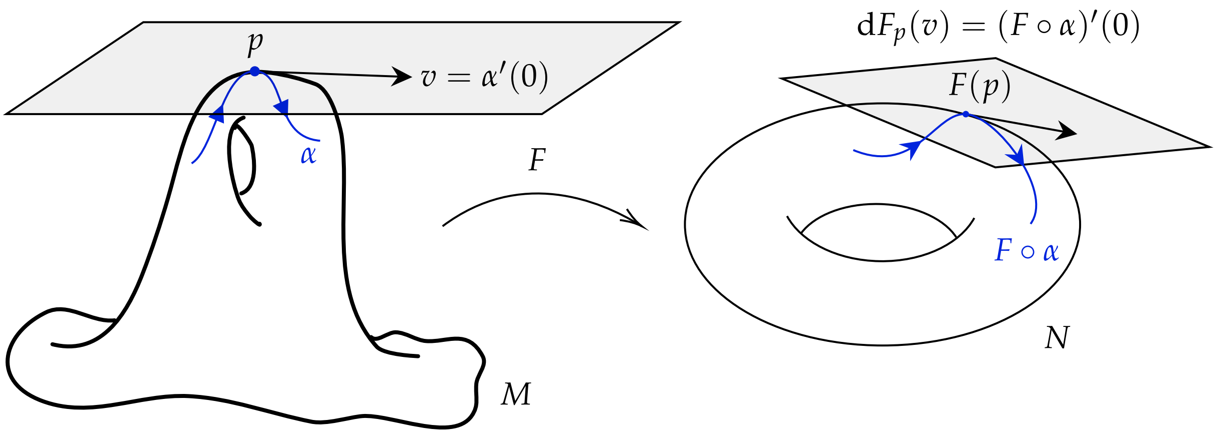


Figure 49: Computing the differential of a function between manifolds via curves.

Indeed,

$$\begin{aligned}\mathrm{d}F_p(v)[g] &= v[g \circ F] = \alpha'(0)[g \circ F] \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (g \circ F)(\alpha(t)) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} g(((F \circ \alpha)(t)) \\ &= (F \circ \alpha)'(0)[g],\end{aligned}$$

for every $[g] \in \mathscr{G}^\infty_{F(p)}(N)$, as claimed.

Proposition 29 (Chain rule redux)

Let M, N and P be smooth manifolds, and $F\colon M \to N$ and $G\colon N \to P$ be smooth mappings. Then, for every $p \in M$, we have that $\mathrm{d}(G\circ F)_p(v) = \mathrm{d}G_{F(p)}(\mathrm{d}F_p(v))$ for all $v \in T_pM$. That is, $\mathrm{d}(G\circ F)_p = \mathrm{d}G_{F(p)} \circ \mathrm{d}F_p$.

Proof: Let $v \in T_pM$ and choose a smooth curve $\alpha\colon (-\varepsilon,\varepsilon) \to M$ with $\alpha(0) = p$ and $\alpha'(0) = v$. Then, as $F\circ\alpha\colon (-\varepsilon,\varepsilon) \to N$ has $(F\circ\alpha)(0) = F(p)$ and $(F\circ\alpha)'(0) = \mathrm{d}F_p(v)$, we may compute

$$\begin{aligned}\mathrm{d}(G\circ F)_p(v) &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(G\circ F)(\alpha(t)) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} G((F\circ\alpha)(t)) \\ &= \mathrm{d}G_{F(p)}((F\circ\alpha)'(0)) = \mathrm{d}G_{F(p)}(\mathrm{d}F_p(v)),\end{aligned}$$

as wanted. □

Let's see a couple of very concrete examples:

Example 71 (The differential of the antipodal mapping)

Let $\mathbb{S}^n$ be the unit sphere, and $\tau\colon \mathbb{S}^n \to \mathbb{S}^n$ be the antipodal mapping, given by $\tau(p) = -p$. The differential is a linear transformation $\mathrm{d}\tau_p\colon T_p(\mathbb{S}^n) \to T_{-p}(\mathbb{S}^n)$, and we have seen in Exercise 77 that $T_p(\mathbb{S}^n)$ is isomorphic to the orthogonal hyperplane $p^\perp \subseteq \mathbb{R}^{n+1}$, so that we may consider $\mathrm{d}\tau_p\colon p^\perp \to p^\perp$ as a linear operator. If $v \in p^\perp$, we may choose a smooth curve $\alpha\colon (-\varepsilon,\varepsilon) \to \mathbb{S}^n$ with $\alpha(0) = p$ and $\alpha'(0) = v$, and compute

$$\mathrm{d}\tau_p(v) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\tau(\alpha(t)) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} -\alpha(t) = -\alpha'(0) = -v.$$

Even though the tangent spaces $T_p(\mathbb{S}^n)$ and $T_{-p}(\mathbb{S}^n)$, regarded as subspaces of $\mathbb{R}^{n+1}$, pass through the origin, we picture them as attached to p and $-p$, so that v and $-v$ should be drawn as starting at p and $-p$, respectively, cf. Figure 50.

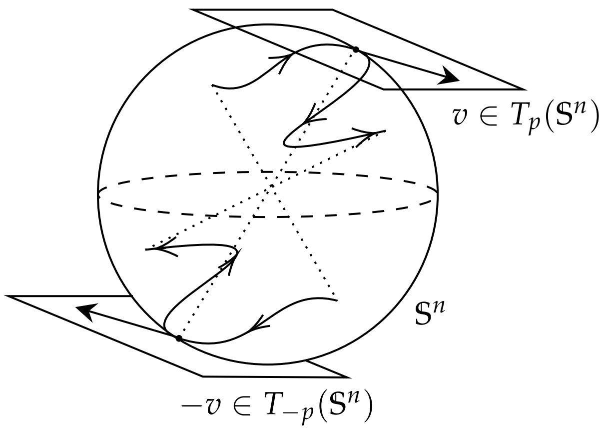


Figure 50: Visualizing the derivative of the antipodal mapping.

Example 72 (Projectivizations of linear transformations)

Let $A\colon \mathbb{R}^{n+1} \to \mathbb{R}^{k+1}$ be an injective linear transformation, and consider the induced mapping $\widetilde{A}\colon \mathbb{RP}^n \to \mathbb{RP}^k$, given by $\widetilde{A}([p]) = [A(p)]$. The injectivity assumption on A ensures that $A(p) \neq 0$ whenever $p \neq 0$, so that $[A(p)]$ makes sense. Being linear, A is homogeneous of degree 1, and so $\widetilde{A}$ is well-defined and smooth by the result of Exercise 62.

We will see ahead in Example 73 that the derivative of $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{RP}^n$ is surjective at each point, so that $T_{[p]}(\mathbb{RP}^n) \cong \mathbb{R}^{n+1}/\mathbb{R}p$ follows from Exercise 74 and the first isomorphism theorem. The problem is that this isomorphism depends on the choice of representative $p \in [p]$, so one must be careful when using it. Instead, we may argue using the chain rule:

$$\begin{array}{ccc} \mathbb{R}^{n+1} \smallsetminus \{0\} & \xrightarrow{\;A\;} & \mathbb{R}^{k+1} \smallsetminus \{0\} \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ \mathbb{RP}^n & \xrightarrow{\;\widetilde{A}\;} & \mathbb{RP}^k \end{array} \quad \Longrightarrow \quad \begin{array}{ccc} \mathbb{R}^{n+1} & \xrightarrow{\;A\;} & \mathbb{R}^{k+1} \\ \downarrow{\scriptstyle \mathrm{d}\pi_p} & & \downarrow{\scriptstyle \mathrm{d}\pi_{A(p)}} \\ T_{[p]}(\mathbb{RP}^n) & \xrightarrow{\;\mathrm{d}\widetilde{A}_{[p]}\;} & T_{[A(p)]}(\mathbb{RP}^k) \end{array}$$

In other words, if $L \in \mathbb{RP}^n$ and $w \in T_L(\mathbb{RP}^n)$ are given, one computes $\mathrm{d}\widetilde{A}_L(w)$ as follows: fix $p \in L \smallsetminus \{0\}$ (so that $[p] = L$), and select any $v \in \mathbb{R}^{n+1}$ such that $\mathrm{d}\pi_p(v) = w$; then $\mathrm{d}\widetilde{A}_L(w) = \mathrm{d}\pi_{A(p)}(A(v))$. The result is independent on the choices of p and v. Morally, the derivative of $\widetilde{A}$ is just A.

The derivatives of the cartesian projections play a central role when computing derivatives of functions on product manifolds.

Exercise 78 (Tangent spaces to product manifolds)

Let M and N be smooth manifolds, and consider the product manifold $M \times N$. Show that for all $(p,q) \in M \times N$, we have that $T_{(p,q)}(M \times N) \cong T_pM \times T_qN$.

Hint: If $\pi\colon M \times N \to M$ and $\sigma\colon M \times N \to N$ are the Cartesian projections, show that $\Phi\colon T_{(p,q)}(M \times N) \to T_pM \times T_qN$ given by $\Phi(v) = (\mathrm{d}\pi_{(p,q)}(v), \mathrm{d}\sigma_{(p,q)}(v))$ works.

Exercise 79 (Fat partial derivatives on product manifolds)

Let M, N, P be smooth manifolds, and consider a smooth mapping $F\colon M \times N \to P$. Fix a point $(p,q) \in M \times N$, and denote by $\iota_p\colon N \to M \times N$ and $\iota_q\colon M \to M \times N$ the inclusions, given by $\iota_p(y) = (p,y)$ and $\iota_q(x) = (x,q)$. With the isomorphism from Exercise 78 in place, show that the derivative $\mathrm{d}F_{(p,q)}\colon T_pM \times T_qN \to T_{F(p,q)}P$ is given by

$$\mathrm{d}F_{(p,q)}(v,w) = \mathrm{d}(F \circ \iota_q)_p(v) + \mathrm{d}(F \circ \iota_p)_q(w)$$

for all $(v, w) \in T_pM \times T_qN$.

Hint: $\pi \circ \iota_q = \mathrm{Id}_M$ and $\sigma \circ \iota_p = \mathrm{Id}_N$. You will have to differentiate $F(\alpha(t), \beta(t))$, where $\alpha'(0) = v$ and $\beta'(0) = w$, at $t = 0$, but this may be tricky. Consider then the auxiliary "surface" $G(t, s) = F(\alpha(t), \beta(s))$ of two real variables, noting that $F(\alpha(t), \beta(t)) = G(t, t)$, and use that the derivative of $G(t, t)$ at $t = 0$ is $(\partial G/\partial t)(0, 0) + (\partial G/\partial s)(0, 0)$.

Note: This is the manifold version of

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(x(t), y(t)) = x'(0)\frac{\partial f}{\partial x}(x(0), y(0)) + y'(0)\frac{\partial f}{\partial y}(x(0), y(0))$$

from baby Calculus, where the differentials $\mathrm{d}(F \circ \iota_q)_p$ and $\mathrm{d}(F \circ \iota_p)_q$ play the role of the partial derivatives, and v and w play the role of $x'(0)$ and $y'(0)$, respectively. For this reason, one usually writes

$$(\partial_1 F)_{(p,q)}(v) = \mathrm{d}(F \circ \iota_q)_p(v) \quad \text{and} \quad (\partial_2 F)_{(p,q)}(w) = \mathrm{d}(F \circ \iota_p)_q(w),$$

so that $(\partial_1 F)_{(p,q)} \colon T_pM \to T_{F(p,q)}P$, and similarly for $\partial_2 F$.

As for quotients, here is what we can establish so far:

Exercise 80

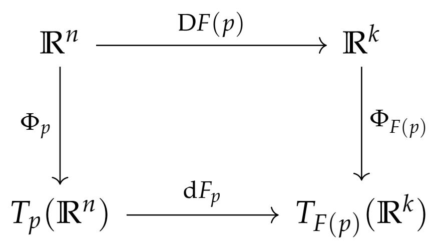
We have in Section 4.2 that if M is a smooth manifold and $\tau \colon M \to M$ is a fixed-point-free involution, the quotient M/τ of M under $p \sim \tau(p)$ admits a unique smooth structure for which $\pi \colon M \to M/\tau$ is a local diffeomorphism. Show, without the aid of coordinates, that $\mathrm{d}\pi_p \colon T_pM \to T_{\pi(p)}(M/\tau)$ is an isomorphism.

Hint: Show that if $v \in \ker \mathrm{d}\pi_p$, then $v[f] = 0$ for every germ $[f] \in \mathscr{G}_p^\infty(M)$. Explain why any $[f] \in \mathscr{G}_p^\infty(M)$ has a representative defined on a τ-small neighborhood of p.

The last question to answer here is how $\mathrm{d}F_p$ relates to the total derivative $\mathrm{D}F(p)$ from Section 2, in the case where M is an open subset of some Euclidean space.

Exercise 81

Let $U \subseteq \mathbb{R}^n$ be open, and $F \colon U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ be a smooth mapping (here, "smooth" and "smooth in the Euclidean sense" are the same thing). Show that for each point $p \in U$, the diagram



commutes, where Φ_p and $\Phi_{F(p)}$ are the isomorphisms introduced in Exercise 74. This explains why we have different notations for $DF(p)$ and dF_p: they are technically different things, but ultimately the "same". Therefore, dF_p generalizes $DF(p)$ to manifolds.

4.5 Immersions, submersions, critical points

Theorem 11 (The Inverse Function Theorem redux)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a smooth function. If $p \in M$ is such that the differential $dF_p\colon T_pM \to T_{F(p)}N$ is an isomorphism, there are open neighborhoods $U \subseteq M$ and $V \subseteq N$ of p and $F(p)$ for which the restriction $F|_U\colon U \to V$ is a diffeomorphism.

The above is a local statement, and we apply the general philosophy of when dealing with manifolds: **any statement which is purely local and true in $\mathbb{R}^n$, will be true on smooth manifolds**. Namely, we take any charts around p and $F(p)$, and apply the Euclidean version of the Inverse Function Theorem to the local representation of F relative to such charts. There is also a version of the Implicit Function Theorem for manifolds, in terms of the partial differentials from Exercise 79. We will use several consequences of Theorem 11 on what follows. To continue the discussion, note that dF_p being an isomorphism directly implies that $\dim M = \dim N$. But what if the dimensions of M and N were not the same? We cannot require dF_p to be an isomorphism anymore—the next best thing is requiring it to have full rank.

Definition 38 (Immersions and submersions)

Let M and N be smooth manifolds, $p \in M$ be any point, and $F\colon M \to N$ be a smooth function. We say that:

(i) F is an **immersion** at p if $dF_p\colon T_pM \to T_{F(p)}N$ is injective.

(ii) F is a **submersion** at p if $dF_p\colon T_pM \to T_{F(p)}N$ is surjective.

If we say that F is an immersion or submersion without specifying the point p, we mean that F is so at *all* points $p \in M$.

If F above is an immersion at some point, then $\dim M \leq \dim N$ by simple linear algebra. Similarly, if F is a submersion at some point, then $\dim M \geq \dim N$. The local form of immersions and submersions (first seen in Propositions 26 and 27, p. 66, for Euclidean spaces) remain valid: immersions locally look like inclusions, while submersions locally look like projections: the only difference is that this time we use charts to express everything.

Proposition 30 (Local form of immersions)

Let M and N be smooth manifolds with $\dim M = n$ and $\dim N = n+k$, and $F\colon M \to N$ be an immersion at a point $p \in M$. Then, there are charts (U, φ) and (V, ψ) centered at p and $F(p)$ for which the local representation

$$\psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \subseteq \mathbb{R}^n \to \psi(V) \subseteq \mathbb{R}^{n+k} = \mathbb{R}^n \times \mathbb{R}^k$$

is given by $(\psi \circ F \circ \varphi^{-1})(x) = (x, 0)$.

Proposition 31 (Local form of submersions)

Let M and N be smooth manifolds with $\dim M = n+k$ and $\dim N = n$, and $F\colon M \to N$ be a submersion at a point $p \in M$. Then, there are charts (U, φ) and (V, ψ) centered at p and $F(p)$ for which the local representation

$$\psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \subseteq \mathbb{R}^{n+k} = \mathbb{R}^n \times \mathbb{R}^k \to \psi(V) \subseteq \mathbb{R}^n$$

is given by $(\psi \circ F \circ \varphi^{-1})(x, y) = x$.

Corollary 7 (Submersions admit smooth local sections)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a surjective submersion. Then, every $q \in N$ has an open neighborhood $V \subseteq N$ and a smooth function $\sigma\colon V \to M$ such that $F \circ \sigma = \mathrm{Id}_V$. In addition, given any $p \in F^{-1}(q)$, we can construct such σ having $\sigma(q) = p$.

Proof: As F is surjective, there is $p \in M$ such that $F(p) = q$. Now let (U, φ) and (V, ψ) be charts for M and N centered at p and q, respectively, relative to which F looks like a standard projection, cf. Proposition 31; define the function $\sigma\colon V \to M$ by $\sigma(z) = \varphi^{-1}(\psi(z), 0)$. Clearly σ is smooth, and

$$(F \circ \sigma)(z) = (F \circ \varphi^{-1})(\psi(z), 0) = \psi^{-1}(\psi(z)) = z$$

for any $z \in V$, as wanted. ☐

Here, "section" is synonymous with "right-inverse". Corollary 7 may in fact be used as a characterization of submersions:

Exercise 82

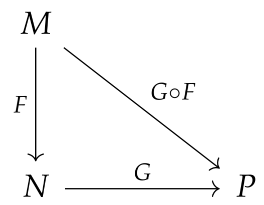
Show that:

(a) σ in Corollary 7 is necessarily an injective immersion.

(b) the converse of Corollary 7 holds: if F is just a smooth surjection which admits local smooth sections around each point in N, then F is a submersion.

Corollary 8 (The characteristic property of surjective submersions)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a surjective submersion. If P is a third manifold and $G\colon N \to P$ is any function, then G is smooth if and only if $G \circ F\colon M \to P$ is smooth, cf. diagram:



Proof: If G is smooth, then $G \circ F$ is a composition of smooth functions, and hence is smooth as well. Conversely, assume that $G \circ F$ is smooth. To establish smoothness of G, we observe that smoothness is a local notion, and so it suffices to argue that G is smooth in some neighborhood of every point in N. But around every point we may find an open neighborhood $V \subseteq N$ together with a local smooth right-inverse $\sigma\colon V \to M$; then $G|_V = (G \circ F) \circ \sigma$ is the composition of smooth functions, and hence smooth as well. □

The moral of the above proof is that if F were a diffeomorphism, we could simply write $G = (G \circ F) \circ F^{-1}$ as a composition of smooth functions and be done with argument. But having smooth local right-inverses is enough (as opposing to having a smooth global two-sided inverse), and F being a submersion is exactly what we need to make it work. In summary, surjective submersions are the smooth analogues of quotient mappings in the setting of topological spaces and continuous functions (i.e., in the "topological category"). Let's see how to use this property in practice.

Example 73

Consider the quotient projection $\pi\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{RP}^n$. We have seen in Example 68 that π is smooth, while it is obviously surjective, but we claim that it is also a submersion. Consider one of the standard charts (U_0, φ_0) for $\mathbb{RP}^n$—recall that

$$U_0 = \{[x_0 : \cdots : x_n] \in \mathbb{RP}^n : x_0 \neq 0\} \quad \text{and} \quad \varphi_0([x_0 : \cdots : x_n]) = \left(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0}\right).$$

The local representation $\varphi_0 \circ \pi\colon (\mathbb{R} \smallsetminus \{0\}) \times \mathbb{R}^n \to \mathbb{R}^n$ (we take the global chart in $\mathbb{R}^{n+1} \smallsetminus \{0\}$ to be the identity function, as usual) is given by

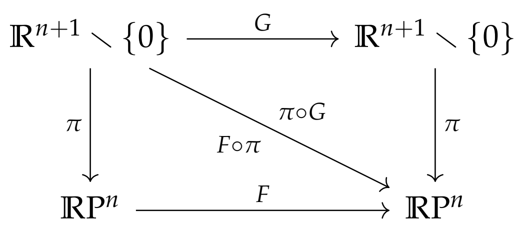
$$(\varphi_0 \circ \pi)(x_0, x_1, \ldots, x_n) = \left(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0}\right),$$

and its Jacobian

$$\mathrm{D}(\varphi_0 \circ \pi)(x_0, x_1, \ldots, x_n) = \begin{bmatrix} -x_1/x_0^2 & 1/x_0 & 0 & \cdots & 0 \\ -x_2/x_0^2 & 0 & 1/x_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -x_n/x_0^2 & 0 & \cdots & 0 & 1/x_0 \end{bmatrix}$$

has full rank due to the $(1/x_0)\mathrm{Id}_n$ block. This means that π is a submersion at all points in U_0, and a similar argument takes care of the points in the remaining charts (U_i, φ_i), for $i = 1, \ldots, n$. This concludes the proof that π is a submersion.

Consider now the function $F\colon \mathbb{R}P^n \to \mathbb{R}P^n$ given by squaring homogeneous coordinates, $F([x_0 : \cdots : x_n]) = [x_0^2 : \cdots : x_n^2]$. We are finally able to argue, without charts, that F is smooth: with the auxilliary function $G\colon \mathbb{R}^{n+1} \smallsetminus \{0\} \to \mathbb{R}^{n+1} \smallsetminus \{0\}$ given by $G(x_0, \ldots, x_n) = (x_0^2, \ldots, x_n^2)$, we have the diagram

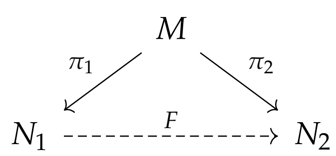


Clearly G is smooth (as its entries are polynomials), and so $F \circ \pi = \pi \circ G$ is also smooth, being the composition of the smooth functions π and G. By the characteristic property of π as a surjective submersion, $F \circ \pi$ being smooth implies that F is smooth, as required.

The characteristic property of surjective submersions can also be used to establish the smooth analogue of Proposition 6 (p. 22):

Exercise 83 (Uniqueness of quotients)

Let M, N_1, and N_2 be smooth manifolds, and $\pi_1\colon M \to N_1$ and $\pi_2\colon M \to N_2$ be surjective submersions. Show that if π_1 and π_2 are constant along each other's fibers, there is a unique diffeomorphism $F\colon N_1 \to N_2$ such that the diagram commutes:



Hint: Draw some inspiration from Claim 5 in the proof of Theorem 9 (p. 89).

As our next step in studying Calculus on manifolds, we may think back to one of the first applications of derivatives: dealing with optimization problems. Doing so relied heavily on the notion of a *critical point* and solving an equation of the form $f'(x) = 0$.

Definition 39 (Critical/regular points/values)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a smooth function. We say that:

(i) $p \in M$ is a **critical point** of F if $\mathrm{d}F_p\colon T_pM \to T_{F(p)}N$ is *not* surjective.

(ii) $p \in M$ is a **regular point** of F if p is not a critical point of F.

(iii) $q \in N$ is a **regular value** of F if $F^{-1}(q)$ consists only of regular points.

(iv) $q \in N$ is a **critical value** of F if $F^{-1}(q)$ contains a critical point.

Above, note that "points" refer to elements of the domain manifold M, while "values" refer to elements of the target manifold N.

Remark. In any smooth manifold M, one can define what the notion of a subset $S \subseteq M$ having **measure zero**: it means that for any chart (U, φ) for M, the image $\varphi(U \cap S) \subseteq \mathbb{R}^n$ has zero Lebesgue measure—for every $\varepsilon > 0$ there exists a sequence $(B_k)_{k\geq 1}$ of open boxes in $\mathbb{R}^n$ such that $\varphi(U\cap S) \subseteq \bigcup_{k\geq 1} B_k$ and $\sum_{k\geq 1} \mathrm{vol}(B_k) < \varepsilon$, where the volume of an open box is explicitly defined as $\mathrm{vol}(\prod_{i=1}^n (a_i, b_i)) = \prod_{i=1}^n (b_i - a_i)$. This definition is "correct" because diffeomorphisms between open subsets of Euclidean spaces send measure zero sets to measure zero sets and, in particular, it implies that checking whether $S \subseteq M$ has measure zero or not requires considering only enough charts to cover S, not M. If S does not have measure zero, it does not mean that we can assign a value to the "measure of S", as that would require equipping M with an actual measure (say, defined on the Borel σ-algebra of M). In any case, **Sard's Theorem** states that the set of critical values of any smooth function between manifolds has zero measure (in the above sense). In other words, *almost all* values are regular.

Definition 39 may sound counter-intuitive at a first glance, as we could reasonably expect that $p \in M$ is a critical point of F if $\ker \mathrm{d}F_p$ is nontrivial. Consider instead the case where $\dim N = 1$, where $\mathrm{d}F_p$ becomes essentially a linear functional, and note the elementary dichotomy: a linear functional on a vector space is either surjective, or the zero functional[7]. Hence, when $\dim N = 1$, p is a critical point if and only if $\mathrm{d}F_p = 0$. And, as a saving grace for such a first expectation, in the case where $\dim M = \dim N$, p is indeed a critical point if and only if $\ker \mathrm{d}F_p$ is nontrivial; diffeomorphisms and local diffeomorphisms (such as $\tau\colon \mathbb{S}^n \to \mathbb{S}^n$ in Example 71) have no critical points. In summary, having "surjective" instead of "injective" in item (i) in Definition 39 is what we need to generalize things correctly.

In any case, this definition of critical point does what it is supposed to do:

[7] If V is a vector space and $\xi \in V^* \smallsetminus \{0\}$, there is $v \in V$ such that $\xi(v) \neq 0$. Then, given any $\lambda \in \mathbb{R}$, we have that $\xi(\lambda v/\xi(v)) = \lambda$, making ξ surjective. Alternatively: $\mathrm{Im}\,\xi$ is a subspace of $\mathbb{R}$, and therefore can only have dimension 0 or 1. These arguments work if we replace $\mathbb{R}$ with any field of scalars $\mathbb{K}$, and also if the dimension of V is infinite.

Proposition 32 (Local maxima or minima are critical points)

Let M be a smooth manifold and $f\colon M \to \mathbb{R}$ be a smooth function. If $p \in M$ is a local maximum (or local minimum) of f, then $\mathrm{d}f_p = 0$.

Proof: Let $v \in T_pM$ be an arbitrary tangent vector, and use Lemma 7 to fix a smooth curve $\alpha\colon (-\varepsilon, \varepsilon) \to M$ be a such that $\alpha(0) = p$ and $\alpha'(0) = v$. As p is a local maximum for f, we have that $t = 0$ is a local maximum of $f \circ \alpha$, and so $\mathrm{d}f_p(v) = (f \circ \alpha)'(0) = 0$ by the corresponding result from single-variable calculus. $\square$

Let's revisit a more elaborate example:

Example 74

Consider again the smooth mapping $F\colon \mathbb{R}\mathrm{P}^n \to \mathbb{R}\mathrm{P}^n$ given by squaring all homogeneous coordinates, $F([x_0 : \cdots : x_n]) = [x_0^2 : \cdots : x_n^2]$. Let's determine the set of all critical points of F. As a general principle, we should try to make an educated guess of what it is before we jump into the calculations. Recall the auxiliary smooth "lift" G of F and the commutative diagram

$$\begin{array}{ccc} \mathbb{R}^{n+1} \smallsetminus \{0\} & \xrightarrow{\quad G \quad} & \mathbb{R}^{n+1} \smallsetminus \{0\} \\ \downarrow \pi & & \downarrow \pi \\ \mathbb{R}\mathrm{P}^n & \xrightarrow{\quad F \quad} & \mathbb{R}\mathrm{P}^n \end{array}$$

As G has Jacobian given by

$$\mathrm{D}G(x_0, \ldots, x_n) = \begin{bmatrix} 2x_0 & & & \\ & 2x_1 & & \\ & & \ddots & \\ & & & 2x_n \end{bmatrix},$$

where the blank entries are equal to zero, is singular if and only if $x_i = 0$ for some index i, we may reasonably guess that

$$\text{the set of critical points of } F \text{ consists of all elements } [x_0 : \cdots : x_n] \in \mathbb{R}\mathrm{P}^n \text{ such that } x_i = 0 \text{ for some index } i. \tag{4.16}$$

This is not immediately obvious from the diagram above, as the differential of π is not invertible. We first claim that, for every $p \in \mathbb{R}^{n+1} \smallsetminus \{0\}$, the equality $\ker \mathrm{d}\pi_p = \mathbb{R}p$ holds. Indeed, as $\mathrm{d}\pi_p\colon \mathbb{R}^{n+1} \to T_{\pi(p)}(\mathbb{R}\mathrm{P}^n)$ is surjective and $\dim \mathbb{R}\mathrm{P}^n = n$, it follows that $\dim \ker \mathrm{d}\pi_p = 1$, and so it suffices to check that

$\mathbb{R}p \subseteq \ker \mathrm{d}\pi_p$. This, in turn, is simple: if $\lambda \in \mathbb{R}$ is arbitrary, we have that

$$\mathrm{d}\pi_p(\lambda p) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \pi(p + \lambda t p) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \pi((1+\lambda t)p) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \pi(p) = 0.$$

Now, we differentiate the equality $F \circ \pi = \pi \circ G$ using the chain rule to obtain

$$\mathrm{d}F_{[x_0:\cdots:x_n]}(\mathrm{d}\pi_{(x_0,\ldots,x_n)}(v_0,\ldots,v_n)) = \mathrm{d}\pi_{(x_0^2,\ldots,x_n^2)}(2x_0v_0,\ldots 2x_nv_n). \tag{4.17}$$

As π is a submersion, (4.17) completely determines $\mathrm{d}F_{[x_0:\cdots:x_n]}$. For example, the derivative $\mathrm{d}F_{[1:\cdots:1]}$ equals twice the identity mapping of the tangent space $T_{[1:\ldots:1]}(\mathbb{R}P^n)$. With this in place and the kernel of $\mathrm{d}\pi$ at each point being known, the following conditions are equivalent:

(i) $[x_0 : \cdots : x_n] \in \mathbb{R}P^n$ is a critical point of F.

(ii) there is $(v_0,\ldots,v_n) \in \mathbb{R}^{n+1}$, not proportional to $(x_0,\ldots,x_n) \in \mathbb{R}^{n+1} \smallsetminus \{0\}$, such that the image $(2x_0v_0,\ldots,2x_nv_n)$ is proportional to $(x_0^2,\ldots,x_n^2)$.

This equivalence is our main tool to prove that (4.16) is in fact true. If $[x_0 : \cdots : x_n]$ is a critical point of F and $(v_0,\ldots,v_n)$ is chosen as in (ii), we have that

$$\operatorname{rank}\begin{bmatrix} 2x_0v_0 & \cdots & 2x_nv_n \\ x_0^2 & \cdots & x_n^2 \end{bmatrix} < 2 \quad \text{and} \quad \operatorname{rank}\begin{bmatrix} v_0 & \cdots & v_n \\ x_0 & \cdots & x_n \end{bmatrix} = 2,$$

which means that $2x_iv_ix_j^2 - 2x_jv_jx_i^2 = 0$ for *all* i and j, that is, $x_ix_j(x_iv_j - x_jv_i) = 0$, while $x_kv_\ell - x_\ell v_k \neq 0$ for *some* k and ℓ. Hence, $x_kx_\ell = 0$, and so at least one of x_k and x_ℓ must vanish. Conversely, let $[x_0 : \cdots : x_n]$ be such that $x_i = 0$ for some i. As any permutation of the coordinates in $\mathbb{R}^{n+1} \smallsetminus \{0\}$ induces a diffeomorphism of $\mathbb{R}P^n$ (why?), we may without loss of generality assume that $[x_0 : \cdots : x_n]$ has the specific form $[1 : 0 : x_3 : \cdots : x_n]$. Then we use (4.17) to compute that

$$\mathrm{d}F_{[1:0:x_3:\cdots:x_n]}(\mathrm{d}\pi_{(1,0,x_3,\ldots,x_n)}(1,1,x_3,\ldots,x_n)) = 2\mathrm{d}\pi_{(1,0,x_3^2,\ldots,x_n^2)}((1,0,x_3^2,\ldots,x_n^2)).$$

The right-hand side clearly vanishes, while $\mathrm{d}\pi_{(1,0,x_3,\ldots,x_n)}(1,1,x_3,\ldots,x_n) \neq 0$, as $(1,0,x_3,\ldots,x_n)$ and $(1,1,x_3,\ldots,x_n)$ are not proportional (compare their second entries). This shows that $[1 : 0 : x_3 : \cdots : x_n]$ is a critical point of F and completes the proof of (4.16).

Exercise 84

In the above setting, show that $\dim\ker \mathrm{d}F_{[x_0:\cdots:x_n]} = |\{i \in \{0,1,\ldots,n\} : x_i = 0\}|$.

Exercise 85

Consider the mapping $F\colon \mathbb{R}^4 \to \mathbb{R}^2$ defined by

$$F(x,y,z,w) = \big(x^2+y\,,\, x^2+y^2+z^2+w^2+y\big).$$

Show that $(0,1) \in \mathbb{R}^2$ is a regular value of F, and that $F^{-1}(0,1)$ is diffeomorphic to $\mathbb{S}^2$ (exhibit the diffeomorphism).

4.6 Submanifolds and embeddings

As we have seen, any n-dimensional smooth manifold M is locally modelled on $\mathbb{R}^n$. A k-dimensional submanifold S of M, however we define the term "submanifold", should be such that the set-inclusion $S \subseteq M$ is locally modelled on the inclusion $\mathbb{R}^k \subseteq \mathbb{R}^n$, where here we of course identify the subspace $\mathbb{R}^k$ with its isomorphic copy $\mathbb{R}^k \times \{0\} \subseteq \mathbb{R}^k \times \mathbb{R}^{n-k} = \mathbb{R}^n$. We now make this precise:

Definition 40 (Regular submanifolds)

Let M be a smooth manifold, with $n = \dim M$, and let $S \subseteq M$ be any subset. We will say that S is **k-dimensional regular submanifold of M** if around each $p \in S$ there is a chart (U,φ) for M centered at p such that $\varphi(U \cap S) = \varphi(U) \cap (\mathbb{R}^k \times \{0\})$. See Figure 51.

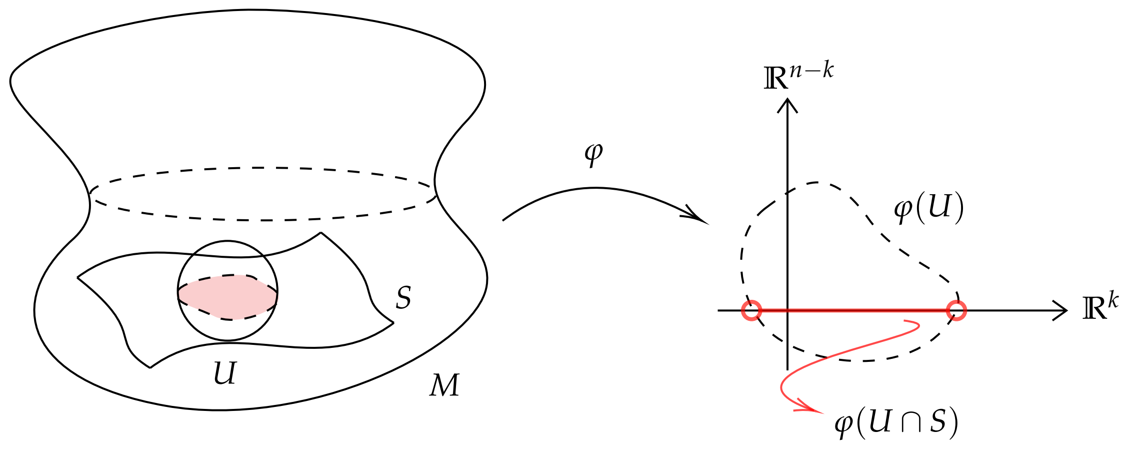


Figure 51: A slice chart for a regular submanifold.

In this setting, (U,φ) is called a **slice chart for M adapted to S**.

Remark. Writing $\varphi = (x^1,\ldots,x^n)$, the slice condition says that S is described by the $n-k$ equations $x^{k+1} = \cdots = x^n = 0$. The integer $n-k$ is called the **codimension** of S in M, and we say that S is a **hypersurface**[8] if its codimension equals 1. Often, slice

[8]As a general principle, the prefix "hyper" means "codimension 1". For example, vector subspaces of a n-dimensional vector space having dimension equal to $n-1$ are called hyperplanes. Regular surfaces in $\mathbb{R}^3$ are hypersurfaces. Curves in surfaces are also technically hypersurfaces, etc.

charts are defined in an apparently broader way, with $\varphi(U \cap S) = \varphi(U) \cap (\mathbb{R}^k \times \{c\})$ for some $c \in \mathbb{R}^{n-k}$, depending on p and φ, so that S gets described by the $n-k$ equations $x^{k+1} = c^{k+1}, \ldots, x^n = c^n$. Of course, this makes no difference: composing a slice chart in this sense with a suitable translation we obtain a slice chart in the sense of the above definition (that is, with $c = 0$). When discussing examples later, we may use this notion of slice chart with general c without further comment.

Example 75 (Vector subspaces are regular submanifolds)

Let V be an n-dimensional vector space, and $W \subseteq V$ be a k-dimensional vector subspace. We have seen that the standard Euclidean structure of V as a manifold is induced by global linear charts $V \to \mathbb{R}^n$ arising from bases for V. Consider instead a basis $(e_1, \ldots, e_k)$ of the subspace W, and complete it to a basis $(e_1, \ldots, e_k, e_{k+1}, \ldots, e_n)$ of V. The corresponding chart $\varphi\colon V \to \mathbb{R}^n$, defined by writing $v \in V$ as a linear combination $v = \sum_{i=1}^n a^i e_i$ and setting $\varphi(v) = (a^1, \ldots, a^n)$, evidently satisfies that $\varphi(W) = \mathbb{R}^k \times \{0\}$. In other words, (V, φ) is a global slice chart for V adapted to W, making W a k-dimensional regular submanifold of V.

Exercise 86

Let M_1 and M_2 be smooth manifolds, and $S_1 \subseteq M_1$ and $S_2 \subseteq M_2$ be regular submanifolds. Show that $S_1 \times S_2$ is a regular submanifold of $M_1 \times M_2$.

Before we see more examples, there are subtleties which need to be addressed.

Let's make a quick analogy. When we have some algebraic structure, for example, a group or a vector space, we have the definition of a subgroup or a vector subspace as being nonempty subsets closed under all the relevant operations. A crucial point here, and the reason why these definitions are appropriate when developing the theory, is that subgroups and vector subspaces are again groups and vector spaces, *on their own right*. This means that, if Definition 40 is to be of any good for us, it should imply (in a somewhat natural way) that S can be made into a smooth manifold as well.

The next concept will play a central role in the discussion:

Definition 41 (Embedding)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a smooth mapping. We say that F is an **embedding** if F is an immersion and a homeomorphism onto its image.

Remark. In the above definition, the image $F(M)$ is equipped with its subspace topology induced from N. We have already seen that a continuous bijection from a compact space to a Hausdorff space is automatically a homeomorphism, so we obtain the following shortcut: if F is an injective immersion and M is compact, then F is automatically an embedding and the image $F(M)$ is closed in N.

Exercise 87 (Proper injective immersions are embeddings with closed images)

Generalize the remark. Let M and N be smooth manifolds, and $F\colon M \to N$ be an injective immersion. Show that if F is **proper**, i.e., $F^{-1}(K) \subseteq M$ is compact whenever $K \subseteq N$ is compact, then F is an embedding and the image $F(M)$ is closed in N.

Exercise 88 (Embedding $\mathbb{RP}^2$ in $\mathbb{R}^4$)

Let $F\colon \mathbb{S}^2 \to \mathbb{R}^4$ be given by $F(x,y,z) = (x^2 - y^2, xy, xz, yz)$. Show that F induces a smooth embedding $\widetilde{F}\colon \mathbb{RP}^2 \to \mathbb{R}^4$ (with $\widetilde{F} \circ \pi = F$, where $\pi\colon \mathbb{S}^2 \to \mathbb{RP}^2$ is the natural projection).

Hint: You may freely use the fact that an injective immersion defined on a compact manifold is an embedding. Can you justify that as well?

Back to the matter at hand:

Theorem 12 (Regular submanifolds are actually manifolds)

Let M be a smooth manifold, with $n = \dim M$, and $S \subseteq M$ be a k-dimensional regular submanifold. Then S is a topological manifold, and it can be made into a smooth manifold so that the inclusion $\iota_S\colon S \hookrightarrow M$ is an embedding.

Proof: We obviously equip S with its subspace topology induced from M, so that S is automatically Hausdorff and second-countable. Writing $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$, we let $\iota\colon \mathbb{R}^k \hookrightarrow \mathbb{R}^n$ and $\pi\colon \mathbb{R}^n \to \mathbb{R}^k$ be given by $\iota(x) = (x, 0)$ and $\pi(x,y) = x$. For any slice chart (U, φ) for M adapted to S, we may define a chart $(U \cap S, \varphi_S)$ for S by letting $\varphi_S\colon U \cap S \to \pi(\varphi(U \cap S)) \subseteq \mathbb{R}^k$ be given by $\varphi_S = \pi \circ \varphi|_{U\cap S}$. Clearly φ_S is continuous, and the image $\pi(\varphi(U \cap S))$ is open in $\mathbb{R}^k$ because $\varphi(U \cap S)$ is open in $\mathbb{R}^k \times \{0\}$, and π restricts to a homeomorphism $\mathbb{R}^k \times \{0\} \cong \mathbb{R}^k$. Then, the inverse function $\varphi_S^{-1}\colon \pi(\varphi(U \cap S)) \to U \cap S$ is explicitly given by $\varphi_S^{-1} = (\varphi|_{U\cap S})^{-1} \circ \iota$, again a composition of continuous functions. This shows that φ_S is a homeomorphism onto its image, so that $(U \cap S, \varphi_S)$ is indeed a chart for S, and S is locally Euclidean.

Hence, S is a topological manifold. To upgrade S to a smooth manifold, we naturally consider the collection

$$\mathfrak{A}_S = \{(U \cap S, \varphi_S) : (U, \varphi) \in \mathfrak{A} \text{ is a slice chart for } M \text{ adapted to } S\},$$

where $\mathfrak{A}$ is the smooth structure of M. To see that $\mathfrak{A}_S$ is a C^∞-atlas for S, it suffices to note that whenever $(U \cap S, \varphi_S)$ and $(V \cap S, \psi_S)$ are induced by slice charts (U, φ) and (V, ψ), the transition $\psi_S \circ \varphi_S^{-1} = \pi \circ (\psi \circ \varphi^{-1}) \circ \iota$ is a composition of smooth mappings between open subsets of Euclidean spaces; reversing the roles of (U, φ) and (V, ψ), we conclude that $(U \cap S, \varphi_S)$ and $(V \cap S, \varphi_S)$ are indeed C^∞-compatible.

Finally, consider the inclusion $\iota_S\colon S \hookrightarrow M$. It is trivially a homeomorphism onto its image, by definition of subspace topology. To see that it is smooth and an immersion, consider its local representation relative to charts $(U \cap S, \varphi_S)$ and (U, φ) for S and M, respectively: $\varphi \circ \iota_S \circ \varphi_S^{-1} \colon \pi(\varphi(U \cap S)) \to \varphi(U)$ is explicitly given by the Euclidean inclusion, $\varphi \circ \iota_S \circ \varphi_S^{-1} = \iota|_{\pi(\varphi(U \cap S))}$, which is the restriction of an injective and linear mapping. Hence ι_S is an embedding, as required. $\square$

With the notation of the result above, we in particular have that the differential $\mathrm{d}(\iota_S)_p \colon T_pS \to T_pM$ is injective for each $p \in S$. It is common to identify the tangent space T_pS with its image $\mathrm{d}(\iota_S)_p(T_pS)$. In practice, this just means that we treat T_pS as a vector subspace of T_pM.

Example 76

Consider the setting of Example 75, with an n-dimensional vector space V, a k-dimensional vector subspace $W \subseteq V$, and the global slice chart $\varphi \colon V \to \mathbb{R}^n$ induced by a basis $\mathcal{B}$ of W completed to a basis of V. The induced chart $\varphi_W \colon W \to \mathbb{R}^k$ in fact equals the linear isomorphism $W \to \mathbb{R}^k$ provided by the initial basis $\mathcal{B}$ chosen for W. This means that the smooth structure induced on W via the result of Theorem 12 agrees with the standard Euclidean one.

Here is one example, certainly familiar to the reader from Multivariable Calculus:

Example 77 (Spherical coordinates are a slice chart for $\mathbb{R}^3$ adapted to $\mathbb{S}^2$)

Consider the unit sphere $\mathbb{S}^2 \subseteq \mathbb{R}^3$. We claim that $\mathbb{S}^2$ is a regular surface in $\mathbb{R}^3$. Indeed, whenever $p_0 \in \mathbb{S}^2$ and we let $\theta_0 \in [0, 2\pi)$ and $\phi_0 \in [0, \pi)$ be the angles between the projection of p_0 to $\mathbb{R}^2 \times \{0\}$ and between p_0 and $(0,0,1)$, we may consider spherical coordinates

$$(\rho, \theta, \phi) \colon U \subseteq \mathbb{R}^3 \to (0, \infty) \times (\theta_0 - \pi, \theta_0 + \pi) \times \left(\phi_0 - \frac{\pi}{2}, \phi_0 + \frac{\pi}{2}\right),$$

characterized as usual by $x = \rho \cos\theta \sin\phi$, $y = \rho \sin\theta \sin\phi$, and $z = \rho\cos\phi$, on a suitable chart domain U—it is roughly an open cone, see Figure 52.

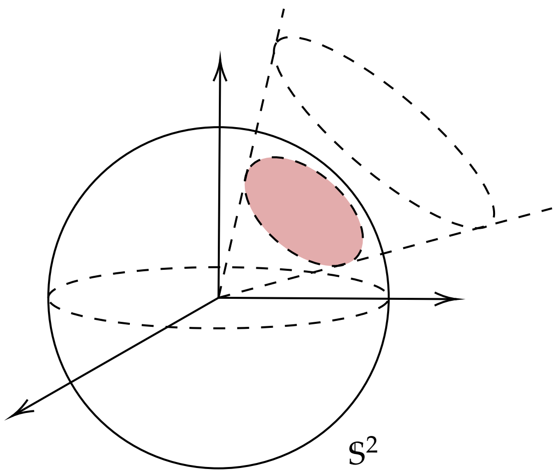


Figure 52: Spherical coordinates are slice charts for the sphere.

As $U \cap \mathbb{S}^2$ is described as $\rho = 1$, $(U; \rho, \theta, \phi)$ is a slice chart for $\mathbb{R}^3$ adapted to $\mathbb{S}^2$.

Applying Theorem 12 to the situation given in the above example, we obtain a smooth structure on the sphere $\mathbb{S}^2$. The issue here is that, a priori, this smooth structure has no reason to agree with the smooth structure we have already been working with. In this case, things do turn out to be ok and the structures in question agree (see Theorem 15 ahead). However, it is perfectly possible for M and S to be smooth manifolds such that $S \subseteq M$, and yet S fails to be a regular submanifold of M: the charts for S need not be induced by slice charts for M adapted to S.

See the next two classical non-examples:

Example 78 (An injective immersion which is not an embedding)

Consider the following curve $S \subseteq \mathbb{R}^2$:

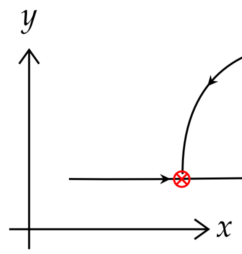


Figure 53: A curve which is immersed but not embedded in $\mathbb{R}^2$.

Such S may be realized as the image of an injective immersion $F\colon \mathbb{R} \to \mathbb{R}^2$, and so we may transfer the topology and smooth structure of $\mathbb{R}$ to S via F. However, S with this structure cannot be a submanifold of $\mathbb{R}^2$, as the inclusion $\iota_S\colon S \hookrightarrow \mathbb{R}^2$ fails to be a homeomorphism onto its image—namely, with the image $\iota(S) = S$ being equipped with its subspace topology induced from $\mathbb{R}^2$, the inverse function $\iota^{-1}|_{\iota(S)}\colon \iota(S) \to S$ is not continuous at the point indicated in Figure 53. (Can you see why?)

Example 79 (The Kronecker flow on $\mathbb{T}^2$)

Consider the torus $\mathbb{T}^2$, seen as the quotient space $\mathbb{R}^2/\mathbb{Z}^2$, as well as the curve $\alpha\colon \mathbb{R} \to \mathbb{T}^2$ given by $\alpha(t) = \pi(t, mt)$, where $m > 0$ is some fixed slope and $\pi\colon \mathbb{R}^2 \to \mathbb{T}^2$ is the quotient projection. We claim that

$$\begin{array}{c}\alpha \text{ is smooth and an immersion, injective or periodic} \\ \text{according to whether } m \notin \mathbb{Q} \text{ or } m \in \mathbb{Q}\text{, respectively.}\end{array} \tag{4.18}$$

Indeed, α is manifestly smooth (being a composition of the two smooth functions π and $\mathbb{R} \ni t \mapsto (t, mt) \in \mathbb{R}$) and the velocity vector $\alpha'(t) = \mathrm{d}\pi_{(t,mt)}(1, m)$ never vanishes, as $\mathrm{d}\pi_{(t,mt)}\colon \mathbb{R}^2 \to T_{\pi(t,mt)}\mathbb{T}^2$ is an isomorphism and $(1, m) \neq (0, 0)$; this shows that α is an immersion. Finally, whenever $t, s \in \mathbb{R}$ are such that $\alpha(t) = \alpha(s)$, there is $(k, \ell) \in \mathbb{Z}^2$ such that $(t, mt) = (s + k, ms + \ell)$, leading to the relation $m(s + k) = ms + \ell$, and hence to $mk = \ell$. As a consequence, if $m \notin \mathbb{Q}$, we must have that $k = \ell = 0$, so that α is injective. Then, if $m = \ell/k \in \mathbb{Q}$, we have that

$k \neq 0$ and so $(t+k, (\ell/k)(t+k)) = (t+k, (\ell/k)t + \ell)$, leading to $\alpha(t+k) = \alpha(t)$ for any $t \in \mathbb{R}$, and making α periodic. This establishes (4.18).

To visualize the image $S = \alpha(\mathbb{R})$, consider it first in the square $[0,1]^2$ with its sides identified, as in Figure 54. Also compare it with its version in the actual torus, in Figure 55.

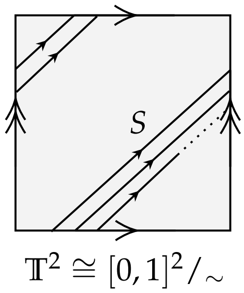


Figure 54: The curve S in the square.

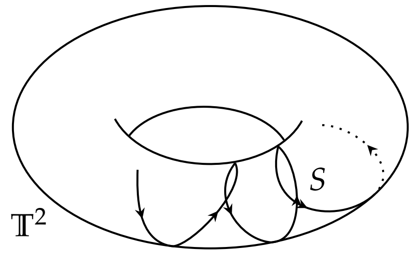


Figure 55: The curve S in the torus.

With (4.18) in mind, we address two cases:

- If $m \in \mathbb{Q}$, the curve α induces an embedding $\widetilde{\alpha}\colon \mathbb{S}^1 \to \mathbb{T}^2$, and we have that $\widetilde{\alpha}(\mathbb{S}^1) = S$. As we will see ahead in Theorem 13, this guarantees that S is a regular submanifold of $\mathbb{T}^2$.

- If $m \notin \mathbb{Q}$, then $\alpha\colon \mathbb{R} \to S$ is a bijection and so S can be made into a smooth manifold, by simply transferring the topology and smooth structure from $\mathbb{R}$ to S via α—this construction makes $\alpha^{-1}\colon S \to \mathbb{R}$ a global chart for S. However, with this topology and smooth structure, the inclusion $\iota_S\colon S \hookrightarrow \mathbb{T}^2$ fails to be a homeomorphism onto its image $\iota(S)$ (equipped with the subspace topology induced from $\mathbb{T}^2$). Namely, **Kronecker's Approximation Theorem**[a] implies that that $\iota(S)$ is dense in $\mathbb{T}^2$, causing the inverse function $\iota(S) \to S$ to be discontinuous. Of course, we can phrase it more succintly by simply saying that if $S \subseteq \mathbb{T}^2$ has its subspace topology from $\mathbb{T}^2$, the curve α is an injective immersion which is not an embedding.

[a] It states that whenever $m \notin \mathbb{Q}$, for every $x \in \mathbb{R}$ and $\varepsilon > 0$ there are integers $a, b \in \mathbb{Z}$ with $a > 0$ such that $|am - b - x| < \varepsilon$.

The main reason for this issue in the above situations is that, in the bad cases, S simply "had the wrong topology"—the correct one being, of course, the subspace topology induced from either $\mathbb{R}^2$ or $\mathbb{T}^2$. This seems like an appropriate moment to mention an alternative definition of smooth manifold: one starts with merely with a *set* M, and a collection $\mathfrak{A}$ of "charts" (U, φ), that is, $U \subseteq M$ is just a subset and $\varphi\colon U \to \varphi(U) \subseteq \mathbb{R}^n$ is a bijection onto its image, assumed to be an open subset of $\mathbb{R}^n$; it does not make sense to ask whether U is open or not at this stage, since M has no topology. One then assumes conditions on the "atlas" $\mathfrak{A}$:

(i) The chart domains in $\mathfrak{A}$ cover M.

(ii) Whenever $(U,\varphi), (V,\psi) \in \mathfrak{A}$ are such that $U \cap V \neq \varnothing$, the images $\varphi(U \cap V)$ and $\psi(U \cap V)$ are open in $\mathbb{R}^n$ and $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ is of class C^∞.

(iii) $\mathfrak{A}$ is maximal for condition (ii).

Once this is in place, we finally define a topology on M by saying that a subset $W \subseteq M$ is open if and only if $\varphi(U \cap W)$ is open in $\mathbb{R}^n$ for every $(U, \varphi) \in \mathfrak{A}$. This topology makes all chart domains open, and all charts become continuous, but it has a crucial flaw: it is not necessarily Hausdorff of second-countable. These conditions must be explicitly included in the definition of a smooth manifold as additional assumptions. The take-away here is that if you start with a set without a topology, but are still able to find "charts" for it, you can almost treat the topology as an afterthought—you just need to show that it is Hausdorff and second-countable before moving on. End of digression.

Proceeding, there is then a general distinction: if $F\colon M \to N$ is an immersion or embedding, its image is called an **immersed submanifold** or **embedded submanifold** of N. The first case can be somewhat pathological when F is not injective, with its image admitting "self-intersections" (or even if it is injective, "almost self-intersections" like in Figure 53). Immersed submanifolds do appear in practice, when dealing with Lie groups and foliations. Embedded submanifolds, on the other hand, agree with what we have defined, as the topology induced on $F(M)$ by F and M equals the subspace topology induced by N:

Theorem 13 (Images of embeddings are regular submanifolds)

Let M and N be smooth manifolds, and $F\colon M \to N$ be an embedding. Then, the image $F(M)$ is a k-dimensional regular submanifold of N, where $k = \dim M$.

Remark. This result is the reason why the terms "embedded submanifold" and "regular submanifold" are used interchangeably.

Proof: The strategy consists in applying the local form of immersions (Proposition 30) to build slice charts for N adapted to $F(M)$. Namely, given $q \in F(M)$ there is a unique $p \in M$ such that $q = F(p)$, and we may choose charts (U, φ) and (V, ψ) for M and N, centered at p and q, for which $F(U) \subseteq V$ and the local representation $\psi \circ F \circ \varphi^{-1}\colon \varphi(U) \subseteq \mathbb{R}^k \to \psi(V) \subseteq \mathbb{R}^k \times \mathbb{R}^{n-k}$ is given by $(\psi \circ F \circ \varphi^{-1})(x) = (x, 0)$; here, $n = \dim N$. However, we cannot say that (V, ψ) is a slice chart adapted to $F(M)$: for that, we would need to know that $\psi(V \cap F(M)) = \psi(V) \cap (\mathbb{R}^k \times \{0\})$, but the above local representation only ensures that, in V, only $F(U)$ is characterized by the vanishing of the last $n-k$ component functions of ψ, and $F(U)$ might be strictly smaller than $V \cap F(M)$. Consider the situation from Example 78 in more detail:

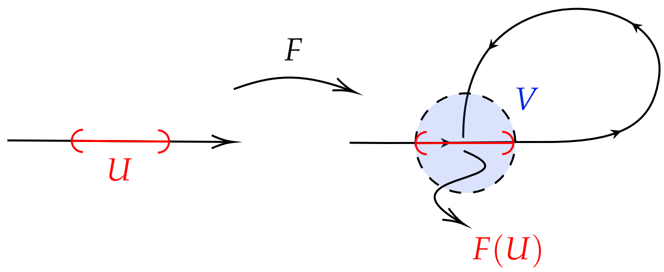


Figure 56: The "tail" of the image of F, not contained in $F(U)$, reenters V.

It is here the assumption that F is an embedding (as opposed to just an immersion) enters. As F is a homeomorphism between M and $F(M)$, when the latter has its subspace topology induced from N, we may use that $F(U)$ is open in $F(N)$ together with the definition of subspace topology to obtain an open subset $V' \subseteq N$ such that $V' \cap F(M) = F(U)$. This does the trick: the restricted chart $(V', \psi|_{V'})$ is a slice chart adapted to $F(M)$, with $\psi(V' \cap F(M)) = \psi(V') \cap (\mathbb{R}^k \times \{0\})$, as required. □

Example 80 (Graphs are embedded submanifolds)

Let M and N be smooth manifolds, with dimensions $\dim M = n$ and $\dim N = m$, and $F\colon M \to N$ be a smooth function. We claim that the graph

$$\mathrm{Gr}(F) = \{(p,q) \in M \times N : q = F(p)\}$$

is a regular submanifold of $M \times N$. To see why this is true, we build slice charts for $M \times N$ adapted to $\mathrm{Gr}(F)$. Let (U, φ) and (V, ψ) be charts for M and N around given points p_0 and $F(p_0)$, respectively, and define $\zeta\colon W \to \mathbb{R}^n \times \mathbb{R}^m$ by

$$\zeta(p,q) = \big(\varphi(p), \psi(q) - \psi(F(p))\big),$$

where $W = (U \cap F^{-1}(V)) \times V$. Then (W, ζ) is a chart for $M \times N$, being written as the composition $\Phi \circ (\varphi|_{U \cap F^{-1}(V)} \times \psi)$, where the mapping

$$\Phi\colon \varphi(U \cap F^{-1}(V)) \times \psi(V) \to \mathbb{R}^n \times \mathbb{R}^m$$

given by $\Phi(x,y) = (x, y - (\psi \circ F \circ \varphi^{-1})(x))$ is also a homeomorphism onto its image; in particular, $\zeta(W)$ is open in $\mathbb{R}^n \times \mathbb{R}^m$. Finally, it is clear that

$$\zeta(W \cap \mathrm{Gr}(F)) = \zeta(W) \cap (\mathbb{R}^n \times \{0\}),$$

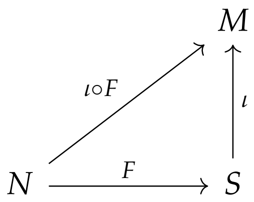
as wanted.

Alternatively, $\mathrm{Gr}(F)$ is the image of the embedding $M \ni p \mapsto (p, F(p)) \in M \times N$. This is indeed a homeomorphism onto $\mathrm{Gr}(F)$, with continuous inverse given by $\mathrm{Gr}(F) \ni (p, F(p)) \mapsto p \in M$. As we will see ahead in Theorem 15, the smooth structure on $\mathrm{Gr}(F)$ induced from M via such embedding agrees with the smooth structure induced by slice charts via Theorem 12.

The smooth structure induced by slice charts turns out to be the precise smooth analogue of the subspace topology, in the sense of the next result:

Theorem 14 (The characteristic property of embeddings)

Let S and M be smooth manifolds, and $\iota\colon S \to M$ be an embedding. If N is a third manifold and $F\colon N \to S$ is any function, then F is smooth if and only if the composition $\iota \circ F\colon N \to M$ is smooth, cf. the next diagram.



Proof: It suffices to argue that if $\iota \circ F$ is smooth, so is F, with the converse statement being obvious from Proposition 28. We identify S with its image $\iota(S)$, as to treat it as a regular submanifold of M. Let $(V \cap S, \psi_S)$ be a chart for S induced from a slice chart (V, ψ) for M adapted to S, and let (U, φ) be any chart for N. Write $n = \dim M$ and $k = \dim S$. By assumption, the components $F^1, \ldots, F^n\colon \varphi(U \cap F^{-1}(V)) \to \mathbb{R}$ of the local representation $\psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \to \psi(V)$, defined via

$$(\psi \circ F \circ \varphi^{-1})(x) = (F^1(x), \ldots, F^k(x), F^{k+1}(x), \ldots, F^n(x)),$$

are all Euclidean-smooth. But $F(\varphi^{-1}(x)) \in S$ implies that $F^{k+1}(x) = \cdots = F^n(x) = 0$, making the local representation $\psi_S \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V \cap S)) \to \psi_S(V \cap S)$ explicitly given by $(\psi_S \circ F \circ \varphi^{-1})(x) = (F^1(x), \ldots, F^k(x))$—it is also Euclidean-smooth, as required. □

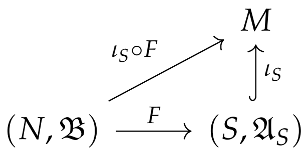
Theorem 14 above combined with the characteristic property of surjective submersions (Corollary 8) provides a huge shortcut for checking whether functions between manifolds are smooth or not. For example, if $F\colon M \to \mathbb{S}^n$ is any function and we wish to show that it is smooth, we can completely ignore $\mathbb{S}^n$, regard F as valued in $\mathbb{R}^{n+1}$, and just check that its $n+1$ component functions $M \to \mathbb{R}$ are individually smooth, which is a much easier task; of course, we can replace $\mathbb{S}^n$ with any embedded submanifold of Euclidean space here.

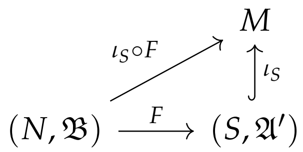
Another consequence of Theorem 14 is the extremely comforting uniqueness result:

Theorem 15 (Uniqueness of embedded-submanifold structures)

Let M be a smooth manifold, and $S \subseteq M$ be a regular submanifold. Then, the topology and smooth structure on S making the inclusion $\iota_S\colon S \hookrightarrow M$ an embedding are unique.

Proof: Let $\mathfrak{A}_S$ be the smooth structure on S induced by slice charts via Theorem 12, and let $\mathfrak{A}'$ be any other smooth structure on S for which $\iota_S\colon S \hookrightarrow M$ is an embedding. As done in the proof of Theorem 9 (more precisely, in Claim 5), we will argue that $\mathrm{Id}_S\colon (S, \mathfrak{A}_S) \to (S, \mathfrak{A}')$ is a diffeomorphism (and therefore $\mathfrak{A}' = \mathfrak{A}_S$). Whenever $(N, \mathfrak{B})$ is a third smooth manifold, Theorem 14 says that a mapping $F\colon (N, \mathfrak{B}) \to (S, \mathfrak{A}_S)$ is smooth if and only if $\iota_S \circ F\colon (N, \mathfrak{B}) \to M$ is smooth; and similarly if we replace $\mathfrak{A}_S$ with $\mathfrak{A}'$.





Set $(N,\mathfrak{B}) = (S,\mathfrak{A}')$ in the first diagram and let $F = \mathrm{Id}_S$, so that smoothness of $\iota_S\colon (S,\mathfrak{A}') \hookrightarrow M$ yields the one of $\mathrm{Id}_S\colon (S,\mathfrak{A}') \to (S,\mathfrak{A}_S)$. Likewise, if $(N,\mathfrak{B}) = (S,\mathfrak{A}_S)$ in the second diagram and $F = \mathrm{Id}_S$, smoothness of $\iota_S\colon (S,\mathfrak{A}_S) \hookrightarrow M$ yields the one of $\mathrm{Id}_S\colon (S,\mathfrak{A}_S) \to (S,\mathfrak{A}')$, concluding the argument. □

To continue the discussion and move towards our next big result on submanifolds, we first register one more consequence of the Inverse Function Theorem:

Corollary 9

Let M be an n-dimensional smooth manifold, $p \in M$ be any point, and $U \subseteq M$ be an open neighborhood of p. If $f_1,\ldots,f_n\colon U \to \mathbb{R}$ are smooth functions and $F = (f_1,\ldots,f_n)\colon U \to \mathbb{R}^n$ is such that $\mathrm{d}F_p\colon T_pM \to \mathbb{R}^n$ is an isomorphism, then, reducing U if needed, we have that (U,F) is a chart for M.

There was nothing special about the specific use of spherical coordinates in Example 77, or the fact that we were dealing with $\mathbb{S}^2$ instead of a more general $\mathbb{S}^n$:

Example 81 (Spheres are submanifolds, redux)

Consider the unit sphere $\mathbb{S}^n \subseteq \mathbb{R}^{n+1}$, and write points in $\mathbb{R}^{n+1}$ as tuples $(x_0,\ldots,x_n)$. Let's show that $\mathbb{S}^n$ is a regular submanifold of $\mathbb{R}^{n+1}$, by exhibiting slice charts around any point $p \in \mathbb{S}^n$. Without loss of generality, assume that $p = (p_0,\ldots,p_n)$ has $p_n > 0$. To define a slice chart, we replace the last coordinate function x_n with a defining function for $\mathbb{S}^n$: if $U = \mathbb{R}^n \times (0,\infty) \subseteq \mathbb{R}^{n+1}$, we let $F\colon U \to \mathbb{R}^{n+1}$ be defined by

$$F(x_0,\ldots,x_n) = \big(x_0, x_1, \ldots, x_{n-1}, x_0^2 + \cdots + x_n^2 - 1\big).$$

Clearly F is smooth, and the Jacobian matrix

$$DF(p) = \begin{bmatrix} & \mathrm{Id}_n & & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 2p_0 & \cdots & 2p_{n-1} & 2p_n \end{bmatrix}$$

is nonsingular precisely because $\det DF(p) = 2p_n \neq 0$. Corollary 9 now allows us to assume, reducing U if needed, that F defines a chart for $\mathbb{R}^{n+1}$. But this is a slice chart adapted to $\mathbb{S}^n$, since the intersection $U \cap \mathbb{S}^n$ is defined by the vanishing of the last component of F, i.e., $F(U \cap \mathbb{S}^n) = F(U) \cap (\mathbb{R}^n \times \{0\})$.

Remark. Strictly speaking, the use of Corollary 9 above was not needed, and we could have directly checked that $F\colon \mathbb{R}^n \times (0,\infty) \to \mathbb{R}^n \times (-1,\infty)$ was already a chart. But the way we argued here is how we will establish a more general result ahead.

The idea used in Example 81 to build slice charts for $\mathbb{R}^{n+1}$ adapted to $\mathbb{S}^n$ by replacing one of the natural coordinates for $\mathbb{R}^{n+1}$ with a defining function for $\mathbb{S}^n$, namely, $(x_0,\ldots,x_n) \mapsto x_0^2 + \cdots + x_n^2 - 1$, works out abstractly:

Theorem 16 (The regular level set theorem – baby version)

Let M be a smooth manifold, $f\colon M \to \mathbb{R}$ be a smooth function, and $c \in \mathbb{R}$ be a regular value of f. Then, whenever $f^{-1}(c) \neq \varnothing$, it is an embedded hypersurface of M, and for every $p \in f^{-1}(c)$ we have that $T_p(f^{-1}(c)) = \ker \mathrm{d}f_p$.

Proof: Let $p_0 \in f^{-1}(c)$ be arbitrary, and fix a chart (U, φ) for M around p_0. As c is a regular value of f, the differential $\mathrm{d}f_{p_0}\colon T_{p_0}M \to \mathbb{R}$ is nonzero. This is equivalent to the Euclidean gradient of the local representation $f \circ \varphi^{-1}\colon \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ being nonzero at $\varphi(p_0)$. Without loss of generality, if $\varphi = (x^1, \ldots, x^n)$, assume that it is its last component $(\partial f/\partial x^n)(p_0) \neq 0$ which survives. Now, define a new smooth mapping $\varphi_f\colon U \to \mathbb{R}^n$ by $\varphi_f(p) = (x^1(p), \ldots, x^{n-1}(p), f(p) - c)$, and note that its differential $\mathrm{d}(\varphi_f)_{p_0}\colon T_{p_0}M \to \mathbb{R}^n$ is an isomorphism. Indeed, its matrix representation

$$\mathrm{D}(\varphi_f \circ \varphi^{-1})(\varphi(p_0)) = \left[\begin{array}{ccc|c} & & & 0 \\ & \mathrm{Id}_{n-1} & & \vdots \\ & & & 0 \\ * & \cdots & * & (\partial f/\partial x^n)(p_0) \end{array}\right]$$

is nonsingular, as $\det \mathrm{D}(\varphi_f \circ \varphi^{-1})(\varphi(p_0)) = (\partial f/\partial x^n)(p_0) \neq 0$. Corollary 9 now tells us that, reducing U if necessary, (U, φ_f) is a chart for M around p_0. This is a slice chart adapted to $f^{-1}(c)$ as $f(p) = c$ if and only if the last component of $\varphi_f(p)$ vanishes, that is, $\varphi_f(U \cap f^{-1}(c)) = \varphi_f(U) \cap (\mathbb{R}^{n-1} \times \{0\})$. We conclude from Theorem 12 that $f^{-1}(c)$ is an embedded hypersurface of M.

It remains to compute, for each $p \in f^{-1}(c)$, the tangent space $T_p(f^{-1}(c))$ (which is identified with a vector subspace of T_pM). We already know that both $T_p(f^{-1}(c))$ and $\ker \mathrm{d}f_p$ have dimension $n-1$, since $\dim f^{-1}(c) = n-1$ and we may apply the rank-nullity law to the linear functional $\mathrm{d}f_p\colon T_pM \to \mathbb{R}$. This means that to establish the equality $T_p(f^{-1}(c)) = \ker \mathrm{d}f_p$, it suffices to show that $T_p(f^{-1}(c)) \subseteq \ker \mathrm{d}f_p$. So, let $v \in T_p((f^{-1}(c))$, and write it (using Lemma 7) as $v = \alpha'(0)$, for some smooth curve $\alpha\colon (-\varepsilon, \varepsilon) \to f^{-1}(c)$ having $\alpha(0) = p$. As $f(\alpha(t)) = c$ for all $t \in (-\varepsilon, \varepsilon)$, we may differentiate it at $t = 0$ to obtain (via formula 4.15 in p. 106) that $\mathrm{d}f_p(v) = 0$, as required. □

In concrete examples, when the ambient space is some Euclidean space and we have identified the correct defining function to use, we may think of its gradient instead of its differential.

Example 82

Consider the set $M = \{(x,y,z) \in \mathbb{R}^3 : xyz = 1\}$. To show that M is an embedded surface in $\mathbb{R}^3$, we simply consider the smooth function $f\colon \mathbb{R}^3 \to \mathbb{R}$ given by $f(x,y,z) = xyz$, so that $M = f^{-1}(1)$, and argue that 1 is a regular value of f: whenever $xyz = 1$, none of the components of $\nabla f(x,y,z) = (yz, xz, xy)$ can vanish, allowing us to apply Theorem 16.

The tangent space $T_{(x,y,z)}M$ is simply the orthogonal complement of the gradient $\nabla f(x,y,z)$, and so $T_{(x,y,z)}M = \{(a,b,c) \in \mathbb{R}^3 : yza + xzb + xyc = 0\}$. Another way to obtain this equation is to start with $xyz = 1$, pretend that x, y, and z all depend on some parameter t, with derivatives a, b, and c, and just differentiate using the product rule. Having checked that $\nabla f(x,y,z)$ is not the zero vector ensures that the resulting equation for the tangent plane is indeed the equation for a plane (instead of degenerating to a line equation, for instance).

With the same argument as above one can also show that any $c \neq 0$ is a regular value of f. However, note that $c = 0$ is not a regular value of f and that the corresponding set $\{(x,y,z) \in \mathbb{R}^3 : xyz = 0\}$ is not an embedded surface in $\mathbb{R}^3$—it is instead the union of the three coordinate planes, which is not even a topological manifold (adapt Example 59, p. 69).

Exercise 89

For each $c \in \mathbb{R}$, consider the set $M_c = \{(x,y,z) \in \mathbb{R}^3 : xe^y + ye^z + ze^x = c\}$. Show that M_c is an embedded surface in $\mathbb{R}^3$ whenever $c \geq 0$, and determine the tangent plane $T_{(0,0,0)}(M_0)$.

Exercise 90

Find the values of $c \in \mathbb{R}$ for which

$$M_c = \{(x,y,z,w,s) \in \mathbb{R}^5 : \sin x + \cos y + \sin z + \cos w + \sin s = c\}$$

is a regular hypersurface of $\mathbb{R}^5$.

Exercise 91

Find the values of $c \in \mathbb{R}$ for which

$$M_c = \{[x : y : z] \in \mathbb{RP}^2 : xy + yz + xz = c(x^2 + y^2 + z^2)\}$$

is an embedded curve in $\mathbb{RP}^2$.

Hint: Which values of c are regular values of the function from Exercise 61?

Exercise 92 (Lagrange multipliers)

Let M be a smooth manifold, and $S \subseteq M$ be an embedded submanifold, globally expressed at $S = \Phi^{-1}(0)$ for some smooth function $\Phi = (\Phi^1, \ldots, \Phi^k) : M \to \mathbb{R}^k$ having the origin $0 \in \mathbb{R}^k$ as a regular value. Let $f : M \to \mathbb{R}$ be a smooth function, and assume that $p \in S$ is a point where the restriction $f|_S : S \to \mathbb{R}$ attains its maximum or minimum value. Show that there are coefficients $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that $\mathrm{d}f_p = \lambda_1 \mathrm{d}\Phi^1_p + \cdots + \lambda_k \mathrm{d}\Phi^k_p$.

Before moving on, here is one more elaborate example:

Example 83 (The special linear group)

Given an integer $n \geq 1$, consider the **special linear group**

$$\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} : \det A = 1\}.$$

It follows from the identities $\det(AB) = \det A \det B$ and $\det(A^{-1}) = (\det A)^{-1}$ that $\mathrm{SL}_n(\mathbb{R})$ is a group. We now claim it is an embedded hypersurface of $\mathbb{R}^{n\times n}$ as well. Clearly we have that $\mathrm{SL}_n(\mathbb{R}) = \det^{-1}(1)$, for the determinant function $\det\colon \mathbb{R}^{n\times n} \to \mathbb{R}$. Note that det is smooth, as $\det A$ is a polynomial function in the n^2 entries of A. We just need to check that $1 \in \mathbb{R}$ is a regular value of det. In other words, we need to verify that whenever $A \in \mathrm{SL}_n(\mathbb{R})$, the differential $\mathrm{d}(\det)_A\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ is not the zero linear functional. Even though having an explicit formula for $\mathrm{d}(\det)_A$ will be convenient for describing the tangent spaces to $\mathrm{SL}_n(\mathbb{R})$ shortly, we don't really need it to apply Theorem 16: just note that

$$\begin{aligned} \mathrm{d}(\det)_A(A) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \det(A + tA) \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \det((1+t)A) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (1+t)^n \det A = n \neq 0. \end{aligned}$$

More generally, we start computing the differential $\mathrm{d}(\det)_{\mathrm{Id}_n}\colon \mathbb{R}^{n\times n} \to \mathbb{R}$. If $e_1, \ldots, e_n$ denotes the standard basis of $\mathbb{R}^n$, and $H = [h_{ij}]_{i,j=1}^n \in \mathbb{R}^{n\times n}$ is arbitrary, we note that the i-th column of H is He_i, and treat det as a multilinear function of the columns of its matrix inputs:

$$\begin{aligned} \mathrm{d}(\det)_{\mathrm{Id}_n}(H) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \det(e_1 + tHe_1, \ldots, e_n + tHe_n) \\ &= \sum_{i=1}^n \det(e_1, \ldots, e_{i-1}, He_i, e_{i+1}, \ldots, e_n) \\ &= \sum_{i=1}^n h_{ii} = \mathrm{tr}(H). \end{aligned}$$

For the general case, we differentiate the relation $\det(AB) = \det A \det B$ at the identity $B = \mathrm{Id}_n$ and evaluate it at the tangent vector $A^{-1}H \in \mathbb{R}^{n\times n}$, using the chain rule: $\mathrm{d}(\det)_A(H) = (\det A)\mathrm{tr}(A^{-1}H)$ whenever $A \in \mathbb{R}^{n\times n}$ is invertible. It follows that

$$T_A\mathrm{SL}_n(\mathbb{R}) = \{H \in \mathbb{R}^{n\times n} : \mathrm{tr}(A^{-1}H) = 0\}.$$

In particular, $T_{\mathrm{Id}_n}\mathrm{SL}_n(\mathbb{R}) = \{H \in \mathbb{R}^{n\times n} : \mathrm{tr}(H) = 0\}$.

Exercise 93

Let $n, k \geq 1$ be integers, and let $M = \{A \in \mathbb{R}^{n\times n} : \operatorname{tr}(A^k) = 1\}$. Show that M is an embedded hypersurface in $\mathbb{R}^{n\times n}$, and determine each tangent space $T_A M$.

Hint: Use cyclic-invariance of the trace function.

The main shortcoming of Theorem 16 is that it only applies to functions which are valued in the real line $\mathbb{R}$. But we will not let this stop us:

Theorem 17 (The regular level set theorem – the true statement)

Let M and N be smooth manifolds, $F\colon M \to N$ be a smooth function, and $q \in N$ be a regular value of F. Then, whenever $F^{-1}(q) \neq \varnothing$, it is an embedded submanifold of M with $\dim F^{-1}(q) = \dim M - \dim N$, and for every $p \in F^{-1}(q)$ we have that $T_p(F^{-1}(q)) = \ker \mathrm{d}F_p$.

Proof: Let $p_0 \in F^{-1}(q)$ be arbitrary, and fix charts (U, φ) and (V, ψ) for M and N around p_0 and q. As q is a regular value of F, the differential $\mathrm{d}F_{p_0}\colon T_{p_0}M \to T_q N$ is surjective. This is equivalent to saying that the Jacobian matrix of the local representation $\psi \circ f \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \subseteq \mathbb{R}^n \to \psi(V) \subseteq \mathbb{R}^m$ at $\varphi(p_0)$ has a nonsingular $m \times m$ submatrix. Without loss of generality, if $\varphi = (x^1, \ldots, x^n)$ and $\psi = (y^1, \ldots, y^m)$, and we write $F^a = y^a \circ F$ so that

$$\mathrm{D}(\psi \circ F \circ \varphi^{-1})(\varphi(p_0)) = \left.\begin{bmatrix} \frac{\partial F^1}{\partial x^1} & \cdots & \frac{\partial F^1}{\partial x^{n-m}} & \frac{\partial F^1}{\partial x^{n-m+1}} & \cdots & \frac{\partial F^1}{\partial x^n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F^m}{\partial x^1} & \cdots & \frac{\partial F^m}{\partial x^{n-m}} & \frac{\partial F^m}{\partial x^{n-m+1}} & \cdots & \frac{\partial F^m}{\partial x^n} \end{bmatrix}\right|_{\varphi(p_0)},$$

assume that it is the right block

$$\mathrm{D}_2 F(p_0) = \left.\begin{bmatrix} \frac{\partial F^1}{\partial x^{n-m+1}} & \cdots & \frac{\partial F^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F^m}{\partial x^{n-m+1}} & \cdots & \frac{\partial F^m}{\partial x^n} \end{bmatrix}\right|_{\varphi(p_0)}$$

which is nonsingular. Now, define a smooth map $\varphi_f\colon U \cap F^{-1}(V) \to \mathbb{R}^n$ by

$$\varphi_f(p) = (x^1(p), \ldots, x^{n-m}(p), F^{n-m+1}(p) - y^{n-m+1}(q), \ldots, F^n(p) - y^n(q)),$$

and note that $\mathrm{d}(\varphi_f)_{p_0}\colon T_{p_0}M \to \mathbb{R}^n$ is an isomorphism. Indeed, its matrix representation (given in block form)

$$\mathrm{D}(\varphi_f \circ \varphi^{-1})(\varphi(p_0)) = \begin{bmatrix} \mathrm{Id}_{n-m} & 0 \\ * & \mathrm{D}_2 F(p_0) \end{bmatrix}$$

is nonsingular, as $\det \mathrm{D}(\varphi_f \circ \varphi^{-1})(\varphi(p_0)) = \det \mathrm{D}_2 F(p_0) \neq 0$. Corollary 9 now tells us that, reducing U if necessary, $(U \cap F^{-1}(V), \varphi_f)$ is a chart for M around p_0. This is a slice chart adapted to $F^{-1}(q)$ as $F(p) = q$ if and only if the last m components of $\varphi_f(p)$ vanishes, that is, $\varphi_f(U \cap F^{-1}(q)) = \varphi_f(U \cap F^{-1}(V)) \cap (\mathbb{R}^{n-m} \times \{0\})$. We conclude from Theorem 12 that $F^{-1}(q)$ is an embedded $(n-m)$-dimensional submanifold of M.

The proof that $T_p(F^{-1}(q)) = \ker \mathrm{d}F_p$ for each $p \in F^{-1}(q)$ is the same proof given in Theorem 16: just replace f, c, and $\mathrm{d}f_p \colon T_pM \to \mathbb{R}$ with F, q, and $\mathrm{d}F_p \colon T_pM \to T_{F(p)}N$, respectively. □

Let's see some more interesting examples.

Example 84 (The orthogonal group)

Given an integer $n \geq 1$, consider the **orthogonal group**:

$$\mathrm{O}(n) = \{A \in \mathbb{R}^{n\times n} : A^{\mathrm{T}}A = \mathrm{Id}_n\}.$$

It is not hard to check that $\mathrm{O}(n)$ is a group. Here, we will check that $\mathrm{O}(n)$ is an embedded $n(n-1)/2$-dimensional submanifold of $\mathbb{R}^{n\times n}$. The obvious guess of which function we should apply Theorem 17 to, would be $F\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ given by $F(A) = A^{\mathrm{T}}A$, since $\mathrm{O}(n) = F^{-1}(\mathrm{Id}_n)$. If this were to work, however, the dimension of $\mathrm{O}(n)$ would be zero (namely, the difference between the dimensions of the domain and target manifolds of F)—and we know that $\mathrm{O}(n)$ is more than a countable and discrete collection of points. We then adjust the target of F, and consider it as a function $F\colon \mathbb{R}^{n\times n} \to \mathrm{Sym}_n(\mathbb{R})$, where

$$\mathrm{Sym}_n(\mathbb{R}) = \{B \in \mathbb{R}^{n\times n} : B^{\mathrm{T}} = B\} \cong \mathbb{R}^{n(n+1)/2}$$

is the vector subspace of all $n \times n$ real symmetric matrices. We are allowed to do this since $A^{\mathrm{T}}A$ is symmetric for any $A \in \mathbb{R}^{n\times n}$. In any case, now we may show that Id_n is a regular value of F, that is, that the differential $\mathrm{d}F_A\colon \mathbb{R}^{n\times n} \to \mathrm{Sym}_n(\mathbb{R})$ is surjective for every $A \in \mathrm{O}(n)$. Indeed, we have that

$$\mathrm{d}F_A(H) = H^{\mathrm{T}}A + A^{\mathrm{T}}H$$

for every $H \in \mathbb{R}^{n\times n}$, so that for any $B \in \mathrm{Sym}_n(\mathbb{R})$ we may compute

$$\mathrm{d}F_A\left(\frac{AB}{2}\right) = \frac{BA^{\mathrm{T}}}{2}A + A^{\mathrm{T}}\frac{AB}{2} = \frac{B}{2} + \frac{B}{2} = B.$$

Hence, $\mathrm{O}(n)$ is an embedded submanifold of $\mathbb{R}^{n\times n}$, and its dimension is equal to $n^2 - n(n+1)/2 = n(n-1)/2$, as claimed.

Finally, we also see that

$$T_A\mathrm{O}(n) = \{H \in \mathbb{R}^{n\times n} : H^{\mathrm{T}}A + A^{\mathrm{T}}H = 0\}$$

and, in particular, $T_{\mathrm{Id}_n}\mathrm{O}(n) = \{H \in \mathbb{R}^{n\times n} : H^{\mathrm{T}} + H = 0\}$ is the space of all skew-symmetric matrices.

Exercise 94 (The unitary group)

Given an integer $n \geq 1$, show that the **unitary group**

$$\mathrm{U}(n) = \{A \in \mathbb{C}^{n\times n} : A^\dagger A = \mathrm{Id}_n\}$$

is an embedded submanifold of $\mathbb{C}^{n\times n}$. What is its dimension? Determine the tangent space $T_A\mathrm{U}(n)$ and, in particular, $T_{\mathrm{Id}_n}\mathrm{U}(n)$. Here, $A^\dagger = \overline{A}^{\mathrm{T}}$ is the conjugate-transpose of A.

The fact that in Example 84 above we had to restrict the target of F to $\mathrm{Sym}_n(\mathbb{R})$ instead of $\mathbb{R}^{n\times n}$ to make things work suggests that Theorem 17 still has room for improvement. The key word here is "rank": recall from Linear Algebra that if $A \in \mathbb{R}^{m\times n}$ has $\mathrm{rank}(A) = k$, there are invertible matrices $P \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{n\times n}$ such that

$$PAQ = \begin{bmatrix} \mathrm{Id}_k & 0 \\ 0 & 0 \end{bmatrix},$$

where 0 stands for blocks of zeros of the several appropriate sizes. Given a smooth function $F\colon M \to N$, the corresponding rank function $p \mapsto \mathrm{rank}(\mathrm{d}F_p)$ is integer-valued and lower-semicontinuous, implying that for every $p_0 \in M$ there is an open neighborhood $U \subseteq M$ of p_0 such that $\mathrm{rank}(\mathrm{d}F_p) \geq \mathrm{rank}(\mathrm{d}F_{p_0})$ for every $p \in U$. If we think back to the motto "*good things happening to* $\mathrm{d}F_p$ *also happen to* F *itself near* p", the next result—which interpolates between the local forms of immersions and submersions—should not be surprising:

Theorem 18 (The constant-rank theorem)

Let M and N be smooth manifolds with $\dim M = n$ and $\dim N = m$, $F\colon M \to N$ be a smooth function, and $p \in M$ be such that the rank of F is constant and equal to k in some neighborhood of p. Then, there are charts (U, φ) and (V, ψ) centered at p and $F(p)$ for which the local representation

$$\psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \subseteq \mathbb{R}^k \times \mathbb{R}^{n-k} \to \psi(V) \subseteq \mathbb{R}^k \times \mathbb{R}^{m-k}$$

is given by $(\psi \circ F \circ \varphi^{-1})(x, y) = (x, 0)$.

It leads, using similar ideas to the ones discussed here, to:

Theorem 19 (The constant-rank level set theorem)

Let M and N be smooth manifolds, $F\colon M \to N$ be a smooth function, $k \geq 1$ be an integer, and $q \in N$ be such that $\mathrm{rank}(\mathrm{d}F_p) = k$ for all $p \in F^{-1}(q)$. Then, whenever $F^{-1}(q) \neq \varnothing$, it is an embedded submanifold of M with $\dim F^{-1}(q) = \dim M - k$ (i.e., the codimension is k) and, for every $p \in F^{-1}(q)$, the tangent space is given by $T_p(F^{-1}(q)) = \ker \mathrm{d}F_p$.

For proofs of Theorems 18 and 19 see, e.g., [23, Theorems 11.1 and 11.2, p. 116] or [15, Theorems 4.12 and 5.12]. There is one last question to consider: if $F: M \to N$ is a smooth function, and we replace the point $q \in N$ with a submanifold $S \subseteq N$, when is the inverse image $F^{-1}(S)$ a submanifold of M, and what is its dimension? This leads us to concept of *transversality*, and you can read more about it in [8], for instance.

4.7 A brief overview of vector fields and flows

Definition 42 (Vector fields)

Let M be a smooth manifold. A **vector field** is an assignment $\mathbf{X}$, to each point $p \in M$, of a tangent vector $\mathbf{X}_p \in T_pM$. We say that $\mathbf{X}$ is smooth if, for every chart $(U; x^1, \ldots, x^n)$ for M, the functions $X^1, \ldots, X^n : U \to \mathbb{R}$ defined by the relation

$$\mathbf{X}_p = \sum_{i=1}^{n} X^i(p) \frac{\partial}{\partial x^i}\bigg|_p, \qquad \text{for all } p \in U,$$

are smooth; note that $X^i(p) = \mathrm{d}x^i|_p(\mathbf{X}_p)$ for each $i = 1, \ldots, n$. We denote the space of all smooth vector fields on M by $\mathfrak{X}(M)$.

Remark. As usual, note that checking that $\mathbf{X}$ is smooth requires checking that the functions X^i are smooth for enough charts to cover M; maximality of the smooth structure of M then takes care of the remaining charts. In general, a vector field which is not necessarily smooth is usually referred to as a **rough vector field**. Otherwise, we will always assume that vector fields are smooth.

Vector fields are ubiquitous in Calculus, Differential Geometry, and Physics. Moreover, vector fields are the generalization to manifolds of systems of first-order differential equations on open subsets of Euclidean spaces. To make precise sense of this, we need one more definition:

Definition 43 (Integral curves)

Let M be a smooth manifold and $\mathbf{X} \in \mathfrak{X}(M)$ be a vector field. A curve $\alpha: I \to M$ (where $I \subseteq \mathbb{R}$ is an open interval) is called an **integral curve** of $\mathbf{X}$ if $\alpha'(t) = \mathbf{X}_{\alpha(t)}$, for every $t \in I$. See Figure 57.

(The vectors belonging to the vector field in $\mathbf{X}$ are indicated in black, while the integral curves indicated in red "follow" $\mathbf{X}$.)

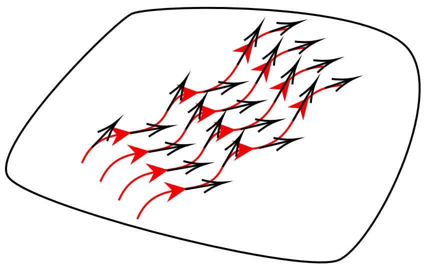


Figure 57: Integral curves of a vector field on a smooth manifold.

Assume, with the setting of the above definition, that the image of α is contained in the domain of a chart $(U; x^1, \ldots, x^n)$ for M, and that we write the coordinates of $\alpha(t)$

as $(x^1(t),\ldots,x^n(t))$. Then, we have that

$$\alpha'(t) = \mathbf{X}_{\alpha(t)} \iff \sum_{i=1}^n (x^i)'(t)\frac{\partial}{\partial x^i}\bigg|_{\alpha(t)} = \sum_{i=1}^n X^i(\alpha(t))\frac{\partial}{\partial x^i}\bigg|_{\alpha(t)}$$

$$\iff \begin{cases} (x^1)'(t) &= X^1(x^1(t),\ldots,x^n(t)) \\ &\vdots \\ (x^n)'(t) &= X^n(x^1(t),\ldots,x^n(t)). \end{cases}$$

This already tells us that, in general, finding integral curves of a given vector field might be a very challenging task. Consider the following concrete example:

Example 85

In $M = \mathbb{R}^2$, consider the vector field

$$\mathbf{X} = x\frac{\partial}{\partial x} + x^2\frac{\partial}{\partial y}.$$

Using the isomorphisms $T_{(x,y)}(\mathbb{R}^2) \cong \mathbb{R}^2$ to write the values of $\mathbf{X}$ as ordered pairs, we have that

$$\mathbf{X}_{(0,0)} = (0,0), \quad \mathbf{X}_{(2,3)} = (2,4), \quad \mathbf{X}_{(-1,0)} = (-1,1), \quad \mathbf{X}_{(-3,1)} = (-3,9),$$

etc.; see Figure 58.

A curve $\alpha\colon I \to \mathbb{R}^2$, written in the form $\alpha(t) = (x(t), y(t))$, is an integral curve of $\mathbf{X}$ if and only if the functions x and y constitute a solution of the system of differential equations

$$\begin{cases} x'(t) = x(t), \\ y'(t) = x(t)^2. \end{cases}$$

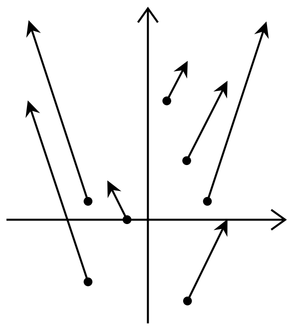


Figure 58: Values of a vector field on the plane.

In this case, imposing initial conditions $x(0) = x_0$ and $y(0) = y_0$, it is not difficult to explicitly obtain the closed form

$$\alpha(t) = \left(x_0 e^t, \frac{x_0^2}{2}(e^{2t} - 1) + y_0\right)$$

for the integral curve α. If the components of $\mathbf{X}$ get more complicated, so will the resulting system for $x(t)$ and $y(t)$ be.

To better understand, from both the geometric and dynamical perspectives, what a vector field $\mathbf{X}$ "is doing", it is convenient to gather all integral curves of $\mathbf{X}$ into a single mapping. The theory of differential equations ensures that this can be done in a smooth way.

Theorem 20 (The fundamental theorem of flows)

Let M be a smooth manifold, and $\mathbf{X}$ be a vector field on M.

(i) For each $p \in M$, there is a maximally-defined[a] integral curve $\alpha_p\colon I_p \to M$ of $\mathbf{X}$ with initial condition $\alpha_p(0) = p$, where $I_p \subseteq \mathbb{R}$ is an open interval with $0 \in I_p$.

(ii) There is an open subset $\mathscr{D} \subseteq \mathbb{R} \times M$ containing $\{0\} \times M$ for which the mapping $\Phi\colon \mathscr{D} \to M$, called the **(local) flow of X** and defined by $\Phi(t,p) = \alpha_p(t)$, is smooth; we also have $\mathscr{D} \cap (\mathbb{R} \times \{p\}) = I_p \times \{p\}$ for each $p \in M$.

(iii) We have that $(t+s,p) \in \mathscr{D}$ whenever $(s,\Phi(t,p)),(t,p) \in \mathscr{D}$, and the relation $\Phi(t+s,p) = \Phi(s,\Phi(t,p))$ holds. In addition, $\Phi(0,p) = p$.

For each $t \in \mathbb{R}$, we set $U_t = \{p \in M : (t,p) \in \mathscr{D}\}$ and $\Phi_t = \Phi(t,\cdot)$.

(iv) For each $t \in \mathbb{R}$, the mapping $\Phi_t\colon U_t \to U_{-t}$ is a diffeomorphism, with inverse explicitly given by $(\Phi_t)^{-1} = \Phi_{-t}$.

The vector field $\mathbf{X}$ is called **complete** if $\mathscr{D} = \mathbb{R} \times M$, that is, if the maximal domain of definition for each of its integral curves is the entire real line $\mathbb{R}$. In this case, Φ is called the **global flow** of $\mathbf{X}$.

[a] This means that if $J \subseteq \mathbb{R}$ is an open interval with $I_p \subseteq J$ and $\beta\colon J \to M$ is an integral curve of $\mathbf{X}$ such that $\beta|_{I_p} = \alpha_p$, then $J = I_p$ and $\beta = \alpha_p$.

For a detailed proof, see [15, Theorem 9.12]. Here, we are interested only in understanding this result and computing some examples. Smoothness in condition (ii) really amounts to saying that the dependence of the integral curve α_p on the initial condition $p \in M$ is also smooth (this is not trivial). The relation given in (iii) says that if we start with a point p, let it flow by time t, and then let the result $\Phi(t,p)$ flow by time s, we obtain the same point as we would be letting the initial point p flow by time $t+s$ to begin with. In particular, $\Phi(0,p) = p$ for every $p \in M$. When the vector field $\mathbf{X}$ is complete, these conditions say that Φ is exactly a group action of $\mathbb{R}$ (equipped with addition) on M: $t \cdot p = \Phi(t,p)$.

Example 86

The vector field $\mathbf{X}$ from Example 85 is complete, and its flow $\Phi\colon \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$\Phi(t,(x,y)) = \left(xe^t, \frac{x^2}{2}(e^{2t}-1)+y\right),$$

for every $(t,(x,y)) \in \mathbb{R} \times \mathbb{R}^2$.

Example 87 (Flows of constant fields)

On $M = \mathbb{R}^n$, consider a single vector $v \in \mathbb{R}^n$ as a constant vector field on $\mathbb{R}^n$, with the aid of the usual isomorphisms $T_p(\mathbb{R}^n) \cong \mathbb{R}^n$. Given $p \in \mathbb{R}^n$, the solution α to the differential equation $\alpha'(t) = v$ with initial condition $\alpha(0) = p$ is clearly given by $\alpha(t) = p + tv$, and its maximal interval of definition is $I_p = \mathbb{R}$. This means that v is complete and its flow $\Phi\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is given by $\Phi(t, p) = p + tv$. See Figure 59.

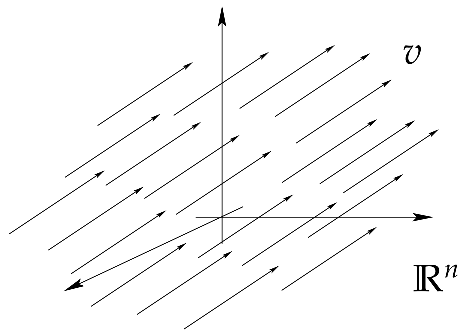


Figure 59: A constant vector field in $\mathbb{R}^n$.

If we replace $\mathbb{R}^n$ with an open subset $U \subseteq \mathbb{R}^n$, the constant vector field may or may not be complete: it depends on whether lines in the direction of v starting at points in U remain forever in U or not.

The simplest manifold one could consider is, of course, the real line $\mathbb{R}$. But even in $\mathbb{R}$, things can get complicated, and vector fields need not be complete.

Example 88 (An incomplete vector field in the real line)

In the manifold $M = \mathbb{R}$, consider the vector field $\mathbf{X} = x^2 \frac{\mathrm{d}}{\mathrm{d}x}$. For each point $x \in \mathbb{R}$, we visualize the tangent vector $\mathbf{X}_x$ as a horizontal arrow of magnitude x^2, always pointing to the right as $x^2 \geq 0$. See Figure 60.

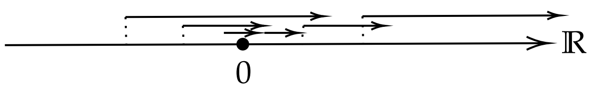


Figure 60: A homogeneous quadratic vector field on the real line.

An integral curve of $\mathbf{X}$ is nothing more than a real-valued function $t \mapsto x(t)$, such that $x'(t) = x(t)^2$. This is a nonlinear ordinary differential equation, but we can solve it explicitly:

$$x'(t) = x(t)^2 \implies \frac{x'(t)}{x(t)^2} = 1 \implies -\frac{1}{x(t)} = t + c \implies x(t) = -\frac{1}{t+c}.$$

If an initial condition $x(0) = x_0$ is given, we necessarily have that $c = -1/x_0$. This means that

$$x(t) = -\frac{1}{t - \frac{1}{x_0}} = \frac{x_0}{1 - tx_0}.$$

The maximal interval of definition of $x(t)$, however, is not the entire real line: we have a problem when $1 - tx_0 = 0$. Renaming x_0 back to x, the flow domain of $\mathbf{X}$ is given by

$$\mathscr{D} = \{(t,x) \in \mathbb{R} \times \mathbb{R} : 1 - tx > 0\},$$

and $\Phi\colon \mathscr{D} \to \mathbb{R}$ is given by

$$\Phi(t,x) = \frac{x}{1-tx}.$$

See the domain $\mathscr{D}$ in Figure 61.

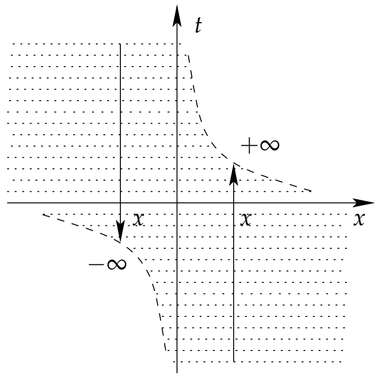


Figure 61: The flow domain for the vector field $\mathbf{X} = x^2 \frac{\mathrm{d}}{\mathrm{d}x}$ on $\mathbb{R}$.

Example 89 (Rotation flow)

Consider in $M = \mathbb{R}^2$ the so-called rotation vector field, given by

$$\mathbf{X} = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}.$$

That is, at each point $(x,y) \in \mathbb{R}^2$, the vector $\mathbf{X}_{(x,y)}$ is obtained by rotating the position vector of (x,y) by $\pi/2$ counterclockwise, and making it start at (x,y) itself. Intuitively, it should be already somewhat clear that the integral curves of $\mathbf{X}$ are circles, cf. Figure 62.

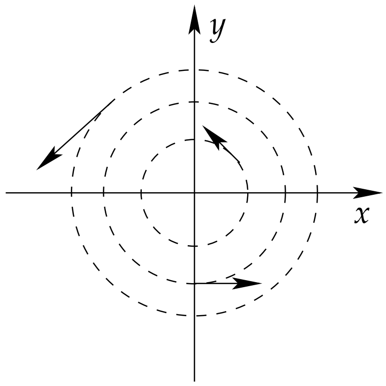


Figure 62: Rotation flow on the plane.

A curve $\alpha(t) = (x(t), y(t))$ is an integral curve of $\mathbf{X}$ if and only if $x'(t) = -y(t)$ and $y'(t) = -x(t)$. If $(x(0), y(0)) = (x_0, y_0)$, we have that

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \implies \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix},$$

so that $\alpha(t) = (x_0 \cos t - y_0 \sin t, x_0 \sin t + y_0 \cos t)$. As (x_0, y_0) is arbitrary and $\alpha(t)$ is defined for all $t \in \mathbb{R}$, we have that $\mathbf{X}$ is complete. To write its flow explicitly, we rename $(x_0, y_0) \mapsto (x, y)$, so that $\Phi\colon \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$\Phi(t,(x,y)) = (x\cos t - y\sin t, x\sin t + y\cos t).$$

Geometrically: given $(x,y) \in \mathbb{R}^2$, the point $\Phi(t,(x,y))$ is obtained by rotating (x,y) by angle t, counterclockwise.

Exercise 95

Is the vector field $\mathbf{X}$ on $\mathbb{R}^2$ given by

$$\mathbf{X} = y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$$

complete? Compute its flow.

Hint: Hyperbolic trigonometric functions are your friends.

Example 90 (Flows of linear vector fields)

Consider the manifold $M = \mathbb{R}^n$, and fix a matrix $A \in \mathbb{R}^{n\times n}$. We can regard it as a linear vector field on $\mathbb{R}^n$, assigning to each $p \in \mathbb{R}^n$ the vector $Ap \in \mathbb{R}^n \cong T_p(\mathbb{R}^n)$, regarded as starting at the point p. The rotation vector field on Example 89 is a particular case of this. A curve $\alpha(t)$ in $\mathbb{R}^n$ is an integral curve of A if and only if $\alpha'(t) = A\alpha(t)$. Once an initial condition $\alpha(0) = p$ is given, we know from the theory of ordinary differential equations that $\alpha(t) = \mathrm{e}^{tA}p$, where

$$\mathrm{e}^{tA} = \sum_{k\geq 0} \frac{(tA)^k}{k!}$$

is the exponential of tA—the series converges absolutely (relative to any norm on $\mathbb{R}^{n\times n}$), defines a smooth function, and e^{tA} commutes with A. In particular, the series converges for all $t \in \mathbb{R}$, making A a complete vector field, and the flow $\Phi\colon \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}^n$ is given by $\Phi(t,p) = \mathrm{e}^{tA}p$.

Example 91 (The simplest radial flow)

In the manifold $M = \mathbb{R}^n \setminus \{0\}$, we consider the radial vector field $\mathbf{X}$ given by $\mathbf{X}_p = p/\|p\|$. It is called radial because for each p, the value $\mathbf{X}_p$ is a multiple of p, cf. Figure 63. Geometrically, it is not hard to see that the integral curves of $\mathbf{X}$ are the (open) rays starting from the origin. What is not immediately clear is how these rays are parametrized. So, we try for integral curves of the form $\alpha(t) = f(t)p$, for some positive function f having $f(0) = 1$. From the condition $\alpha'(t) = \mathbf{X}_{\alpha(t)}$, we have that

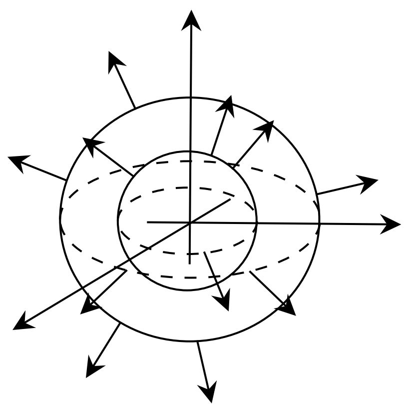


Figure 63: A unit radial vector field on $\mathbb{R}^n \setminus \{0\}$.

$$f'(t)p = \frac{f(t)p}{\|f(t)p\|} = \frac{p}{\|p\|} \implies f'(t) = \frac{1}{\|p\|} \implies f(t) = \frac{t}{\|p\|} + c,$$

and so $f(0) = 1$ implies that $c = 1$. Hence we may write $\alpha(t) = (1 + t/\|p\|)p$. The maximal interval of definition of α, however, is not the entire real line: α cannot cross the origin, and so the multiple $1 + t/\|p\|$ has to stay positive. This means that the flow domain of $\mathbf{X}$ is $\mathcal{D} = \{(t, p) \in \mathbb{R} \times (\mathbb{R}^n \smallsetminus \{0\}) : t > -\|p\|\}$, and the flow $\Phi\colon \mathcal{D} \to \mathbb{R}^n \smallsetminus \{0\}$ is given by $\Phi(t, p) = (1 + t/\|p\|)p$.

Exercise 96

With the setting of the previous example, compute the flow of the vector field $\mathbf{X}$ given by $\mathbf{X}_p = p/\|p\|^k$, where $k \geq 1$ is a fixed odd integer, and write its flow domain $\mathcal{D}$ explicitly. Is $\mathbf{X}$ complete? What happens differently if k is even?

In general, determining whether a vector field is complete or not without actually computing its flow can be very difficult. There are a few results in this direction, here is the most friendly of them:

Proposition 33

Let M be a smooth manifold, and $\mathbf{X}$ be a vector field on M whose flow domain $\mathcal{D}$ contains a strip $(-\varepsilon, \varepsilon) \times M$. Then, $\mathcal{D} = \mathbb{R} \times M$ (i.e., $\mathbf{X}$ is complete).

Proof: Let $p \in M$ and $\alpha\colon (a, b) \to M$ be a maximally-defined integral curve of $\mathbf{X}$, with $0 \in (a, b)$ and $\alpha(0) = p$, and assume by contradiction that $b < \infty$. Then, if $\beta\colon (-\varepsilon, \varepsilon) \to M$ is an integral curve of $\mathbf{X}$ with $\beta(0) = \alpha(b - \varepsilon/2)$ (we do not assume that β is maximally-defined), consider $\widetilde{\alpha}\colon (a, b + \varepsilon/2) \to M$ given by

$$\widetilde{\alpha}(t) = \begin{cases} \alpha(t), & \text{if } t \in (a, b), \\ \beta\left(t - b + \frac{\varepsilon}{2}\right), & \text{if } t \in \left(b - \varepsilon, b + \frac{\varepsilon}{2}\right). \end{cases}$$

Since $\alpha(t) = \beta(t - b + \varepsilon/2)$ on the intersection $(b - \varepsilon, b)$ (they are both integral curves of $\mathbf{X}$ which agree at $t = b - \varepsilon/2$), $\widetilde{\alpha}$ is a well-defined, smooth, integral curve of $\mathbf{X}$ starting at $t = 0$ and with domain strictly larger than the domain of α. This contradicts the maximality of α, and shows that $b = \infty$. A similar argument shows that $a = -\infty$, and so α must be defined on all of $\mathbb{R}$, as required. ☐

Corollary 10

Every compactly-supported vector field on a smooth manifold is complete.

Corollary 11

Every vector field on a compact manifold is complete.

Example 92 (Vector fields on $\mathbb{RP}^n$ induced from ones on $\mathbb{S}^n$)

Consider the real projective space $\mathbb{RP}^n$, but regarded as the quotient $\mathbb{S}^n/\mathbb{Z}_2$, where $\mathbb{Z}_2 = \{\mathrm{Id}_{\mathbb{S}^n}, \tau\}$ for the antipodal mapping $\tau\colon \mathbb{S}^n \to \mathbb{S}^n$, cf. the remark after Theorem 9. When does a vector field $\mathbf{X}$ on $\mathbb{S}^n$ survive on the quotient $\mathbb{RP}^n$? It seems natural to expect that if such a vector field $\widetilde{\mathbf{X}} \in \mathfrak{X}(\mathbb{RP}^n)$ exists, it should satisfy the relation $\mathrm{d}\pi_p(\mathbf{X}_p) = \widetilde{\mathbf{X}}_{\{p,-p\}}$, for all $p \in \mathbb{S}^n$. At the same time, replacing p with $-p$ leads to $\mathrm{d}\pi_{-p}(\mathbf{X}_{-p}) = \widetilde{\mathbf{X}}_{\{p,-p\}}$, meaning that the equality

$$\mathrm{d}\pi_{-p}(\mathbf{X}_{-p}) = \mathrm{d}\pi_p(\mathbf{X}_p), \quad \text{for all } p \in \mathbb{S}^n, \tag{4.19}$$

is a necessary condition for the existence of $\widetilde{\mathbf{X}}$. In terms of the antipodal mapping, (4.19) is equivalent to saying that

$$\mathrm{d}\tau_p(\mathbf{X}_p) = \mathbf{X}_{-p}, \quad \text{for all } p \in \mathbb{S}^n. \tag{4.20}$$

Indeed, applying $\mathrm{d}\pi_{-p}$ to both sides of (4.20) and using the chain rule together with the relation $\pi \circ \tau = \pi$ leads to (4.19); conversely, (4.19) together with the chain rule and relation $\pi \circ \tau = \pi$ leads to $\mathbf{X}_{-p} - \mathrm{d}\tau_p(\mathbf{X}_p) \in \ker \mathrm{d}\pi_{-p} = \{0\}$, so that (4.20) holds (here we use Exercise 80).

Now, is (4.19)–(4.20) also sufficient for the existence of $\widetilde{\mathbf{X}}$? Yes! Given any element $\{p,-p\} \in \mathbb{RP}^n$, we *define* $\widetilde{\mathbf{X}}$ by $\widetilde{\mathbf{X}}_{\{p,-p\}} = \mathrm{d}\pi_p(\mathbf{X}_p)$. Choosing $-p$ instead of p leads to the same element of $T_{\{p,-p\}}(\mathbb{RP}^n)$ due to (4.19), so that $\widetilde{\mathbf{X}}$ is well-defined. Smoothness of $\widetilde{\mathbf{X}}$ follows from the one of $\mathbf{X}$ as π is a local diffeomorrphism. (In fact, π being a surjective submersion is enough, due to Corollary 8.)

The above situation motivates the next definition:

Definition 44 (F-related fields)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a smooth mapping. Two vector fields $\mathbf{X} \in \mathfrak{X}(M)$ and $\mathbf{Y} \in \mathfrak{X}(N)$ are said to be F**-related** if $\mathrm{d}F_p(\mathbf{X}_p) = \mathbf{Y}_{F(p)}$, for every $p \in M$.

Exercise 97

Let M be a smooth manifold, and Γ be a finite group of diffeomorphisms acting freely on M, in the sense of Exercise 68, so that M/Γ becomes a smooth manifold and $\pi\colon M \to M/\Gamma$ is a local diffeomorphism. Let $\mathbf{X} \in \mathfrak{X}(M)$ be a vector field such that $\mathbf{X}$ is γ-related to itself, for every $\gamma \in \Gamma$. Show that there is a unique vector field $\widetilde{\mathbf{X}} \in \mathfrak{X}(M/\Gamma)$, automatically smooth, which is π-related to $\mathbf{X}$.

Hint: Revisit Example 92.

The flows of F-related vector fields are also related, in the sense of the next result:

Proposition 34

Let M and N be smooth manifolds, and $F\colon M \to N$ be a smooth mapping. The flows $\Phi_{\mathbf{X}}\colon \mathscr{D}_{\mathbf{X}} \subseteq \mathbb{R} \times M \to M$ and $\Phi_{\mathbf{Y}}\colon \mathscr{D}_{\mathbf{Y}} \subseteq \mathbb{R} \times N \to N$ of two F-related vector fields $\mathbf{X} \in \mathfrak{X}(M)$ and $\mathbf{Y} \in \mathfrak{X}(N)$ are related via

$$F(\Phi_{\mathbf{X}}(t,p)) = \Phi_{\mathbf{Y}}(t, F(p)),$$

for all $(t,p) \in \mathscr{D}_{\mathbf{X}}$.

Proof: We just need to show that, fixed $p \in M$, the curve $t \mapsto F(\Phi_{\mathbf{X}}(t,p))$ is an integral curve of $\mathbf{Y}$, starting at $F(p)$ for $t = 0$. The latter condition is clear: $F(\Phi_{\mathbf{X}}(0,p)) = F(p)$, since $\Phi_{\mathbf{X}}(0,p) = p$. As for the former, we compute using the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}t}F(\Phi_{\mathbf{X}}(t,p)) = \mathrm{d}F_{\Phi_{\mathbf{X}}(t,p)}\left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\mathbf{X}}(t,p)\right) = \mathrm{d}F_{\Phi_{\mathbf{X}}(t,p)}\left(\mathbf{X}_{\Phi_{\mathbf{X}}(t,p)}\right) = \mathbf{Y}_{F(\Phi_{\mathbf{X}}(t,p))}.$$

□

We conclude this section with one last concept which plays a central role in the study of vector fields and flows—you will certainly come across it if you continue your studies in differential geometry.

Exercise 98 (A crash course on Lie brackets)

Let M be a smooth manifold, and let $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M)$.

(a) Show that for each $p \in M$, the expression

$$[\mathbf{X},\mathbf{Y}]_p[f] = \mathbf{X}_p[\mathbf{Y}(f)] - \mathbf{Y}_p[\mathbf{X}(f)],$$

where $[\mathbf{Y}(f)]$ denotes the germ at p of the function $x \mapsto \mathbf{Y}_x[f]$, defines an element of $\mathrm{Der}(\mathscr{C}^\infty_p(M), \delta_p) = T_pM$; the resulting vector field $[\mathbf{X},\mathbf{Y}]$ is called the **Lie bracket** of $\mathbf{X}$ and $\mathbf{Y}$.

Note: the assignment $[f] \mapsto \mathbf{X}_p[\mathbf{Y}(f)]$ alone (i.e., composition of vector fields) does not define a tangent vector — it contains "second derivatives". This is why the skew-symmetrization happens, so they cancel. It is similar to a "commutator" of matrices, and it has $[\mathbf{Y},\mathbf{X}] = -[\mathbf{X},\mathbf{Y}]$.

(b) Show that whenever $(U, (x^1, \ldots, x^n))$ is a chart for M and we write

$$\mathbf{X} = \sum_{i=1}^n X^i \frac{\partial}{\partial x^i} \quad \text{and} \quad \mathbf{Y} = \sum_{i=1}^n Y^i \frac{\partial}{\partial x^i},$$

the Lie bracket is given by

$$[\mathbf{X},\mathbf{Y}] = \sum_{i,j=1}^n \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}\right) \frac{\partial}{\partial x^i}.$$

(c) Show that $[f\mathbf{X}, g\mathbf{Y}] = fg[\mathbf{X}, \mathbf{Y}] + f\mathbf{X}[g]\mathbf{Y} - g\mathbf{Y}[f]\mathbf{X}$ for all $f, g \in C^\infty(M)$.

Note: The Lie bracket is not bilinear over $C^\infty(M)$ (i.e., it is not a tensor), just over the scalar field $\mathbb{R}$ — it instead picks up an error term which is a linear combination of $\mathbf{X}$ and $\mathbf{Y}$. You can even use (c) together with the fact that coordinate vectors commute to prove (b) out of order if you want.

(d) In $M = \mathbb{R}^3$, given two smooth functions $f, g\colon \mathbb{R}^2 \to \mathbb{R}$, compute the Lie bracket of

$$\mathbf{X} = \frac{\partial}{\partial x} + f(x,y)\frac{\partial}{\partial z} \quad \text{and} \quad \mathbf{Y} = \frac{\partial}{\partial y} + g(x,y)\frac{\partial}{\partial z}.$$

What is the condition necessary and sufficient for $[\boldsymbol{X}, \boldsymbol{Y}] = 0$? What does it mean, geometrically?

(e) Let N be a second smooth manifold, with vector fields $\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}} \in \mathfrak{X}(N)$, and $F\colon M \to N$ be a smooth mapping. Show that if $\mathbf{X}$ and $\mathbf{Y}$ are F-related to $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$, respectively, then $[\mathbf{X}, \mathbf{Y}]$ is F-related to $[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]$.

We move on to the final part of the course, whose goal is to establish Stokes's Theorem: the ultimate generalization of the Fundamental Theorem of Calculus.

5 Integration on manifolds and Stokes's theorem

5.1 Some exterior algebra

In vector calculus, we deal mostly with two types of objects: scalar fields (that is, real-valued functions) and vector fields. There are multiple ways we can multiply them:

- the product of two scalar fields is a scalar field: $(fg)(p) = f(p)g(p)$.
- the product of a scalar field and a vector field is a vector field: $(f\mathbf{F})(p) = f(p)\mathbf{F}(p)$.
- the dot product of two vector fields is a scalar field: $(\mathbf{F}\cdot\mathbf{G})(p) = \mathbf{F}(p)\cdot\mathbf{G}(p)$.
- the cross product of two vector fields is a vector field: $(\mathbf{F}\times\mathbf{G})(p) = \mathbf{F}(p)\times\mathbf{G}(p)$.

In a similar manner, there are three main differential operators to consider:

- the gradient of a scalar field is a vector field: $\nabla f(p) = (f_x, f_y, f_z)|_p$.
- the divergence of a vector field is a function: $(\nabla\cdot\mathbf{F})(p) = ((F_1)_x + (F_2)_y + (F_3)_z)|_p$.
- the curl of a vector field is a vector field:

$$(\nabla\times\mathbf{F})(p) = ((F_3)_y - (F_2)_z, (F_1)_z - (F_3)_x, (F_2)_x - (F_1)_y)|_p.$$

We want to develop a formalism which allows us to treat all these products and differential operators at the same time. In other words, we want to consider scalar and vector fields as particular examples of a larger class of objects, for which a general product $\wedge$ and differential operator d exist, and reduce to the ones described above when restricted to scalar and vector fields. The remaining part of vector calculus, however, deals with line and surface integrals. While everything can be made formal with Riemann sums, here is the gist of it for surface integrals: if $M \subseteq \mathbb{R}^3$ is a regular surface, the area of M is computed by adding the areas $\mathrm{d}A$ of infinitesimal parallelograms in each tangent plane to M, cf. Figure 64 below.

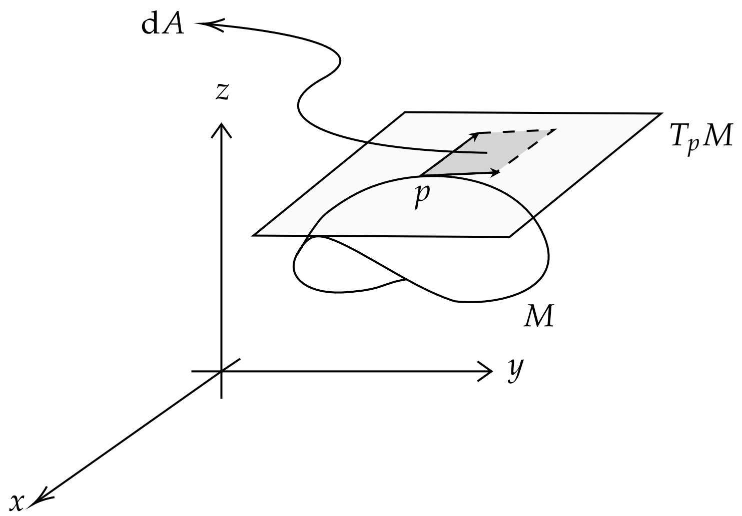


Figure 64: Infinitesimal areas of parallelograms in the tangent planes to M.

Note that if $\mathrm{A}(v, w)$ denotes the area of the parallelogram spanned by two vectors v, w, then we have the relations

$$\mathrm{A}(\lambda v, w) = \lambda \mathrm{A}(v, w) \quad \text{and} \quad \mathrm{A}(v_1 + v_2, w) = \mathrm{A}(v_1, w) + \mathrm{A}(v_2, w), \tag{5.1}$$

and another similar one in relative to the w-argument; see Figure 65.

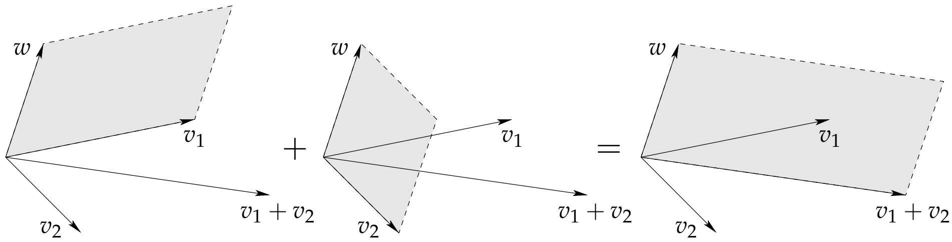


Figure 65: The linear relation $\mathrm{A}(v_1 + v_2, w) = \mathrm{A}(v_1, w) + \mathrm{A}(v_2, w)$.

In higher dimensional manifolds, we still want to add up infinitesimal volumes of parallelepipeds on each tangent space. The problem is that, abstractly, there is no device in an abstract vector space which can be used to compute volumes. Still, if we had one, it would have to be multilinear, just like our area function A above is. Here, we apply for the first time a general philosophy:

> When trying to understand some concept about smooth manifolds, which does not involve derivatives and concerns mainly tangent vectors, we forget about the topology and return to linear algebra: once we finally make sense of whatever it is in a single abstract vector space, we return to manifolds and apply our understanding to all tangent spaces simultaneously. (5.2)

For this reason, we start with the following definition:

Definition 45 (Covariant tensors)

Let V be a real vector space. A **(covariant)** k**-tensor** on V is a multilinear mapping $T\colon V^{\times k} = V \times \cdots \times V \to \mathbb{R}$, that is, whenever any $k-1$ of its arguments are fixed, we obtain a linear functional. In other words, given $v_1, \ldots, v_k \in V$, for each $i = 1, \ldots, n$ we have that $T(v_1, \ldots, v_{i-1}, \_, v_{i+1}, \ldots, v_k) : V \to \mathbb{R}$ is a linear functional. The set of all k-tensors in V will be denoted by $\mathcal{T}^k(V)$; it is clearly a vector space with pointwise operations.

Remark. When $k = 0$, we set $\mathcal{T}^0(V) = \mathbb{R}$. Next, $\mathcal{T}^1(V) = V^*$. Then, $\mathcal{T}^2(V)$ is the space of all bilinear forms $V \times V \to \mathbb{R}$. And so on.

To exhibit a basis for $\mathcal{T}^k(V)$ when V is finite-dimensional, and to ultimately arrive at the product $\wedge$ mentioned earlier, we need an additional operation between tensors.

Definition 46 (Tensor product between covariant tensors)

Let V be a real vector space, and fix tensors $T \in \mathcal{T}^k(V)$ and $S \in \mathcal{T}^\ell(V)$. The **tensor product** between T and S is the element $T \otimes S \in \mathcal{T}^{k+\ell}(V)$ defined by

$$(T \otimes S)(v_1, \ldots, v_{k+\ell}) = T(v_1, \ldots, v_k)S(v_{k+1}, \ldots, v_\ell),$$

for all $v_1, \ldots, v_{k+\ell} \in V$; note that $T \otimes S$ is multilinear since both T and S are.

The above definition is natural: T eats k vectors and S eats ℓ vectors, so how can $T \otimes S$ eat $k + \ell$ vectors? We feed the first k to T, and the last ℓ to S, in order. For example, if $f, g \in V^*$, we have that $f \otimes g \in \mathcal{T}^2(V)$ is given by $(f \otimes g)(v, w) = f(v)g(w)$. Here are some properties of the tensor product operation:

Exercise 99 (Properties of $\otimes$)

Let V be a real vector space. Show that:

(a) $\otimes \colon \mathcal{T}^k(V) \times \mathcal{T}^\ell(V) \to \mathcal{T}^{k+\ell}(V)$ is bilinear, that is, whenever $T, T_1, T_2 \in \mathcal{T}^k(V)$ and $S, S_1, S_2 \in \mathcal{T}^\ell(V)$, and $\lambda \in \mathbb{R}$, we have that

$$\begin{aligned}(T_1 + \lambda T_2) \otimes S &= T_1 \otimes S + \lambda T_2 \otimes S,\\ T \otimes (S_1 + \lambda S_2) &= T \otimes S_1 + \lambda T \otimes S_2.\end{aligned}$$

(b) $\otimes$ is associative, that is, $(T \otimes S) \otimes R = T \otimes (S \otimes R)$ for all tensors $T \in \mathcal{T}^k(V)$, $S \in \mathcal{T}^\ell(V)$, and $R \in \mathcal{T}^r(V)$.

(c) in general, $T \otimes S \neq S \otimes T$.

As a consequence of associativity, parentheses are no longer needed when taking tensor products of several elements at once. For example, if $f^1, \ldots, f^k \in V^*$, we have that $f^1 \otimes \cdots \otimes f^k \in \mathcal{T}^k(V)$ is given by

$$(f^1 \otimes \cdots \otimes f^k)(v_1, \ldots, v_k) = f^1(v_1) \cdots f^k(v_k)$$

for all $v_1, \ldots, v_k \in V$. This is exactly what we need to build a basis for $\mathcal{T}^k(V)$ from a basis of V when V is finite-dimensional:

Proposition 35 (A basis for $\mathcal{T}^k(V)$)

Let V be a real vector space, and $(e_1, \ldots, e_n)$ be a basis for V. Then, if $(\varphi^1, \ldots, \varphi^n)$ is its dual basis in V^*, we have that

$$\{\varphi^{i_1} \otimes \cdots \otimes \varphi^{i_k} : i_1, \ldots, i_k = 1, \ldots, n\} \tag{5.3}$$

is a basis for $\mathcal{T}^k(V)$. In particular, $\dim \mathcal{T}^k(V) = n^k$.

Proof: The argument is morally the same one used in the case where $k = 1$, that is, when showing that the dual basis is indeed a basis. One observes that if $\sum_{i=1}^n a_i\varphi^i = 0$, evaluating both sides at e_j yields $a_j = 0$ (establishing linear independence of $\varphi^1, \ldots, \varphi^n$), and that $v = \sum_{i=1}^n \varphi^i(v)e_i$ for every $v \in V$ implies that $f = \sum_{i=1}^n f(e_i)\varphi^i$ for every $f \in V^*$ (so that $\varphi^1, \ldots, \varphi^n$ spans V^*).

With that said, we first show that (5.3) is linearly independent. Given n^k real coefficients $a_{i_1\cdots i_k} \in \mathbb{R}$, assume that

$$\sum_{i_1\ldots i_k=1}^n a_{i_1\ldots i_k}\varphi^{i_1}\otimes\cdots\otimes\varphi^{i_k} = 0.$$

Evaluating it at the k-tuple $(e_{j_1}, \ldots, e_{j_k})$, we obtain that

$$\begin{aligned} 0 &= \sum_{i_1,\ldots,i_k=1}^n a_{i_1\ldots i_k}(\varphi^{i_1}\otimes\cdots\otimes\varphi^{i_k})(e_{j_1},\ldots,e_{j_k}) = \sum_{i_1,\ldots,i_k=1}^n a_{i_1\ldots i_k}\varphi^{i_1}(e_{j_1})\cdots\varphi^{i_k}(e_{j_k}) \\ &= \sum_{i_1,\ldots,i_k=1}^n a_{i_1\ldots i_k}\delta^{i_1}_{j_1}\cdots\delta^{i_k}_{j_k} = a_{j_1\ldots j_k}. \end{aligned}$$

As the indices $j_1, \ldots, j_k = 1, \ldots, n$ are also arbitrary, this yields the desired conclusion.

It remains to show that (5.3) spans $\mathcal{T}^k(V)$. So, let $T \in \mathcal{T}^k(V)$, and $v_1, \ldots, v_k \in V$ be arbitrary. We simply compute, using multilinearity of T:

$$\begin{aligned} T(v_1,\ldots,v_k) &= T\left(\sum_{i_1=1}^n \varphi^{i_1}(v_1)e_{i_1}, \cdots, \sum_{i_k=1}^n \varphi^{i_k}(v_k)e_{i_k}\right) \\ &= \sum_{i_1,\ldots,i_k=1}^n \varphi^{i_1}(v_1)\cdots\varphi^{i_k}(v_k)T(e_{i_1},\ldots,e_{i_k}) \\ &\overset{(*)}{=} \sum_{i_1\ldots,i_k=1}^n T_{i_1\ldots i_k}\varphi^{i_1}(v_1)\cdots\varphi^{i_k}(v_k) \\ &= \sum_{i_1,\ldots,i_k=1}^n T_{i_1\ldots i_k}(\varphi^{i_1}\otimes\cdots\otimes\varphi^{i_k})(v_1,\ldots,v_k), \end{aligned}$$

where in $(*)$ we define $T_{i_1\ldots i_k} = T(e_{i_1}, \ldots, e_{i_k})$—the n^k **components** of the tensor T relative to the basis $e_1, \ldots, e_n$ of V. This shows that

$$T = \sum_{i_1,\ldots,i_k=1}^n T_{i_1\ldots i_k}\varphi^{i_1}\otimes\cdots\otimes\varphi^{i_k}, \tag{5.4}$$

concluding the proof. □

Formula (5.4) in the above proof is particularly important: it actually tells us how to write a given tensor as a linear combination of the tensor products of basis covectors.

Example 93

If $T \in \mathcal{T}^2(\mathbb{R}^3)$ is given by $T((x_1, y_1, z_1), (x_2, y_2, z_2)) = 2x_1y_2 - x_1y_3 + 5x_2y_2$ and we let (e_1, e_2, e_3) denote the standard basis of $\mathbb{R}^3$, and $(\varphi^1, \varphi^2, \varphi^3)$ be its dual basis, then we compute $T(e_1, e_2) = 2$, $T(e_1, e_3) = -1$, and $T(e_2, e_2) = 5$, while we have $T(e_i, e_j) = 0$ for all choices of i and j not previously listed. These are the coefficients needed to write T as a linear combination of $\{\varphi^i \otimes \varphi^j : i, j = 1, 2, 3\}$. Namely, we have $T = 2\varphi^1 \otimes \varphi^2 - \varphi^1 \otimes \varphi^3 + 5\varphi^2 \otimes \varphi^2$.

Exercise 100

If $T \in \mathcal{T}^2(\mathbb{R}^2)$ is given by $T((x_1, y_1), (x_2, y_2)) = 3x_1y_1 + 5x_2y_1 - 2x_1y_2 - 3x_2y_2$, write T as a linear combination of $\{\varphi^i \otimes \varphi^j : i, j = 1, 2\}$, where (φ^1, φ^2) is the dual basis of the standard basis of $\mathbb{R}^2$.

Consider again the discussion on areas used to motivate Definition 45. In addition to the properties given in (5.1), there is a third one which should be taken into account: sensitivity to orientation, $\mathrm{A}(w, v) = -\mathrm{A}(v, w)$. In terms of higher-dimensional volumes, thinking that a positively-oriented basis of $\mathbb{R}^n$ should give rise to a parallelepiped having positive volume, and a negatively-oriented basis gives a negative volume, the natural conclusion is that switching the order of two vectors switchs the sign of the resulting oriented volume, since this process makes a positively-oriented basis negatively oriented, and vice-versa.

To formulate a more elegant definition from the above, note that if V is a real vector space, for each integer $k \geq 1$ we have an action of the **permutation group**[9] S_k on $\mathcal{T}^k(V)$: given $\sigma \in \mathrm{S}_k$ and $T \in \mathcal{T}^k(V)$, the element $\sigma T \in \mathcal{T}^k(V)$ is defined by

$$(\sigma T)(v_1, \ldots, v_k) = T(v_{\sigma(1)}, \ldots, v_{\sigma(k)}),$$

for all $v_1, \ldots, v_k \in V$. As the group S_k is generated by **transpositions** (i.e., permutations switching only two elements), the following definition captures what we need:

Definition 47 (Alternating tensors)

Let V be a real vector space. A tensor $T \in \mathcal{T}^k(V)$ is called **alternating** if, for every $\sigma \in \mathrm{S}_k$, we have that $\sigma T = (\operatorname{sgn} \sigma)T$. Explicitly, this means that

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

for all $v_1, \ldots, v_k \in V$. We set $\mathcal{A}^k(V) = \{T \in \mathcal{T}^k(V) : T \text{ is alternating}\}$.

[9] That is, the set of all bijections $\{1, \ldots, k\} \to \{1, \ldots, k\}$, with composition of functions as its group operation. More generally, for any set X, we have the group $\mathrm{S}(X)$ of all bijections $X \to X$. When $|X| = k$, fixing a bijection $\varphi\colon X \to \{1, \ldots, k\}$ yields an isomorphism $\Phi\colon \mathrm{S}(X) \to \mathrm{S}_k$, given by $\Phi(\alpha) = \varphi\alpha\varphi^{-1}$. If $\psi\colon X \to \{1, \ldots, k\}$ is another bijection, inducing $\Psi\colon \mathrm{S}(X) \to \mathrm{S}_k$, the "transition" $\Phi \circ \Psi^{-1}\colon \mathrm{S}_k \to \mathrm{S}_k$ is simply conjugation by the element $\varphi \circ \psi^{-1} \in \mathrm{S}_k$. Think that X and $\mathrm{S}(X)$ are abstract (like manifolds) while $\{1, \ldots, k\}$ and S_k are concrete models (like $\mathbb{R}^n$), while φ and ψ are like "charts".

Above, the **sign** $\operatorname{sgn}\sigma \in \{+1,-1\}$ of an element $\sigma \in S_k$ is usually defined as the number of transpositions used in some factorization of σ as a product of transpositions; the factorization is not unique in general, but the parity of the number of transpositions used is well-defined. Alternatively, you can think of $\sigma \in S_k$ as a linear transformation $\sigma\colon \mathbb{R}^k \to \mathbb{R}^k$, given by $\sigma(x_1,\ldots,x_k) = (x_{\sigma(1)},\ldots,x_{\sigma(k)})$, and then $\operatorname{sgn}\sigma$ is just the determinant $\det\sigma$. It is also helpful to think of it as $\operatorname{sgn}\sigma = (-1)^m$, where m is whichever number of transpositions you manage to use to shuffle $(x_{\sigma(1)},\ldots,x_{\sigma(k)})$ back to the original arrangement $(x_1,\ldots,x_k)$. For example, if $\sigma \in S_4$ is given by $(\sigma(1),\sigma(2),\sigma(3),\sigma(4)) = (3,4,2,1)$, we have that

$$(3,4,2,1) \xrightarrow{1} (1,4,2,3) \xrightarrow{2} (1,2,3,4),$$

so $\operatorname{sgn}\sigma = (-1)^2 = +1$. One of the main properties of sgn is that it defines a homomorphism $S_k \to \{+1,-1\}$, that is, $\operatorname{sgn}(\sigma\tau) = (\operatorname{sgn}\sigma)(\operatorname{sgn}\tau)$ for every $\sigma,\tau \in S_k$

The tensor product of alternating tensors in not alternating as well, in general. Fortunately, there is a process to convert any k-tensor into an alternating k-tensor:

Definition 48 (The alternator operator)

Let V be a real vector space. The **alternator** of a tensor $T \in \mathcal{T}^k(V)$ is the tensor $\operatorname{Alt}(T)$ defined by $\operatorname{Alt}(T) = \sum_{\sigma\in S_k}(\operatorname{sgn}\sigma)\sigma T$

We will see ahead in Proposition 36 that $\operatorname{Alt}(T)$ is indeed an alternating tensor, i.e., $\operatorname{Alt}(T) \in \mathcal{A}^k(V)$. But first, an example to build intuition:

Example 94 (Low-dimensional alternators)

For $k=1$, every $T \in \mathcal{T}^1(V)$ is alternating by default, as $\operatorname{Alt}(T) = T$.

For $k=2$, $\operatorname{Alt}\colon \mathcal{T}^2(V) \to \mathcal{A}^2(V)$ is given by $\operatorname{Alt}(T)(v_1,v_2) = T(v_1,v_2) - T(v_2,v_1)$.

For $k=3$, we compute $\operatorname{Alt}\colon \mathcal{T}^3(V) \to \mathcal{A}^3(V)$ with the aid of the table on the right, listing $\operatorname{sgn}\sigma$ for all $\sigma \in S_3$:

$\operatorname{sgn}\sigma$	$\sigma(1)$	$\sigma(2)$	$\sigma(3)$
+	1	2	3
-	1	3	2
-	2	1	3
+	2	3	1
+	3	1	2
-	3	2	1

$$\begin{aligned}\operatorname{Alt}(T)(v_1,v_2,v_3) = &+T(v_1,v_2,v_3) - T(v_1,v_3,v_2)\\ &- T(v_2,v_1,v_3) + T(v_2,v_3,v_1)\\ &+ T(v_3,v_1,v_2) - T(v_3,v_2,v_1).\end{aligned}$$

For $k=4$, we have $4! = 24$ terms, and we'll hardly have any use for the resulting explicit formula for $\operatorname{Alt}(T)(v_1,v_2,v_3,v_4)$.

Proposition 36 ($\operatorname{Alt}(T)$ is alternating)

For any $T \in \mathcal{T}^k(V)$, we indeed have that $\operatorname{Alt}(T) \in \mathcal{A}^k(V)$.

Proof: Let $\tau \in \mathsf{S}_k$ be arbitrary, and compute

$$\begin{aligned}\tau \mathrm{Alt}(T) &= \tau\left(\sum_{\sigma\in\mathsf{S}_k}(\operatorname{sgn}\sigma)\,\sigma T\right) = \sum_{\sigma\in\mathsf{S}_k}(\operatorname{sgn}\sigma)\,\tau(\sigma T) \\ &= \sum_{\sigma\in\mathsf{S}_k}(\operatorname{sgn}\sigma)\,(\tau\sigma)T = (\operatorname{sgn}\tau)\sum_{\sigma\in\mathsf{S}_k}(\operatorname{sgn}\tau\sigma)(\tau\sigma)T \\ &\stackrel{(*)}{=} (\operatorname{sgn}\tau)\mathrm{Alt}(T),\end{aligned}$$

as required. In $(*)$ we use a general principle: as $\mathsf{S}_k \ni \sigma \mapsto \tau\sigma \in \mathsf{S}_k$ is a bijection, we have that $\sum_{\sigma\in\mathsf{S}_k} F(\tau\sigma) = \sum_{\sigma\in\mathsf{S}_k} F(\sigma)$ for any function $F\colon \mathsf{S}_k \to \mathbb{R}$; as σ ranges over S_k, so does the product $\tau\sigma$, and without repetitions. Think of it as relabeling dummy indices, or as doing a change of variables. □

The operator Alt has another relevant property, which is worth registering:

Proposition 37 (Alt is almost a projection)

Let V be a real vector space. Then, if $T \in \mathcal{T}^k(V)$ is any tensor, we have that $\mathrm{Alt}(\mathrm{Alt}(T)) = k!\mathrm{Alt}(T)$. In particular, $\mathrm{Alt}(T) = k!T$ when T is already alternating.

Remark. If not for the factor of $k!$, the mapping $\mathrm{Alt}\colon \mathcal{T}^k(V) \to \mathcal{T}^k(V)$ would be a projection[10] onto the subspace $\mathcal{A}^k(V)$ of $\mathcal{T}^k(V)$. In fact, many people define the Alt operator as $\mathrm{Alt}(T) = \frac{1}{k!}\sum_{\sigma\in\mathsf{S}_k}(\operatorname{sgn}\sigma)\,\sigma T$, as to actually make it a projection. There are competing conventions in the literature and you must pay close attention to them everytime you consult a new source. For example, we are following the convention adopted in [23], while [15] has the normalizing factor of $\frac{1}{k!}$.

Proof: The argument again consists of a direct computation:

$$\begin{aligned}\mathrm{Alt}(\mathrm{Alt}(T)) &= \sum_{\sigma\in\mathsf{S}_k}(\operatorname{sgn}\sigma)\,\sigma\mathrm{Alt}(T) = \sum_{\sigma\in\mathsf{S}_k}(\operatorname{sgn}\sigma)\,\sigma\left(\sum_{\tau\in\mathsf{S}_k}(\operatorname{sgn}\tau)\tau T\right) \\ &= \sum_{\sigma\in\mathsf{S}_k}\sum_{\tau\in\mathsf{S}_k}(\operatorname{sgn}\sigma)(\operatorname{sgn}\tau)\sigma(\tau T) = \sum_{\sigma\in\mathsf{S}_k}\sum_{\tau\in\mathsf{S}_k}(\operatorname{sgn}\sigma\tau)(\sigma\tau)T \\ &\stackrel{(*)}{=} \sum_{\sigma\in\mathsf{S}_k}\mathrm{Alt}(T) = k!\mathrm{Alt}(T).\end{aligned}$$

In $(*)$, we use the same general principle mentioned in the previous proof to say that $\sum_{\tau\in\mathsf{S}_k}(\operatorname{sgn}\sigma\tau)(\sigma\tau)T = \mathrm{Alt}(T)$: this time σ is fixed and the product $\sigma\tau$ ranges over S_k as τ ranges over S_k, without repetitions. □

One last (rather technical) property that will need later is given next:

[10] If Z is a vector space and $P\colon Z \to Z$ is a linear operator, P is called a **projection operator** onto some subspace $W \subseteq Z$ if $P \circ P = P$ and $\mathrm{Im}(P) = W$. It then follows that $V = \ker P \oplus \mathrm{Im}\,P$, with $v = (v - Pv) + Pv$ for every $v \in V$. Conversely, if a direct sum decomposition $Z = W \oplus W'$ is given, the direct-sum projection $Z \to W$ is a projection operator in the above sense.

Lemma 8 ("Pre-associativity" of Alt)

Let V be a real vector space, and consider tensors $T \in \mathcal{T}^k(V)$ and $S \in \mathcal{T}^\ell(V)$. Then $\mathrm{Alt}(\mathrm{Alt}(T) \otimes S) = k!\mathrm{Alt}(T \otimes S)$ and, similarly, $\mathrm{Alt}(T \otimes \mathrm{Alt}(S)) = \ell!\mathrm{Alt}(T \otimes S)$.

Proof: We establish just the first claimed relation, as the second one is dealt with similarly. Given $\tau \in \mathrm{S}_k$, we consider $\hat{\tau} \in \mathrm{S}_{k+\ell}$ defined by

$$\hat{\tau}(j) = \begin{cases} \tau(j), & \text{if } j \in \{1, \ldots, k\} \\ j, & \text{if } j \in \{k+1, \ldots k+\ell\}. \end{cases}$$

Clearly $\mathrm{S}_k \ni \tau \mapsto \hat{\tau} \in \mathrm{S}_{k+\ell}$ is an injective homomorphism, with $\operatorname{sgn} \hat{\tau} = \operatorname{sgn} \tau$ for every $\tau \in \mathrm{S}_k$, and such that $\hat{\tau}(\_ \otimes \_) = (\tau \_) \otimes \_$. With this in place, we compute

$$\begin{aligned} \mathrm{Alt}(\mathrm{Alt}(T) \otimes S) &= \sum_{\sigma \in \mathrm{S}_{k+\ell}} (\operatorname{sgn} \sigma)\, \sigma(\mathrm{Alt}(T) \otimes S) \\ &= \sum_{\sigma \in \mathrm{S}_{k+\ell}} (\operatorname{sgn} \sigma)\, \sigma\left(\left(\sum_{\tau \in \mathrm{S}_k} (\operatorname{sgn} \tau)\, (\tau T)\right) \otimes S\right) \\ &= \sum_{\sigma \in \mathrm{S}_{k+\ell}} \sum_{\tau \in \mathrm{S}_k} (\operatorname{sgn} \sigma)(\operatorname{sgn} \tau)\, \sigma((\tau T) \otimes S) \\ &= \sum_{\sigma \in \mathrm{S}_{k+\ell}} \sum_{\tau \in \mathrm{S}_k} (\operatorname{sgn} \sigma)(\operatorname{sgn} \hat{\tau})\, (\sigma\hat{\tau})(T \otimes S) \\ &= \sum_{\tau \in \mathrm{S}_k} \sum_{\sigma \in \mathrm{S}_{k+\ell}} (\operatorname{sgn}(\sigma\hat{\tau}))\, (\sigma\hat{\tau})(T \otimes S) \\ &\overset{(*)}{=} \sum_{\tau \in \mathrm{S}_k} \mathrm{Alt}(T \otimes S) = k!\mathrm{Alt}(T \otimes S), \end{aligned} \tag{5.5}$$

as required; the step in $(*)$ being hopefully clear by now. □

Exercise 101

Mimic (5.5) to show that $\mathrm{Alt}(T \otimes \mathrm{Alt}(S)) = \ell!\mathrm{Alt}(T \otimes S)$.

We may now modify the product $\otimes$ to obtain the desired product $\wedge$:

Definition 49 (Exterior product)

Let V be a real vector space. The **exterior product** of $\omega \in \mathcal{A}^k(V)$ and $\eta \in \mathcal{A}^\ell(V)$ is the element $\omega \wedge \eta \in \mathcal{A}^{k+\ell}(V)$ defined by

$$\omega \wedge \eta = \frac{1}{k!\ell!}\mathrm{Alt}(\omega \otimes \eta).$$

The presence of Alt ensures that $\omega \wedge \eta$ is alternating.

Remark. There are again different conventions on how the $\wedge$ operation is defined. If one defined Alt with the normalization factor of $\frac{1}{k!}$ mentioned in the previous remark, defining

$$\omega \wedge \eta = \frac{(k+\ell)!}{k!\ell!}\mathrm{Alt}(\omega \otimes \eta)$$

makes the resulting tensor $\omega \wedge \eta$ agree with the one in Definition 49. In [15], the consistency is maintained, but there are other texts which are not too careful with that. In any case, the factor $k!\ell!$ in the denominator simply accounts for repetitions (as you can see in Example 96 below).

Example 95 (Exterior products of linear functionals)

Let V be a vector space. If $f, g \in V^*$, then $f \wedge g \in \mathscr{A}^2(V)$ is given by

$$\begin{aligned}(f \wedge g)(v, w) &= \frac{1}{1!1!}\mathrm{Alt}(f \otimes g)(v, w) \\ &= (f \otimes g)(v, w) - (f \otimes g)(w, v) \\ &= f(v)g(w) - f(w)g(v),\end{aligned}$$

for all $v, w \in V$. When $V = \mathbb{R}^2$ and e^1, e^2 is the dual basis of the standard basis of $\mathbb{R}^2$, we have that $e^1 \wedge e^2 = \mathrm{A} = \det$ is the "area function" on the plane. Also note that $g \wedge f = -f \wedge g$.

Example 96 (Exterior products of linear functionals and skew-symmetric forms)

Let V be a real vector space, $f \in V^*$, and $\omega \in \mathscr{A}^2(V)$. Using the result from Example 94, we may compute

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

Since ω is skew-symmetric, we may rewrite the above as

$$\mathrm{Alt}(f \otimes \omega)(v_1, v_2, v_3) = 2(f(v_1)\omega(v_2, v_3) + f(v_2)\omega(v_3, v_1) + f(v_3)\omega(v_1, v_2)).$$

Dividing out the factor of 2, we conclude that

$$(f \wedge \omega)(v_1, v_2, v_3) = f(v_1)\omega(v_2, v_3) + f(v_2)\omega(v_3, v_1) + f(v_3)\omega(v_1, v_2), \tag{5.6}$$

for every $v_1, v_2, v_3 \in V$.

Exercise 102

In the same setting as Example 96 above, show that $\omega \wedge f = f \wedge \omega$.

Exercise 103

In Example 96, we presented the result as in (5.6) because then all terms present the cyclic order 1-2-3. This makes (5.6) easy to remember, but it is particular to the case where $\omega \in \mathscr{A}^2(V)$ (which, incidentally, is one of the most frequently encountered in practice). Another option is to write all the arguments of ω in order, starting from the smaller index to the largest, skipping the one taken care of by f:

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

The signs alternate, and this is how we generalize it all from $k = 2$ to arbitrary k.

Let V be a real vector space, and $k \geq 1$ be any integer. Show that whenever $f \in V^*$ and $\omega \in \mathscr{A}^k(V)$ are given, the exterior product $f \wedge \omega \in \mathscr{A}^{k+1}(V)$ is given by

$$(f \wedge \omega)(v_1, \ldots, v_{k+1}) = \sum_{i=1}^{k+1} (-1)^{i+1} f(v_i)\omega(v_1, \ldots, \widehat{v_i}, \ldots, v_{k+1})$$

for every $v_1, \ldots, v_{k+1} \in V$, where $\widehat{v_i}$ means that the argument v_i is being omitted. For example, $\omega(v_1, \ldots, \widehat{v_3}, \ldots, v_{k+1})$ means $\omega(v_1, v_2, v_4, \ldots, v_{k+1})$.

Hint: One option is to use induction. Alternatively, you can directly argue that for each fixed $i = 1, \ldots, k+1$ there are $k!$ terms of the form $f(v_i)\omega(\ldots, \widehat{v_i}, \ldots)$ in the expression for $\mathrm{Alt}(f \otimes \omega)(v_1, \ldots, v_{k+1})$, and all of them agree (up to a sign which may be adjusted by permuting the arguments) as ω is alternating.

We have seen in Exercise 99 that $\otimes$ is not commutative. As for $\wedge$, we had that it was skew-symmetric when operating on linear functionals, but symmetric when acting on a linear function and an alternating 2-tensor. This suggests that whether $\wedge$ is a commutative or anti-commutative product depends on the "degree" of the objects on which it acts. Here are the main properties of the exterior product:

Proposition 38 (Properties of $\wedge$)

Let V be a real vector space.

(a) $\wedge\colon \mathscr{A}^k(V) \times \mathscr{A}^\ell(V) \to \mathscr{A}^{k+\ell}(V)$ is bilinear, that is, whenever $\omega, \omega_1, \omega_2 \in \mathscr{A}^k(V)$ and $\eta, \eta_1, \eta_2 \in \mathscr{A}^\ell(V)$, and $\lambda \in \mathbb{R}$, we have that

$$(\omega_1 + \lambda\omega_2) \wedge \eta = \omega_1 \wedge \eta + \lambda\omega_2 \wedge \eta,$$
$$\omega \wedge (\eta_1 + \lambda\eta_2) = \omega \wedge \eta_1 + \lambda\omega \wedge \eta_2.$$

(b) $\wedge$ is associative, that is, $(\omega \wedge \eta) \wedge \zeta = \omega \wedge (\eta \wedge \zeta)$ for all alternating tensors $\omega \in \mathscr{A}^k(V)$, $\eta \in \mathscr{A}^\ell(V)$, and $\zeta \in \mathscr{A}^r(V)$.

(c) $\wedge$ is **graded-commutative**, that is, $\omega \wedge \eta = (-1)^{k\ell}\eta \wedge \omega$ whenever $\omega \in \mathscr{A}^k(V)$ and $\eta \in \mathscr{A}^\ell(V)$.

Proof: Property (a) is an immediate consequence of $\wedge$ being (up to the normalization factor) the composition of the bilinear mapping $\otimes\colon \mathscr{A}^k(V) \times \mathscr{A}^\ell(V) \to \mathscr{T}^{k+\ell}(V)$ (cf. Exercise 99) and the linear mapping $\mathrm{Alt}\colon \mathscr{T}^{k+\ell}(V) \to \mathscr{A}^{k+\ell}(V)$.

For property (b), we use Lemma 8: on one hand, we have that

$$\begin{aligned}(\omega \wedge \eta) \wedge \zeta &= \frac{1}{(k+\ell)!r!}\mathrm{Alt}((\omega \wedge \eta) \otimes \zeta) \\ &= \frac{1}{(k+\ell)!r!}\mathrm{Alt}\left(\left(\frac{1}{k!\ell!}\mathrm{Alt}(\omega \otimes \eta)\right) \otimes \zeta\right) \\ &= \frac{1}{k!\ell!r!}\frac{1}{(k+\ell)!}\mathrm{Alt}(\mathrm{Alt}(\omega \otimes \eta) \otimes \zeta) \\ &= \frac{1}{k!\ell!r!}\frac{1}{(k+\ell)!}(k+\ell)!\mathrm{Alt}((\omega \otimes \eta) \otimes \zeta) \\ &= \frac{1}{k!\ell!r!}\mathrm{Alt}((\omega \otimes \eta) \otimes \zeta).\end{aligned}$$

A similar calculation shows that

$$\omega \wedge (\eta \otimes \zeta) = \frac{1}{k!\ell!r!}\mathrm{Alt}(\omega \otimes (\eta \otimes \zeta)),$$

so that $(\omega \wedge \eta) \wedge \zeta = \omega \wedge (\eta \wedge \zeta)$ follows from $(\omega \otimes \eta) \otimes \zeta = \omega \otimes (\eta \otimes \zeta)$.

Finally, consider (c). The particular permutation $\alpha \in \mathrm{S}_{k+\ell}$ given by

$$\alpha(j) = \begin{cases} j+\ell, & \text{if } j \in \{1, \ldots, k\}, \\ j-k, & \text{if } j \in \{k+1, \ldots, k+\ell\} \end{cases}$$

has $\operatorname{sgn}\alpha = (-1)^{k\ell}$ and satisfies that $\alpha(\omega \otimes \eta) = \eta \otimes \omega$. Applying Alt on both sides of this last relation, the conclusion immediately follows. Here, of course, we use the following general fact: $\mathrm{Alt}(\sigma T) = (\operatorname{sgn}\sigma)\,\mathrm{Alt}(T)$, valid for any tensor $T \in \mathscr{T}^k(V)$ and permutation $\sigma \in \mathrm{S}_k$. □

Exercise 104

Establish the general fact stated at the end of the above proof.

Associativity of the exterior product has a few important consequences:

Corollary 12

Let V be a real vector space, an integer $r \geq 2$, and consider alternating tensors $\omega_j \in \mathcal{A}^{k_j}(V)$, for $j = 1, \ldots, r$. Then the relation

$$\omega_1 \wedge \cdots \wedge \omega_r = \frac{1}{k_1! \cdots k_r!} \operatorname{Alt}(\omega_1 \otimes \cdots \otimes \omega_r)$$

holds. In particular, $f^1 \wedge \cdots \wedge f^r = \operatorname{Alt}(f^1 \otimes \cdots \otimes f^r)$ for any $f^1, \ldots, f^r \in V^*$.

Proof: The argument is by induction on r. When $r = 1$ there is nothing to do, and when $r = 2$ it is just the definition of the exterior product. Assume that $r \geq 3$ and that the conclusion holds for products of $r - 1$ alternating tensors. Writing $\omega_1 \wedge \cdots \wedge \omega_r$ as $\omega_1 \wedge (\omega_2 \wedge \cdots \wedge \omega_r)$, we may compute:

$$\begin{aligned}
\omega_1 \wedge \cdots \wedge \omega_r &= \frac{1}{k_1!(k_2 + \cdots + k_r)!} \operatorname{Alt}(\omega_1 \otimes (\omega_2 \wedge \cdots \wedge \omega_r)) \\
&= \frac{1}{k_1!(k_2 + \cdots + k_r)!} \operatorname{Alt}\left(\omega_1 \otimes \left(\frac{1}{k_2! \cdots k_r!} \operatorname{Alt}(\omega_2 \otimes \cdots \otimes \omega_r)\right)\right) \\
&= \frac{1}{k_1! \cdots k_r!} \frac{1}{(k_2 + \cdots + k_r)!} \operatorname{Alt}(\omega_1 \otimes \operatorname{Alt}(\omega_2 \otimes \cdots \otimes \omega_r)) \\
&\overset{(\dagger)}{=} \frac{1}{k_1! \cdots k_r!} \frac{1}{(k_2 + \cdots + k_r)!} (k_2 + \cdots + k_r)! \operatorname{Alt}(\omega_1 \otimes \omega_2 \otimes \cdots \otimes \omega_r) \\
&\overset{(\ddagger)}{=} \frac{1}{k_1! \cdots k_r!} \operatorname{Alt}(\omega_1 \otimes \cdots \otimes \omega_r),
\end{aligned}$$

as required. In $(\dagger)$ we use Lemma 8, and in $(\ddagger)$ we use associativity of the tensor product operation. □

Corollary 13 ("Repeats kill")

If V is a real vector space and $k \geq 1$ is an odd integer, then $\omega \wedge \omega = 0$ for every $\omega \in \mathcal{A}^k(V)$.

Proof: We use that $\wedge$ is graded commutative and that, for odd k, $(-1)^k = -1$. Thus $\omega \wedge \omega = -\omega \wedge \omega$ implies that $\omega \wedge \omega = 0$. □

Probably the most frequent application of Corollary 13 is when we have a product $\omega = f^1 \wedge \cdots \wedge f^k$ of $f^1, \ldots, f^k \in V^*$: if there are two distinct indices i and j such that $f^i = f^j$, then $\omega = 0$.

Exercise 105

Exhibit some $\omega \in \mathcal{A}^2(\mathbb{R}^4)$ such that $\omega \wedge \omega \neq 0$.

Corollary 14

Let V be a real vector space, and $f^1, \ldots, f^k \in V^*$ be given. Then the exterior product $f^1 \wedge \cdots \wedge f^k \in \mathcal{A}^k(V)$ is given by

$$(f^1 \wedge \cdots \wedge f^k)(v_1, \ldots, v_k) = \det[f^i(v_j)]_{i,j=1}^k,$$

for all $v_1, \ldots, v_k \in V$.

Proof: We once again argue by induction on k. When $k = 1$ there is nothing to do; when $k = 2$ this is what we have seen in Example 95. Assume that $k \geq 3$ and that the desired conclusion holds for the product of $k-1$ linear functionals. By Exercise 103 and writing $f^1 \wedge \cdots \wedge f^k = f^1 \wedge (f^2 \wedge \cdots \wedge f^k)$, we have that

$$\begin{aligned}(f^1 \wedge \cdots \wedge f^k)(v_1, \ldots, v_k) &= \sum_{i=1}^k (-1)^{i+1} f^1(v_i)(f^2 \wedge \cdots \wedge f^k)(v_1, \ldots, \widehat{v_i}, \ldots, v_k) \\ &= \sum_{i=1}^k (-1)^{i+1} f^1(v_i) \det \begin{bmatrix} f^2(v_1) & \cdots & \widehat{f^2(v_i)} & \cdots & f^2(v_k) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ f^k(v_1) & \cdots & \widehat{f^k(v_i)} & \vdots & f^k(v_k) \end{bmatrix} \\ &= \det[f^i(v_j)]_{i,j=1}^k,\end{aligned}$$

as required. The expression after the second equality is simply the row-expansion of $\det[f^i(v_j)]_{i,j=1}^k$ through its first row. □

Corollary 14 illustrates a general trend: exterior products and determinants go hand-in-hand.

Example 97 (Determinants as sums over permutation groups)

The determinant of a matrix $A = [a_{ij}]_{i,j=1}^n$ can also be expressed as a sum over the permutation group S_n:

$$\det A = \sum_{\sigma \in \mathrm{S}_n} (\operatorname{sgn} \sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}. \tag{5.7}$$

Equation (5.7) is often taken as the definition of determinant—there are several different equivalences. For instance, one can directly prove that there is a unique $\omega \in \mathcal{A}^n(\mathbb{R}^n)$ such that $\omega(e_1, \ldots, e_n) = 1$, where $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$, and then baptize ω as "determinant". If one identifies $A \in \mathbb{R}^{n \times n}$ with $(Ae_1, \ldots, Ae_n) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n$, the equality $\det A = \omega(Ae_1, \ldots, Ae_n)$ we have implicitly used in Example 83 (p. 129) is in fact a theorem.

Proposition 39

Let V be a real vector space, $\omega \in \mathcal{A}^k(V)$, and $A = [a^i_j]^k_{i,j=1}$ be a matrix. Then, we have that

$$\omega(\widetilde{v}_1, \ldots, \widetilde{v}_k) = (\det A)\, \omega(v_1, \ldots, v_k), \quad \text{where } \widetilde{v}_j = \sum_{i=1}^k a^i_j v_i \text{ for } j = 1, \ldots, k,$$

for every $v_1, \ldots, v_k \in V$.

Proof: We use multilinearity of ω to write

$$\omega(\widetilde{v}_1, \ldots, \widetilde{v}_k) = \omega\left(\sum_{i_1=1}^k a^{i_1}_1 v_{i_1}, \ldots, \sum_{i_1=1}^k a^{i_k}_k v_{i_k} \right) = \sum_{i_1,\ldots,i_k=1}^n a^{i_1}_1 \cdots a^{i_k}_k \omega(v_{i_1}, \ldots, v_{i_k}).$$

Now, for each k-tuple $(i_1, \ldots, i_k)$ as above, consider the permutation $\sigma \in \mathrm{S}_k$ defined by $\sigma(j) = i_j$ for every $j = 1, \ldots, k$. As every element of S_k is obviously of this form for some k-tuple $(i_1, \ldots, i_k)$, we may rewrite the last summation as a sum over S_k and apply the alternating property of ω:

$$\begin{aligned}
\omega(\widetilde{v}_1, \ldots, \widetilde{v}_k) &= \sum_{\sigma \in \mathrm{S}_k} a^{\sigma(1)}_1 \cdots a^{\sigma(k)}_k \omega(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) \\
&= \sum_{\sigma \in \mathrm{S}_k} a^{\sigma(1)}_1 \cdots a^{\sigma(k)}_k (\operatorname{sgn} \sigma) \omega(v_1, \ldots, v_k) \\
&= \left(\sum_{\sigma \in \mathrm{S}_k} (\operatorname{sgn} \sigma) a^{\sigma(1)}_1 \cdots a^{\sigma(k)}_k \right) \omega(v_1, \ldots, v_k) \\
&= (\det A^{\mathrm{T}})\, \omega(v_1, \ldots, v_k) \\
&= (\det A)\, \omega(v_1, \ldots, v_k),
\end{aligned}$$

as claimed. □

Exercise 106

"Dualize" Proposition 39: show that if V is a real vector space and $A = [a^i_j]^k_{i,j=1}$ is a matrix, then

$$\widetilde{f}^1 \wedge \cdots \wedge \widetilde{f}^k = (\det A)\, f^1 \wedge \cdots \wedge f^k, \quad \text{for } \widetilde{f}^i = \sum_{j=1}^k a^i_j f^j \text{ for } i = 1, \ldots, k,$$

for every $f^1, \ldots, f^k \in V^*$.

Proposition 40 (A basis for $\mathscr{A}^k(V)$)

Let V be a real vector space, and $(e_1,\ldots,e_n)$ be a basis for V. Then, if $(\varphi^1,\ldots,\varphi^n)$ is its dual basis in V^*, we have that

$$\{\varphi^{i_1}\wedge\cdots\wedge\varphi^{i_k} : 1\leq i_1<\cdots<i_k\leq n\} \tag{5.8}$$

is a basis for $\mathscr{A}^k(V)$. In particular, $\dim\mathscr{A}^k(V)=\binom{n}{k}$.

Proof: We start by verifying that (5.8) is linearly independent. So, consider $\binom{n}{k}$ coefficients $a_{i_1\cdots i_k}\in\mathbb{R}$, always with $1\leq i_1<\cdots<i_k\leq n$, such that

$$\sum_{1\leq i_1<\ldots<i_k\leq n} a_{i_1\cdots i_k}\varphi^{i_1}\wedge\cdots\wedge\varphi^{i_k}=0. \tag{5.9}$$

Once we have shown that

$$(\varphi^{i_1}\wedge\cdots\wedge\varphi^{i_k})(e_{j_1},\ldots,e_{j_k})=\begin{cases}1, & \text{if } i_r=j_r \text{ for every } r=1,\ldots,k,\\ 0, & \text{else,}\end{cases} \tag{5.10}$$

whenever $1\leq j_1<\cdots<j_k\leq n$, i.e., that $(\varphi^{i_1}\wedge\cdots\wedge\varphi^{i_k})(e_{j_1},\ldots,e_{j_k})=\delta^{i_1}_{j_1}\cdots\delta^{i_k}_{j_k}$, we obtain that $a_{j_1\cdots j_k}=0$ by evaluating both sides of (5.9) at $(e_{j_1},\ldots,e_{j_k})$, just as in the proof of Proposition 35. The reasons why (5.10) holds, at the end of the day, are Corollary 14 and induction: if $j_1<i_1$ the first column of $[\varphi^{i_r}(e_{j_s})]_{r,s=1}^k$ vanishes, while if $j_1>i_1$ it is its first row that vanishes, and in either case $\det[\varphi^{i_r}(e_{j_s})]_{r,s=1}^k=0$. If such determinant is to be nonzero we must have that $j_1=i_1$, and now we may proceed inductively.

Finally, to establish that (5.8) spans $\mathscr{A}^k(V)$, we may use Proposition 35 as a shortcut: if $\omega\in\mathscr{A}^k(V)$, then in particular $\omega\in\mathscr{T}^k(V)$, and so it can be written as

$$\omega=\sum_{i_1,\ldots,i_k=1}^{n}\omega_{i_1\ldots i_k}\varphi^{i_1}\otimes\cdots\otimes\varphi^{i_k}, \tag{5.11}$$

where $\omega_{i_1\ldots i_k}=\omega(e_{i_1},\ldots,e_{i_k})$ are the components of ω relative to the basis $(e_1,\ldots,e_n)$. Applying Alt to both sides of (5.11), and using Proposition 37 together with Corollary 12, it follows that

$$k!\omega=\sum_{i_1,\ldots,i_k=1}^{n}\omega_{i_1\ldots i_k}\varphi^{i_1}\wedge\cdots\wedge\varphi^{i_k}. \tag{5.12}$$

For each k-tuple $(i_1,\ldots,i_k)$ there are $k!$ components of ω containing the indices $i_1,\ldots,i_k$, but they are all equal to each other up to a sign, since ω is alternating. Hence, rearranging common terms as to have the summation be over only increasing k-tuples of indices, we have that

$$k!\omega=\sum_{1\leq i_1<\ldots<i_k\leq n} k!\omega_{i_1\ldots i_k}\varphi^{i_1}\wedge\cdots\wedge\varphi^{i_k},$$

and so

$$\omega = \sum_{1 \le i_1 < \ldots < i_k \le n} \omega_{i_1 \ldots i_k} \varphi^{i_1} \wedge \cdots \wedge \varphi^{i_k}, \tag{5.13}$$

as required. □

Remark. Note how the *same* coefficients are used in both basis-expansions (5.11) and (5.13). But if you write ω as a linear combination of the k-fold products $\varphi^{i_1} \wedge \cdots \wedge \varphi^{i_k}$ without restricting yourself to increasing k-tuples of indices, you must pay the price of $k!$, cf. (5.12).

Corollary 15

Let V be a real vector space, and assume that $n = \dim V$. Then, whenever $\varphi^1, \ldots, \varphi^n \in V^*$ are linearly independent, $\mathbb{R} \ni \lambda \mapsto \lambda\, \varphi^1 \wedge \cdots \wedge \varphi^n \in \mathcal{A}^n(V)$ is an isomorphism.

The next two exercises contain two of the main results about "$\wedge$-divisibility".

Exercise 107

Let V be a finite-dimensional vector space, $\omega \in \mathcal{A}^k(V)$, and $\varphi \in V^* \setminus \{0\}$. Show that if $\omega \wedge \varphi = 0$, there is $\eta \in \mathcal{A}^{k-1}(V)$ such that $\omega = \eta \wedge \varphi$.

Hint: Complete φ to a basis of V^* and use (5.13) to determine η explicitly.

Exercise 108 (Cartan's Lemma)

Let V be a finite-dimensional real vector space, and $\varphi^1, \ldots, \varphi^k \in V^*$ be linearly independent. Show that, for any $\alpha_1, \ldots, \alpha_k \in V^*$ such that

$$\alpha_1 \wedge \varphi^1 + \cdots + \alpha_k \wedge \varphi^k = 0,$$

there is a symmetric matrix $[h_{ij}]_{i,j=1}^k$ such that $\alpha_i = \sum_{j=1}^k h_{ij}\varphi^j$ for all $i = 1, \ldots, k$.

Note: This result does not assume that $k = \dim V$. It has a generalization: if we instead assume that $\alpha^i \in \mathcal{A}^p(V)$ for each i, then we have $h_{ij} \in \mathcal{A}^{p-1}(V)$ instead of $h_{ij} \in \mathbb{R}$, still with $h_{ij} = h_{ji}$, and $\alpha_i = \sum_{j=1}^k h_{ij} \wedge \varphi^j$. See [1, Lemma 1].

5.2 Differential forms and the exterior derivative

Now, we return to smooth manifolds. We have seen in Exercise 75 that whenever M is a smooth manifold and $(U; x^1, \ldots, x^n)$ is a chart for U, the differentials $\mathrm{d}x^1|_p, \ldots, \mathrm{d}x^n|_p$ form the basis of T_p^*M dual to the coordinate basis of T_pM induced by the chart, for all $p \in U$. We may now consider "fields of alternating k-tensors":

Definition 50 (Differential forms)

Let M be a smooth manifold, $n = \dim M$, and $k \geq 0$ be an integer. A **differential form of degree** k (or, a **differential** k**-form**) on M is an assignment ω, to each point $p \in M$, of an element $\omega_p \in \mathcal{A}^k(T_pM)$. The integer k is called the **degree** of ω. We say that ω is smooth if, for every chart $(U; x^1, \ldots, x^n)$ for M, the $\binom{n}{k}$ functions $\omega_{i_1 \ldots i_k} \colon U \to \mathbb{R}$ defined by the relation

$$\omega_p = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \omega_{i_1 \ldots i_k}(p) \mathrm{d}x^{i_1}|_p \wedge \cdots \wedge \mathrm{d}x^{i_k}|_p, \qquad \text{for all } p \in U,$$

are smooth; note that

$$\omega_{i_1 \cdots i_k}(p) = \omega_p \left(\left.\frac{\partial}{\partial x^{i_1}}\right|_p, \cdots, \left.\frac{\partial}{\partial x^{i_k}}\right|_p \right)$$

for each increasing k-tuple $1 \leq i_1 < \cdots < i_k \leq n$. We let $\Omega^k(M)$ denote the space of all smooth differential forms of degree k on M.

Remark. As done for vector fields, we will always assume that differential forms are smooth, unless otherwise stated.

In other words, a differential form $\omega \in \Omega^k(M)$ is something that takes as input a point $p \in M$ and k tangent vectors $v_1, \ldots, v_k \in T_pM$, and outputs a number $\omega_p(v_1, \ldots, v_k) \in \mathbb{R}$, which depends on the vectors in a multilinear and alternating fashion. For $k = 0$, we have that $\Omega^0(M) = C^\infty(M)$.

Example 98

Whenever M is a smooth manifold, and $f \colon M \to \mathbb{R}$ is a smooth function, we have that $\mathrm{d}f \in \Omega^1(M)$. This is because $\mathrm{d}f_p \in (T_pM)^* = \mathcal{T}^1(T_pM) = \mathcal{A}^1(T_pM)$ for each $p \in M$—linear functionals are alternating by default.

Example 99

Let $U \subseteq \mathbb{R}^3$ be an open subset. We have global coordinates x, y, z on U, and therefore $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z \in \Omega^1(U)$.

- Any element of $\Omega^1(U)$ is of the form $\omega_1 = f\,\mathrm{d}x + g\,\mathrm{d}y + h\,\mathrm{d}z$.
- Any element of $\Omega^2(U)$ is of the form $\omega_2 = f\,\mathrm{d}y \wedge \mathrm{d}z + g\,\mathrm{d}z \wedge \mathrm{d}x + h\,\mathrm{d}x \wedge \mathrm{d}y$.
- Any element of $\Omega^3(U)$ is of the form $\omega_3 = f\,\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$.

Above, $f, g, h \colon U \to \mathbb{R}$ are real functions.

Concretely, consider $\omega = xy^2 \mathrm{d}x \wedge \mathrm{d}z \in \Omega^2(\mathbb{R}^3)$. Say we want to compute the value $\omega_{(1,2,3)}((1,1,1), (2,1,3))$. The coefficients in front of the "basic" 2-form are evaluated with the point $p = (1,2,3)$, while the exterior products receive the tangent vector arguments. As common practice, we write simply $\mathrm{d}x$, $\mathrm{d}y$, and $\mathrm{d}z$ for what should technically be written as $\mathrm{d}x_{(1,2,3)}$, $\mathrm{d}y_{(1,2,3)}$, and $\mathrm{d}z_{(1,2,3)}$, respectively. This way, we have that

$$\omega_{(1,2,3)}((1,1,1),(2,1,3)) = 1\cdot 2^2\,(\mathrm{d}x\wedge\mathrm{d}z)((1,1,1),(2,1,3)) = 4\det\begin{bmatrix}\mathrm{d}x(1,1,1) & \mathrm{d}x(2,1,3)\\ \mathrm{d}z(1,1,1) & \mathrm{d}z(2,1,3)\end{bmatrix} = 4\det\begin{bmatrix}1 & 2\\ 1 & 3\end{bmatrix} = 4.$$

Of course, we have the identifications

$$(1,1,1) = \frac{\partial}{\partial x}\bigg|_{(1,2,3)} + \frac{\partial}{\partial y}\bigg|_{(1,2,3)} + \frac{\partial}{\partial z}\bigg|_{(1,2,3)},$$

and a similar one for $(2,1,3)$, in place.

Exercise 109

In each item, evaluate the given differential form at the given point and tangent vectors:

(a) $\omega = yz\,\mathrm{d}x + xz\,\mathrm{d}y + xy\,\mathrm{d}z \in \Omega^1(\mathbb{R}^3)$, at $(3,2,1)\in T_{(1,1,2)}(\mathbb{R}^3)$.

(b) $\eta = xyz\,\mathrm{d}x\wedge\mathrm{d}y + yz^2\,\mathrm{d}y\wedge\mathrm{d}z \in \Omega^2(\mathbb{R}^3)$, at $(2,0,1),(3,3,1)\in T_{(0,1,1)}(\mathbb{R}^3)$.

(c) $\zeta = \sin\big(xe^y + \arctan(\sqrt[3]{z})\big)\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z \in \Omega^3(\mathbb{R}^3)$, at $(1,2,4),(3,2,2),(1,0,1)\in T_{(3,0,0)}(\mathbb{R}^3)$.

On manifolds, we are now able to take derivatives of differential forms. We will first do so in open subsets of $\mathbb{R}^n$, and then transplant it onto smooth manifolds.

Definition 51 (Exterior derivative in $\mathbb{R}^n$)

Let $U\subseteq\mathbb{R}^n$ be an open subset. The **exterior derivative** is $\mathrm{d}\colon \Omega^k(U)\to\Omega^{k+1}(U)$, defined as follows: if $\omega\in\Omega^k(U)$ is written as

$$\omega = \sum_{1\le i_1<\cdots<i_k\le n} \omega_{i_1\ldots i_k}\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k},$$

then $\mathrm{d}\omega\in\Omega^{k+1}(U)$ is given by

$$\mathrm{d}\omega = \sum_{1\le i_1<\cdots<i_k\le n} \mathrm{d}\omega_{i_1\ldots i_k}\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k},$$

where $\mathrm{d}\omega_{i_1\ldots i_k}$ is the differential of the component function $\omega_{i_1\ldots i_k}\colon U\to\mathbb{R}$.

Remark. Technically, the exterior derivative is not a single operator, but instead a collection of operators $\mathrm{d}\colon \Omega^k(U) \to \Omega^{k+1}(U)$, one for each k. We denote all of them with the same letter d.

Example 100

Consider in $\mathbb{R}^3$ the differential form $\omega = xy^2\,\mathrm{d}x \wedge \mathrm{d}z$ from Example 98. Using Corollary 13, we have that

$$\begin{aligned}\mathrm{d}\omega &= \mathrm{d}(xy^2) \wedge \mathrm{d}x \wedge \mathrm{d}z = (y^2\,\mathrm{d}x + 2xy\,\mathrm{d}y) \wedge \mathrm{d}x \wedge \mathrm{d}z \\ &= y^2\,\mathrm{d}x \wedge \mathrm{d}x \wedge \mathrm{d}z + 2xy\,\mathrm{d}y \wedge \mathrm{d}x \wedge \mathrm{d}z = -2xy\,\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z.\end{aligned}$$

Exercise 110

Compute the exterior derivatives of the differential forms given in Exercise 109:

(a) $\omega = yz\,\mathrm{d}x + xz\,\mathrm{d}y + xy\,\mathrm{d}z \in \Omega^1(\mathbb{R}^3)$.

(b) $\eta = xyz\,\mathrm{d}x \wedge \mathrm{d}y + yz^2\,\mathrm{d}y \wedge \mathrm{d}z \in \Omega^2(\mathbb{R}^3)$.

(c) $\zeta = \sin\big(xe^y + \arctan(\sqrt[3]{z})\big)\,\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \in \Omega^3(\mathbb{R}^3)$.

We register next the main properties of the exterior derivative:

Proposition 41 (Properties of d)

Let $U \subseteq \mathbb{R}^n$ be open, and $k, \ell \geq 0$ be integers. Then:

(i) On $\Omega^0(U) = C^\infty(U)$, the exterior derivative d agrees with the differential.

(ii) $\mathrm{d}(\omega \wedge \eta) = \mathrm{d}\omega \wedge \eta + (-1)^k \omega \wedge \mathrm{d}\eta$, for all $\omega \in \Omega^k(U)$ and $\eta \in \Omega^\ell(U)$.

(iii) $\mathrm{d}^2 = 0$.

Proof: Item (i) is a consequence of the expression $\mathrm{d}f = \sum_{i=1}^n (\partial f/\partial x^i)\,\mathrm{d}x^i$, seen in Example 69. For items (ii) and (iii), the obvious linearity of d allows us to simply consider monomials $\omega = f\,\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}$ and $\eta = g\,\mathrm{d}x^{j_1} \wedge \cdots \wedge \mathrm{d}x^{j_\ell}$, where the k-tuple $(i_1, \ldots, i_k)$ and the ℓ-tuple $(j_1, \ldots, j_\ell)$ are fixed.

To prove (ii), first note that $\omega \wedge \eta = fg\,\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k} \wedge \mathrm{d}x^{j_1} \wedge \cdots \wedge \mathrm{d}x^{j_\ell}$. The product rule for differentials now yields $\mathrm{d}(fg) = g\,\mathrm{d}f + f\,\mathrm{d}g$. We address the two resulting terms separately:

$$\begin{aligned} g\,\mathrm{d}f \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k} \wedge \mathrm{d}x^{j_1} \wedge \cdots \wedge \mathrm{d}x^{j_\ell} &= (\mathrm{d}f \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}) \wedge (g\,\mathrm{d}x^{j_1} \wedge \cdots \wedge \mathrm{d}x^{j_\ell}), \\ f\,\mathrm{d}g \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k} \wedge \mathrm{d}x^{j_1} \wedge \cdots \wedge \mathrm{d}x^{j_\ell} &= (-1)^k (f\,\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}) \wedge (\mathrm{d}g \wedge \mathrm{d}x^{j_1} \wedge \cdots \wedge \mathrm{d}x^{j_\ell}). \end{aligned}$$

The first one equals $\mathrm{d}\omega \wedge \eta$, while the second one equals $(-1)^k \omega \wedge \mathrm{d}\eta$. Adding them up yields (ii).

Finally, (iii) consists of a direct computation:

$$\begin{aligned}
\mathrm{d}^2\omega &= \mathrm{d}(\mathrm{d}\omega) \\
&= \mathrm{d}\left(\mathrm{d}f \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}\right) \\
&= \mathrm{d}\left(\sum_{j=1}^n \frac{\partial f}{\partial x^j}\,\mathrm{d}x^j \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}\right) \\
&= \sum_{j=1}^n \mathrm{d}\left(\frac{\partial f}{\partial x^j}\right) \wedge \mathrm{d}x^j \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k} \\
&= \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j}\,\mathrm{d}x^i \wedge \mathrm{d}x^j \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}.
\end{aligned}$$

The terms in this last summation having $i = j$ all vanish due to Corollary 13, and so we may break it all into two summations:

$$\begin{aligned}
\mathrm{d}^2\omega = \sum_{1\leq i<j\leq n} \frac{\partial^2 f}{\partial x^i \partial x^j}\,\mathrm{d}x^i \wedge \mathrm{d}x^j \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k} \\
+ \sum_{1\leq j<i\leq n} \frac{\partial^2 f}{\partial x^i \partial x^j}\,\mathrm{d}x^i \wedge \mathrm{d}x^j \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}.
\end{aligned}$$

They cancel each other because $\mathrm{d}x^i \wedge \mathrm{d}x^j = -\mathrm{d}x^j \wedge \mathrm{d}x^i$ and second-order mixed partial derivatives commute in Euclidean space. □

We may now show that $\wedge$ and d do generalize all products from vector calculus, as well as the gradient, divergence, and curl operators:

Theorem 21 (Vector calculus redux)

Let $U \subseteq \mathbb{R}^3$ be an open subset, and consider the identifications $\alpha\colon \mathfrak{X}(U) \to \Omega^1(U)$ and $\beta\colon \mathfrak{X}(U) \to \Omega^2(U)$ given by

$$\alpha_{\mathbf{X}} = P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z \quad \text{and} \quad \beta_{\mathbf{X}} = P\,\mathrm{d}y \wedge \mathrm{d}z + Q\,\mathrm{d}z \wedge \mathrm{d}x + R\,\mathrm{d}x \wedge \mathrm{d}y,$$

whenever $\mathbf{X} \in \mathfrak{X}(U)$ is written as

$$\mathbf{X} = P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y} + R\frac{\partial}{\partial z},$$

for suitable smooth functions $P, Q, R : U \to \mathbb{R}$. Similarly, we consider a third identification $\gamma\colon C^\infty(U) \to \Omega^3(U)$ given by $\gamma_f = f\,\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$. Then we have that the diagrams

$$\begin{array}{ccc} \mathfrak{X}(U)\times\mathfrak{X}(U) & \xrightarrow{\ \times\ } & \mathfrak{X}(U) \\ {\scriptstyle \alpha\times\alpha}\big\downarrow & & \big\downarrow{\scriptstyle \beta} \\ \Omega^1(U)\times\Omega^1(U) & \xrightarrow{\ \wedge\ } & \Omega^2(U) \end{array} \quad\text{and}\quad \begin{array}{ccc} \mathfrak{X}(U)\times\mathfrak{X}(U) & \xrightarrow{\ \langle\cdot,\cdot\rangle\ } & C^\infty(U) \\ {\scriptstyle \alpha\times\beta}\big\downarrow & & \big\downarrow{\scriptstyle \gamma} \\ \Omega^1(U)\times\Omega^2(U) & \xrightarrow{\ \wedge\ } & \Omega^3(U) \end{array}$$

as well as

$$\begin{array}{ccccccc} C^0(U) & \xrightarrow{\ \nabla\ } & \mathfrak{X}(U) & \xrightarrow{\ \mathrm{curl}\ } & \mathfrak{X}(U) & \xrightarrow{\ \mathrm{div}\ } & C^\infty(U) \\ \Big\| & & \big\downarrow{\scriptstyle \alpha} & & \big\downarrow{\scriptstyle \beta} & & \big\downarrow{\scriptstyle \gamma} \\ \Omega^0(U) & \xrightarrow{\ \mathrm{d}\ } & \Omega^1(U) & \xrightarrow{\ \mathrm{d}\ } & \Omega^2(U) & \xrightarrow{\ \mathrm{d}\ } & \Omega^3(U) \end{array}$$

commute. Explicitly, all of the relations (i) $\alpha_{\mathbf{X}}\wedge\alpha_{\mathbf{Y}} = \beta_{\mathbf{X}\times\mathbf{Y}}$, (ii) $\alpha_{\mathbf{X}}\wedge\beta_{\mathbf{Y}} = \gamma_{\langle\mathbf{X},\mathbf{Y}\rangle}$, (iii) $\alpha_{\nabla f} = \mathrm{d}f$, (iv) $\mathrm{d}(\alpha_{\mathbf{X}}) = \beta_{\mathrm{curl}\,\mathbf{X}}$, and (v) $\mathrm{d}(\beta_{\mathbf{X}}) = \gamma_{\mathrm{div}\,\mathbf{X}}$ hold, for all $\mathbf{X},\mathbf{Y}\in\mathfrak{X}(U)$ and $f\in C^\infty(U)$.

In particular, the property $\mathrm{d}^2 = 0$ corresponds to the well-known facts from vector calculus: $\mathrm{div}\circ\mathrm{curl} = 0$ and $\mathrm{curl}\circ\nabla = 0$.

Proof: Write the components of $\mathbf{Y}$ as A, B, C. Then

$$\begin{aligned} \alpha_{\mathbf{X}}\wedge\alpha_{\mathbf{Y}} &= (P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z)\wedge(A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z) \\ &= PA\,\mathrm{d}x\wedge\mathrm{d}x + PB\,\mathrm{d}x\wedge\mathrm{d}y + PC\,\mathrm{d}x\wedge\mathrm{d}z \\ &\quad + QA\,\mathrm{d}y\wedge\mathrm{d}x + QB\,\mathrm{d}y\wedge\mathrm{d}y + QC\,\mathrm{d}y\wedge\mathrm{d}z \\ &\quad + RA\,\mathrm{d}z\wedge\mathrm{d}x + RB\,\mathrm{d}z\wedge\mathrm{d}y + RC\,\mathrm{d}z\wedge\mathrm{d}z \\ &= (QC-RB)\mathrm{d}y\wedge\mathrm{d}z + (RA-PC)\mathrm{d}z\wedge\mathrm{d}x + (PB-QA)\mathrm{d}x\wedge\mathrm{d}y. \end{aligned}$$

But $QC-RB$, $RA-PC$, and $PB-QA$ are precisely the components of the cross product $\mathbf{X}\times\mathbf{Y}$. So, by definition of β, we conclude that $\alpha_{\mathbf{X}}\wedge\alpha_{\mathbf{Y}} = \beta_{\mathbf{X}\times\mathbf{Y}}$, proving (i). Relation (ii) is obtained in a similar manner. Then, (iii) is clear. As for (iv), using subscript notation for partial derivatives, we compute

$$\begin{aligned} \mathrm{d}(\alpha_{\mathbf{X}}) &= \mathrm{d}(P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z) \\ &= \mathrm{d}P\wedge\mathrm{d}x + \mathrm{d}Q\wedge\mathrm{d}y + \mathrm{d}R\wedge\mathrm{d}z \\ &= (P_x\,\mathrm{d}x + P_y\,\mathrm{d}y + P_z\,\mathrm{d}z)\wedge\mathrm{d}x \\ &\quad + (Q_x\,\mathrm{d}x + Q_y\,\mathrm{d}y + Q_z\,\mathrm{d}z)\wedge\mathrm{d}y \\ &\quad + (R_x\,\mathrm{d}x + R_y\,\mathrm{d}y + R_z\,\mathrm{d}z)\wedge\mathrm{d}z \\ &= P_y\,\mathrm{d}y\wedge\mathrm{d}x + P_z\,\mathrm{d}z\wedge\mathrm{d}x + Q_x\,\mathrm{d}x\wedge\mathrm{d}y \\ &\quad + Q_z\,\mathrm{d}z\wedge\mathrm{d}y + R_x\,\mathrm{d}x\wedge\mathrm{d}z + R_y\,\mathrm{d}y\wedge\mathrm{d}z \\ &= (R_y - Q_z)\mathrm{d}y\wedge\mathrm{d}z + (P_z - R_x)\mathrm{d}z\wedge\mathrm{d}x + (Q_x - P_y)\mathrm{d}x\wedge\mathrm{d}y \\ &= \beta_{\mathrm{curl}\,\mathbf{X}}. \end{aligned}$$

Finally,

$$\begin{aligned}
d(\beta_{\mathbf{X}}) &= d(P\,dy\wedge dz + Q\,dz\wedge dx + R\,dx\wedge dy) \\
&= dP\wedge dy\wedge dz + dQ\wedge dz\wedge dx + dR\wedge dx\wedge dy \\
&= (P_x\,dx + P_y\,dy + P_z\,dz)\wedge dy\wedge dz \\
&\quad + (Q_x\,dx + Q_y\,dy + Q_z\,dz)\wedge dz\wedge dx \\
&\quad + (R_x\,dx + R_y\,dy + R_z\,dz)\wedge dx\wedge dy \\
&= P_x\,dx\wedge dy\wedge dz + Q_y\,dy\wedge dz\wedge dx + R_z\,dz\wedge dx\wedge dy \\
&= (P_x + Q_y + R_z)dx\wedge dy\wedge dz \\
&= \gamma_{\operatorname{div}\mathbf{X}}
\end{aligned}$$

establishes (v). □

The properties of the exterior derivative seen in Proposition 41 actually characterize it completely – this is a crucial step in moving d from open subsets of $\mathbb{R}^n$ back to smooth manifolds:

Proposition 42 ("Uniqueness" of the exterior derivative in $\mathbb{R}^n$)

Let $U \subseteq \mathbb{R}^n$ be an open subset, and assume that we have a collection of linear mappings $D\colon \Omega^k(U) \to \Omega^{k+1}(U)$, satisfying properties (i)–(iii) in Proposition 41. Then, we necessarily have that $D = d$.

Proof: First, note that $D(dx^i) = D(Dx^i) = D^2x^i = 0$, in view of (i) and (iii). By induction and (ii), it follows that $D(dx^{i_1}\wedge\cdots\wedge dx^{i_k}) = 0$ for any k-tuple of indices $(i_1,\ldots,i_k)$. Now, as both d and D are linear, it suffices to show that they agree on "monomials" $f\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}$. And indeed, using (ii) and (i) again it follows that

$$\begin{aligned}
D(f\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}) &= Df\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k} \\
&= df\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k} = d(f\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}),
\end{aligned}$$

as wanted. Hence $D = d$. □

Theorem 22 (Definition of exterior derivative in smooth manifolds)

Let M be a smooth manifold. The **exterior derivative** is $d\colon \Omega^k(M) \to \Omega^{k+1}(M)$ is defined as follows: if $\omega \in \Omega^k(M)$ and $p \in M$, we let $(U; x^1,\ldots,x^n)$ be a chart for M centered at p, write

$$\omega = \sum_{1\le i_1<\cdots<i_k\le n} \omega_{i_1\ldots i_k}\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}$$

for suitable component functions $\omega_{i_1\ldots i_k}\colon U \to \mathbb{R}$, and set

$$(d\omega)_p = \sum_{1\le i_1<\cdots<i_k\le n} d(\omega_{i_1\ldots i_k})_p \wedge dx^{i_1}|_p\wedge\cdots\wedge dx^{i_k}|_p.$$

The resulting element $(d\omega)_p \in \mathcal{A}^{k+1}(T_pM)$ does not depend on the choice of chart $(U; x^1,\ldots,x^n)$ centered at p, and properties (i)–(iii) in Proposition 41 still hold.

Proof: With the notation as above, if $(\widetilde{U}, \widetilde{\varphi} = (\widetilde{x}^1, \ldots, \widetilde{x}^n))$ is a second chart centered at p and we write

$$\omega = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \widetilde{\omega}_{i_1 \ldots i_k} \, \mathrm{d}\widetilde{x}^{i_1} \wedge \cdots \wedge \mathrm{d}\widetilde{x}^{i_k},$$

consider the open neighborhood $W = \varphi(U) \cap \widetilde{\varphi}(\widetilde{U})$ of $\varphi(p) = \widetilde{\varphi}(p) = 0$ in $\mathbb{R}^n$. The restrictions to W of the unique exterior derivative operators in $\varphi(U)$ and $\widetilde{\varphi}(\widetilde{U})$ necessarily agree, by Proposition 42. This means that if we provisorily denote by $u^1, \ldots, u^n$ the Euclidean coordinate functions, we have that

$$\mathrm{d}\left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} (\omega_{i_1 \ldots i_k} \circ \varphi^{-1}) \, \mathrm{d}u^{i_1} \wedge \cdots \wedge \mathrm{d}u^{i_k} \right) = $$
$$= \mathrm{d}\left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} (\widetilde{\omega}_{i_1 \ldots i_k} \circ \widetilde{\varphi}^{-1}) \, \mathrm{d}u^{i_1} \wedge \cdots \wedge \mathrm{d}u^{i_k} \right)$$

in W. Evaluating it at 0 and moving the result back to $\mathscr{A}^{k+1}(T_pM)$, using the two isomorphisms $\mathscr{A}^{k+1}(\mathbb{R}^n) \cong \mathscr{A}^{k+1}(T_pM)$ induced by both charts, it follows that

$$\sum_{1 \leq i_1 < \cdots < i_k \leq n} \mathrm{d}(\omega_{i_1 \ldots i_k})_p \wedge \mathrm{d}x^{i_1}|_p \wedge \cdots \wedge \mathrm{d}x^{i_k}|_p =$$
$$= \sum_{1 \leq i_1 < \cdots < i_k \leq n} \mathrm{d}(\widetilde{\omega}_{i_1 \ldots i_k})_p \wedge \mathrm{d}\widetilde{x}^{i_1}|_p \wedge \cdots \wedge \mathrm{d}\widetilde{x}^{i_k}|_p.$$

Once this is in place, properties (i)–(iii) immediately follow from their corresponding versions in (iii). □

Remark. It is now also possible to show that the uniqueness established in Proposition 42 remains valid if we replace U with M. Proposition 41 remains valid for manifolds.

Example 101 (Area forms on regular surfaces in $\mathbb{R}^3$)

Let $M \subseteq \mathbb{R}^3$ be a regular surface, and assume that there exists a smooth unit normal vector field $\boldsymbol{N}$ along M, that is, for each point $p \in M$ we assign a vector $\boldsymbol{N}(p) \in \mathbb{R}^3$ such that $\|\boldsymbol{N}(p)\| = 1$ and $\langle \boldsymbol{N}(p), v \rangle = 0$ for every $v \in T_pM$; here, $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^3$. We may define $\omega \in \Omega^2(M)$ by setting $\omega_p(v, w) = \langle \boldsymbol{N}(p), v \times w \rangle$. In other words, $\omega_p(v, w)$ is the (oriented) area of the parallelogram in T_pM spanned by v and w, cf. Figure 64.

Generalizing the above for a regular hypersurface $M \subseteq \mathbb{R}^n$, we instead define $\omega \in \Omega^{n-1}(M)$ by $\omega_p(v_1, \ldots, v_{n-1}) = \det(\boldsymbol{N}(p), v_1, \ldots, v_{n-1})$. In either case, we have that $\mathrm{d}\omega = 0$ for dimensional reasons.

There are ways of computing exterior derivatives on manifolds without relying on coordinate systems. Here is the simplest case:

Proposition 43 (The exterior derivative of a 1-form)

Let M be a smooth manifold, and $\omega \in \Omega^1(M)$. Then

$$\mathrm{d}\omega(\mathbf{X},\mathbf{Y}) = \mathbf{X}(\omega(\mathbf{Y})) - \mathbf{Y}(\omega(\mathbf{X})) - \omega([\mathbf{X},\mathbf{Y}]),$$

for all vector fields $\mathbf{X},\mathbf{Y} \in \mathfrak{X}(M)$.
Here, $\mathbf{X}(\omega(\mathbf{Y}))$ is the function given by $p \mapsto \mathbf{X}_p[\omega(\mathbf{Y})]$, where $[\omega(\mathbf{Y})]$ denotes the germ at p of the function $p \mapsto \omega_p(\mathbf{Y}_p)$, while the Lie bracket $[\mathbf{X},\mathbf{Y}]$ is as in Exercise 98 (p. 141).

Proof: Let $(U; x^1,\ldots,x^n)$ be a chart on M, and write

$$\omega = \sum_{i=1}^n \omega_i\,\mathrm{d}x^i, \quad \mathbf{X} = \sum_{i=1}^n X^i \frac{\partial}{\partial x^i}, \quad \text{and} \quad \mathbf{Y} = \sum_{i=1}^n Y^i \frac{\partial}{\partial x^i},$$

where all indicated component functions are smooth. On one hand, we have that

$$\mathrm{d}\omega(\mathbf{X},\mathbf{Y}) = \sum_{i,j=1}^n \frac{\partial \omega_i}{\partial x^j}\mathrm{d}x^j \wedge \mathrm{d}x^i(\mathbf{X},\mathbf{Y}) = \sum_{i,j=1}^n \frac{\partial \omega_i}{\partial x^j}(X^jY^i - X^iY^j).$$

On the other hand,

$$\mathbf{X}(\omega(\mathbf{Y})) = \sum_{j=1}^n X^j \frac{\partial}{\partial x^j}\left(\sum_{i=1}^n \omega_i Y^i\right) = \sum_{i,j=1}^n \frac{\partial \omega_i}{\partial x^j} X^jY^i + \sum_{i,j=1}^n \omega_i X^j \frac{\partial Y^i}{\partial x^j},$$

and similarly for $\mathbf{Y}(\omega(\mathbf{X}))$. Putting all of it together, we have that

$$\mathbf{X}(\omega(\mathbf{Y})) - \mathbf{Y}(\omega(\mathbf{X})) = \mathrm{d}\omega(\mathbf{X},\mathbf{Y}) + \sum_{i=1}^n \omega_i\left(X^j\frac{\partial Y^i}{\partial x^j} - Y^j\frac{\partial X^i}{\partial x^j}\right).$$

By item (b) of Exercise 98, the last term above is exactly $\omega([\mathbf{X},\mathbf{Y}])$. As the chosen chart $(U; x^1,\ldots,x^n)$ was arbitrary, we are done. □

Remark. The generalization of Proposition 43 is called **Palais's formula**: given any $\omega \in \Omega^k(M)$, the exterior derivative $\mathrm{d}\omega \in \Omega^{k+1}(M)$ is given by

$$\begin{aligned}\mathrm{d}\omega(\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k) &= \sum_{i=0}^k (-1)^{k+1}\mathbf{X}_i(\omega(\mathbf{X}_0,\ldots,\widehat{\mathbf{X}_i},\ldots,\mathbf{X}_k)) \\ &\quad + \sum_{0\leq i<j\leq k} (-1)^{i+j}\omega([\mathbf{X}_i,\mathbf{X}_j],\mathbf{X}_0,\ldots,\widehat{\mathbf{X}_i},\ldots,\widehat{\mathbf{X}_j},\ldots,\mathbf{X}_k),\end{aligned}$$

for all $\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k \in \mathfrak{X}(M)$, where hats indicate omission (as in Exercise 103, p. 152). See [15, Proposition 14.32] for a proof (or try it yourself, if you're feeling daring).

Example 102 (Gluing differential forms on chart overlaps for $\mathbb{RP}^1$)

Consider the projective line $\mathbb{RP}^1$. Its standard atlas consists of the two charts $\varphi_0 \colon U_0 \to \mathbb{R}$ and $\varphi_1 \colon U_1 \to \mathbb{R}$ given by

$$\varphi_0([x:y]) = \frac{y}{x} \quad \text{and} \quad \varphi_1([x:y]) = \frac{x}{y},$$

where $U_0 = \{[x:y] \in \mathbb{RP}^1 : x \neq 0\}$ and $U_1 = \{[x:y] \in \mathbb{RP}^1 : y \neq 0\}$. Denoting the coordinates in the images of φ_0 and φ_1 by t and s, respectively, we have that $t = y/x$ and $s = x/y$, so that $\varphi_0 \circ \varphi_1^{-1} \colon \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$ is given by $s \mapsto t = 1/s$. Say that $\omega \in \Omega^1(\mathbb{RP}^n)$ is expressed relative to such charts as

$$\omega = f(t)\,\mathrm{d}t \quad \text{and} \quad \omega = g(s)\,\mathrm{d}s,$$

respectively. What must be the relation between the coefficient functions f and g on the overlap $\mathbb{R} \setminus \{0\}$? As $t = 1/s$ leads to $\mathrm{d}t = -\mathrm{d}s/s^2$, we must have that

$$g(s)\,\mathrm{d}s = f(t)\,\mathrm{d}t = f\left(\frac{1}{s}\right)\left(\frac{-\mathrm{d}s}{s^2}\right) = -\frac{1}{s^2} f\left(\frac{1}{s}\right)\mathrm{d}s \implies g(s) = -\frac{1}{s^2} f\left(\frac{1}{s}\right).$$

Conversely, this condition is also sufficient for one to "glue" the 1-forms $f(t)\,\mathrm{d}t$ and $g(s)\,\mathrm{d}s$ on $\mathbb{R} \setminus \{0\}$ into a 1-form ω on $\mathbb{RP}^1$. Namely, if $[x:y] \in \mathbb{RP}^1$ and $v \in T_{[x:y]}(\mathbb{RP}^1)$, one proceeds as follows: write

$$v = a\frac{\partial}{\partial t}\Big|_{[x:y]} \quad \text{or} \quad v = b\frac{\partial}{\partial s}\Big|_{[x:y]}$$

with $a, b \in \mathbb{R}$, according to whether $[x:y] \in U_0$ or $[x:y] \in U_1$, and set $\omega_{[x:y]}(v)$ to be equal to $af(y/x)$ or $bg(x/y)$, respectively. If $[x:y] \in U_0 \cap U_1$, the condition relating f and g implies that the definition is consistent, as $a = -by^2/x^2$ in view of Exercise 72 (p. 99), and so

$$af\left(\frac{y}{x}\right) = -b\frac{y^2}{x^2} f\left(\frac{y}{x}\right) = b\left(-\frac{1}{(y/x)^2} f\left(\frac{y}{x}\right)\right) = bg\left(\frac{x}{y}\right),$$

as required. (You could alternatively use the previous relation $\mathrm{d}t = -\mathrm{d}s/s^2$ with $a = \mathrm{d}t(v)$, $b = \mathrm{d}s(v)$, and $s = x/y$.)

A similar reasoning works for $\mathbb{RP}^n$ with $n > 1$ too, but the algebra gets considerably more cumbersome. A more efficient way to study differential forms on $\mathbb{RP}^n$ is to regard it as the quotient $\mathbb{S}^n/\mathbb{Z}_2$ and see which forms on $\mathbb{S}^n$ survive in the quotient. We will not pursue this now, and instead see a brief overview of how this works in a more general context (Theorem 29 in page 197 ahead).

In the above example, we related the two differential forms $f(t)\,\mathrm{d}t$ and $g(s)\,\mathrm{d}s$ on $\mathbb{R} \setminus \{0\}$ to one another, using the transition function $\varphi_0 \circ \varphi_1^{-1}$. However, a subtle point here is that we were dealing with two different copies of $\mathbb{R} \setminus \{0\}$. In general, given a smooth mapping $F\colon M \to N$ and some $\omega \in \Omega^k(N)$, is it possible to define a k-form on M instead, using F and ω? A moment of thought will tell that there is only one natural way to do it: we first push tangent vectors to M onto tangent vectors to N using $\mathrm{d}F$, and then evaluate ω at the images.

Definition 52 (Pullbacks of differential forms)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a smooth mapping. For each $\omega \in \Omega^k(N)$, we define the **pullback of** ω **under** F to be $F^*\omega \in \Omega^k(M)$ given by

$$(F^*\omega)_p(v_1, \ldots, v_n) = \omega_{F(p)}(\mathrm{d}F_p(v_1), \ldots, \mathrm{d}F_p(v_k)),$$

for all $p \in M$ and $v_1, \ldots, v_k \in T_pM$.

For instance, the condition relating f and g in Example 102 is equivalent to saying that $g(s)\,\mathrm{d}s$ equals the pullback of $f(t)\,\mathrm{d}t$ under $\varphi_0 \circ \varphi_1^{-1}$. Here's a very concrete example so you can get a feel for how pullbacks work:

Example 103

Let $F\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by $F(u,v) = (u^2 - v^2, uv, u+v+1)$, and consider

$$\omega = y\,\mathrm{d}x + z\,\mathrm{d}y + x\,\mathrm{d}z$$

in $\mathbb{R}^3$. How to compute the pullback $F^*\omega$ in practice? Following the prescription given by F, we set

$$x = u^2 - v^2, \quad y = uv, \quad z = u + v + 1,$$

and differentiate to obtain

$$\mathrm{d}x = 2u\,\mathrm{d}u - 2v\,\mathrm{d}v, \quad \mathrm{d}y = v\,\mathrm{d}u + u\,\mathrm{d}v, \quad \mathrm{d}z = \mathrm{d}u + \mathrm{d}v.$$

Now, plug everything into ω:

$$\begin{aligned} F^*\omega &= uv(2u\,\mathrm{d}u - 2v\,\mathrm{d}v) + (u+v+1)(v\,\mathrm{d}u + u\,\mathrm{d}v) + (u^2 - v^2)(\mathrm{d}u + \mathrm{d}v) \\ &= (2u^2v + (u+v+1)v + u^2 - v^2)\mathrm{d}u + (-2uv^2 + (u+v+1)u + u^2 - v^2)\mathrm{d}v \\ &= (2u^2v + uv + v + u^2)\mathrm{d}u + (-2uv^2 + 2u^2 + uv + u - v^2)\mathrm{d}v. \end{aligned}$$

Check your understanding so far:

Exercise 111

Compute the pullbacks $F^*\omega$ in the following cases:

(a) $\omega = f(u,v)\,\mathrm{d}u + g(u,v)\,\mathrm{d}v \in \Omega^1(\mathbb{R}^2)$,
with $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by $F(x,y,z) = (x,y)$.

(b) $\omega = dx \wedge dy \in \Omega^2(\mathbb{R}^2)$,
with $F\colon (0,\infty)\times(0,2\pi) \to \mathbb{R}^2$ given by $F(r,\theta) = (r\cos\theta, r\sin\theta)$.

(c) $\omega = dx \wedge dy \wedge dz \in \Omega^3(\mathbb{R}^3)$,
with $F\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $F(u,v) = (u,v,u^2+v^2)$.

We register next some general properties of pullbacks.

Proposition 44 (Algebraic properties of pullbacks)

Let M, N and P be smooth manifolds, $F\colon M \to N$ and $G\colon N \to P$ be a smooth mappings, and $k,\ell \geq 0$ be integers. Then:

(i) $F^*(\omega+\eta) = F^*\omega + F^*\eta$ for all $\omega,\eta \in \Omega^k(N)$.

(ii) $F^*(f\omega) = (f\circ F)F^*\omega$ for all $\omega \in \Omega^k(N)$ and $f \in C^\infty(N)$

(iii) $\mathrm{Id}_M^*\alpha = \alpha$ for every $\alpha \in \Omega^k(M)$,

(iv) $F^*G^*\zeta = (G\circ F)^*\zeta$ for every $\zeta \in \Omega^k(P)$.

(v) $F^*(\omega\wedge\eta) = F^*\omega \wedge F^*\eta$ for every $\omega \in \Omega^k(N)$ and $\eta \in \Omega^\ell(N)$.

Proof: Properties (i) and (ii) are clear from the definition of pullback, while (iii) easily follows from the relation $d(\mathrm{Id}_M)_p = \mathrm{Id}_{T_pM}$. We now justify (iv) for $k=1$, with the general case being an exercise in notation:

$$\begin{aligned}(F^*G^*\zeta)_p(v) &= (G^*\zeta)_{F(p)}(dF_p(v)) = \zeta_{G(F(p))}(dG_{F(p)}(dF_p(v))) \\ &= \zeta_{(G\circ F)(p)}(d(G\circ F)_p(v)) = ((G\circ F)^*\zeta)_p(v)\end{aligned}$$

Finally, we observe that both sides of (v) are bilinear in the variables ω and η, meaning that we may simply establish it with $\otimes$ instead of[11] $\wedge$. As done for (iv), we show how it is done in the case where $k=\ell=1$:

$$\begin{aligned}(F^*(\omega\otimes\eta))_p(v,w) &= (\omega\otimes\eta)_{F(p)}(dF_p(v), dF_p(w)) = \omega_{F(p)}(dF_p(v))\eta_{F(p)}(dF_p(w)) \\ &= (F^*\omega)_p(v)(F^*\eta)_p(w) = (F^*\omega\otimes F^*\eta)_p(v,w).\end{aligned}$$

□

Exercise 112

Verify items (iv) and (v) of the above result for general k and ℓ.

With these properties in place, we can establish another very important property of surjective submersions.

[11] One could argue that, with how things are written here, the pullback $F^*(\omega\otimes\eta)$ is not defined since $\omega\otimes\eta$ might fail to be alternating even if both ω and η are. This indeed is a correct observation, but the definition of pullback in fact makes sense for arbitrary covariant tensor fields Θ on M, that is, we smoothly assign to each $p \in M$ an element $\Theta_p \in \mathcal{T}^k(T_pM)$.

Proposition 45 (The pullback operation of a surjective submersion is injective)

Let M and N be smooth manifolds, and $F\colon M \to N$ be a surjective submersion. Then $F^*\colon \Omega^k(N) \to \Omega^k(M)$ is injective for each $k \geq 0$, that is, if $\omega, \eta \in \Omega^k(N)$ are such that $F^*\omega = F^*\eta$, then $\omega = \eta$.

Proof: Let $q \in N$ be arbitrary, and use Corollary 7 (p. 111) to fix an open neighborhood $U \subseteq N$ of q and a local section $\sigma\colon U \to M$ of F. As $F \circ \sigma = \mathrm{Id}_U$, we may use items (iii) and (iv) of Proposition 44 above: applying σ^* to both sides of the initial equality $F^*\omega = F^*\eta$ and evaluating the result at q yields $\omega_q = \eta_q$. As the point q was arbitrary, the argument is concluded. □

What Proposition 44 does not tell us is how pullbacks relate to exterior derivatives. This needs a separate statement:

Proposition 46

Let M and N be smooth manifolds, and $F\colon M \to N$ be any smooth mapping. Then, for any $\omega \in \Omega^k(N)$, we have that $F^*(\mathrm{d}\omega) = \mathrm{d}(F^*\omega)$. That is, pullbacks and exterior derivatives commute.

Proof: As both sides of the proposed relation are additive in the variable ω (but obviously not $C^\infty(M)$-linear), we may simply work with arbitrary chart $(U; x^1, \ldots, x^n)$ and $(V; y^1, \ldots, y^m)$ for M and N, and assume that $\omega = f\,\mathrm{d}y^{a_1} \wedge \cdots \wedge \mathrm{d}y^{a_k}$ is a monomial. When $k = 0$, the conclusion is clear.

We write, as before, $F^a = y^a \circ F$ for the components of F. On one hand, we have that

$$\mathrm{d}\omega = \sum_{b=1}^m \frac{\partial f}{\partial y^b}\,\mathrm{d}y^b \wedge \mathrm{d}y^{a_1} \wedge \cdots \wedge \mathrm{d}y^{a_k}$$

so that, by item (v) of Proposition 44, we have

$$\begin{aligned} F^*(\mathrm{d}\omega) &= \sum_{b=1}^m \left(\frac{\partial f}{\partial y^b} \circ F\right) \mathrm{d}F^b \wedge \mathrm{d}F^{a_1} \wedge \cdots \wedge \mathrm{d}F^{a_k} \\ &= \sum_{b=1}^m \left(\frac{\partial f}{\partial y^b} \circ F\right) \left(\sum_{j=1}^n \frac{\partial F^b}{\partial x^j}\,\mathrm{d}x^j\right) \wedge \left(\sum_{i_1=1}^n \frac{\partial F^{a_1}}{\partial x^{i_1}}\mathrm{d}x^{i_1}\right) \wedge \cdots \wedge \left(\sum_{i_k=1}^n \frac{\partial F^{a_k}}{\partial x^{i_k}}\mathrm{d}x^{i_k}\right) \\ &= \sum_{b=1}^m \sum_{i_1,\ldots,i_k,j=1}^n \left(\frac{\partial f}{\partial y^b} \circ F\right) \frac{\partial F^b}{\partial x^j} \frac{\partial F^{a_1}}{\partial x^{i_1}} \cdots \frac{\partial F^{a_k}}{\partial x^{i_k}}\,\mathrm{d}x^j \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}. \end{aligned} \tag{5.14}$$

On the other hand, again by item (v) of Proposition 44, we have that

$$F^*\omega = (f \circ F)\mathrm{d}F^{a_1} \wedge \cdots \wedge \mathrm{d}F^{a_k},$$

and hence

$$\begin{aligned} \mathrm{d}(F^*\omega) &= \mathrm{d}(f \circ F) \wedge \mathrm{d}F^{a_1} \wedge \cdots \wedge \mathrm{d}F^{a_k} \\ &\quad + \sum_{r=1}^k (f \circ F)\,\mathrm{d}F^{a_1} \wedge \cdots \wedge \mathrm{d}(\mathrm{d}F^{a_r}) \wedge \cdots \wedge \mathrm{d}F^{a_k}, \end{aligned} \tag{5.15}$$

as a consequence of item (ii) of Proposition 41. Now, the second sum in (5.15) vanishes by item (iii) of Proposition 41, and substituting

$$\frac{\partial}{\partial x^j}(f\circ F) = \left(\frac{\partial f}{\partial y^b}\circ F\right)\frac{\partial F^b}{\partial x^j}$$

into (5.15) finally yields

$$\mathrm{d}(F^*\omega) = \sum_{b=1}^{m}\sum_{i_1,\ldots,i_k,j=1}^{n}\left(\frac{\partial f}{\partial y^b}\circ F\right)\frac{\partial F^b}{\partial x^j}\frac{\partial F^{a_1}}{\partial x^{i_1}}\cdots\frac{\partial F^{a_k}}{\partial x^{i_k}}\,\mathrm{d}x^j\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}$$

which agrees with (5.14), as required. □

We are on our way to generalizing the Fundamental Theorem of Calculus:

$$\int_a^b f(x)\,\mathrm{d}x = F(b)-F(a),$$

where $f\colon [a,b]\to\mathbb{R}$ is a continuous function, and F is an antiderivative of f. The integrand $f(x)\,\mathrm{d}x$ will certainly get replaced with a differential form, but what about the integral? Before we develop some integration theory for manifolds, there is one issue we must address: the formula above involves the values of F at the points $x=a$ and $x=b$, and so they must be included in the interval over which the integration is happening. In other words, even though singletons have measure zero, we should formally regard the integration as being carried over the closed interval $[a,b]$ instead of the open interval (a,b). Well, $[a,b]$ does not have any natural smooth structure (and it is not a submanifold of the real line $\mathbb{R}$ either). The workaround is to extend our definition of manifold.

5.3 Manifolds with boundary

Topological manifolds are locally modeled in Euclidean space $\mathbb{R}^n$. If we want to include things such as a half-sphere $\{(x,y,z)\in\mathbb{R}^3 : x^2+y^2+z^2=1 \text{ and } z\geq 0\}$ in our theory, it becomes natural to consider manifolds "with boundary". The standard way to do this is by introducing

the **closed half-space** of dimension n, that is, $\mathscr{H}^n = \mathbb{R}^{n-1}\times[0,\infty)$, (5.16)

cf. Figure 66. It is equipped with its subspace topology induced from $\mathbb{R}^n$. In Differential Geometry, the open half-space $\mathbb{R}^{n-1}\times(0,\infty)$ is usually denoted by $\mathbb{H}^n$, with $\mathbb{H}$ having the same blackboard font as $\mathbb{R}$ in $\mathbb{R}^n$. Thus we changed the font from $\mathbb{H}$ to $\mathscr{H}$ in order to avoid any confusion. We can now mirror Definitions 24 and 25 (p. 68).

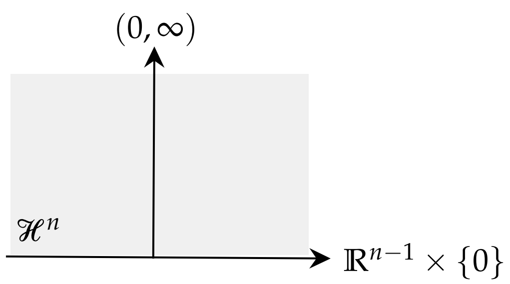


Figure 66: The closed half-space $\mathscr{H}^n$.

Definition 53 (Topological manifold with boundary)

A **n-dimensional topological manifold with boundary** is a topological space M which is Hausdorff, second-countable, and **locally** $\mathscr{H}^n$: for every $p \in M$ there is an open neighborhood $U \subseteq M$ of p and a homeomorphism $\varphi\colon U \to \varphi(U) \subseteq \mathscr{H}^n$, where $\varphi(U)$ is open in $\mathscr{H}^n$. Denoting the topological interior and boundary of $\mathscr{H}^n$ by $\mathring{\mathscr{H}}^n$ and $\partial\mathscr{H}^n$, the pair (U,φ) is called an **interior-chart** if $\varphi(U) \subseteq \mathring{\mathscr{H}}^n$, and a **boundary-chart** if $\varphi(U) \cap \partial\mathscr{H}^n \neq \varnothing$. The integer n is called the **dimension** of M.

Example 104

The closed half-space $\mathscr{H}^n$ itself is a topological manifold with boundary: the identity function $\mathrm{Id}_{\mathscr{H}^n}\colon \mathscr{H}^n \to \mathscr{H}^n$ serves as a global boundary-chart.

Example 105

The closed half-sphere $M = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1 \text{ and } z \geq 0\}$ mentioned above is a topological manifold with boundary: we can get away with four boundary-charts, naturally obtained by restricting the ones in Exercise 54 (p. 74).

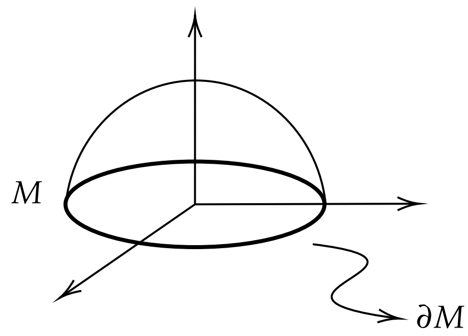


Figure 67: Half-spheres are topological manifolds with boundary.

Example 106

The closed disk

$$\mathbb{D}^2 = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 1\}$$

is a topological manifold with boundary: the identity chart on the open disk takes care of all points other than in the circle, whereas modified polar coordinates $(x,y) \mapsto (\theta, 1-r)$ (on the appropriate domain) are boundary-charts around points in $\mathbb{S}^1$.

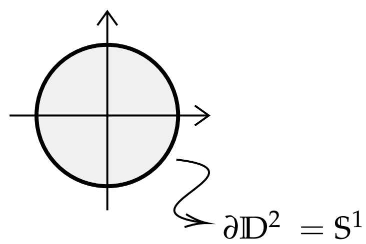


Figure 68: The closed unit disk as a manifold with boundary.

There are two issues to address with this definition: how to make sense of smoothness between chart-transitions, and how to consistently sort out points of M into interior points or boundary points.

We first elaborate on the second issue: if $p \in M$, the most natural thing to do is to say that p is an **interior point** of M if there is an interior-chart (U, φ) for M around p, and that p is a **boundary point** of M if there is a boundary-chart (U, φ) for M around p for which $\varphi(p) \in \partial\mathscr{H}^n$. It is not at all obvious, although it turns out to be true, that p cannot simultaneously be an interior point and a boundary point. The proof of this fact for smooth manifolds with boundary, to be defined ahead, turns out to be more reasonable—at the topological level, homological tools enter the playground. In any case:

Definition 54 (Interior and boundary)

Let M be a topological manifold with boundary. The **interior** and **boundary** of M are defined as the subsets $\mathrm{Int}(M) = \{p \in M : p \text{ is an interior point of } M\}$ and $\partial M = \{p \in M : p \text{ is a boundary point of } M\}$, respectively.

Example 107

If a topological manifold with boundary M is a subspace of a larger topological space X, the "manifold-boundary" of M in the sense of Definition 54 above does not need to agree with the topological boundary of M as a subset of X.

Consider for instance the half-sphere M from Example 105. The topological boundary of M as a subset of $\mathbb{R}^3$ is just M itself, but its manifold-boundary is the circle $\mathbb{S}^1 \times \{0\}$. In a similar manner, the topological interior of M is empty, while its manifold-interior equals $\mathbb{S}^2 \cap \mathring{\mathscr{H}}^2 = M \setminus (\mathbb{S}^1 \times \{0\}) \neq \varnothing$.

As for the first issue: the boundary $\partial\mathscr{H}^n$ is not an open subset of $\mathbb{R}^n$, so what does it mean for a function defined on $\partial\mathscr{H}^n$ (or, more generally, on a subset of it) to be smooth?

Definition 55

Let $S \subseteq \mathbb{R}^n$ be any set, and $f\colon S \to \mathbb{R}$ be any function. We say that f is **smooth** on S if, for every $p \in S$, there is an open neighborhood $U \subseteq \mathbb{R}^n$ of p and a smooth function $F\colon U \to \mathbb{R}$ such that $F|_{U\cap S} = f|_{U\cap S}$.

If the subset S in the above definition is already open in $\mathbb{R}^n$, this new definition of smoothness agrees with the classical definition of Euclidean-smoothness, which is a local notion.

With this in place, we may put down the definition we need to proceed:

Definition 56 (Smooth manifolds with boundary)

Let M be a topological manifold with boundary. Two charts (U, φ) and (V, ψ) are called C^∞-compatible if either $U \cap V = \varnothing$, or $U \cap V \neq \varnothing$ and

$$\psi \circ \varphi^{-1} \colon \varphi(U \cap V) \to \psi(U \cap V) \quad \text{and} \quad \varphi \circ \psi^{-1} \colon \psi(U \cap V) \to \varphi(U \cap V)$$

are smooth functions between open subsets of $\mathscr{H}^n$, in the sense of Definition 55. A **smooth manifold with boundary** is a topological manifold with boundary M equipped with a maximal atlas of C^∞-compatible charts.

As before, any C^∞-atlas is contained in a maximal one.

Theorem 23

Let M be a smooth manifold with boundary, and $n = \dim M$. The boundary of ∂M can be made into a smooth manifold without boundary, with $\dim \partial M = n - 1$.

Proof: For every boundary-chart (U, φ) for M, let $\varphi_{\partial M} \colon U \cap \partial M \to \varphi(U) \cap \partial\mathscr{H}^n$ be the restriction of φ to $U \cap \partial M$. Note that the intersection $\varphi(U) \cap \partial\mathscr{H}^n$ is an open subset of $\partial\mathscr{H}^n = \mathbb{R}^{n-1} \times \{0\} \cong \mathbb{R}^{n-1}$. If (V, ψ) is a second boundary-chart for M, the transition $\psi_{\partial M} \circ \varphi_{\partial M}^{-1} \colon \varphi(U \cap V) \cap \partial\mathscr{H}^n \to \psi(U \cap V) \cap \partial\mathscr{H}^n$ is a smooth mapping between open subsets of $\mathbb{R}^{n-1}$, being a suitable restriction of $\psi \circ \varphi^{-1}$. □

Virtually everything we have seen so far for smooth manifolds also works out for smooth manifolds with boundary. Namely, the definition of smooth functions and smooth mappings (Definitions 30 and 31, pp. 82–83) makes sense and, consequently, so does the definition of germs of smooth functions (Definition 33, p. 95). This allows us to define tangent spaces as being the spaces of derivations of germs (Definition 35, p. 97), and also derivatives of smooth mappings between manifolds (Definition 36, p. 101).

Here is one example:

Example 108

Consider the half-plane $\mathscr{H}^2$. For each $p \in \mathscr{H}^2$, we will have that $T_p(\mathscr{H}^2) \cong \mathbb{R}^2$, because $\dim \mathscr{H}^2 = 2$, this will not change. If $p \in \operatorname{Int}(\mathscr{H}^2)$, then $T_p(\mathscr{H}^2)$ can be seen as the space of all vectors in $\mathbb{R}^2$ starting at the point p. If $p \in \partial\mathscr{H}^2$ instead, the same applies, and tangent vectors are still allowed to point outside of $\mathscr{H}^2$, that is, even if its endpoint does not correspond to a point in $\mathscr{H}^2$.

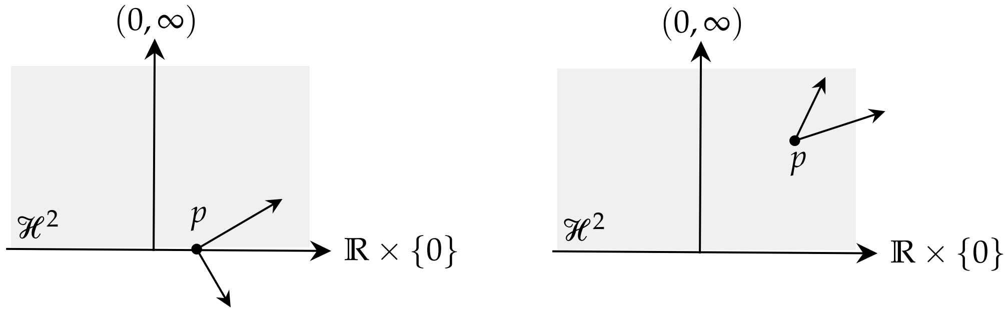


Figure 69: The tangent space $T_p(\mathscr{H}^2)$ for $p \in \mathrm{Int}(\mathscr{H}^2)$ and $p \in \partial\mathscr{H}^2$.

The tangent space $T_p(\partial\mathscr{H}^2)$ for $p \in \partial\mathscr{H}^2$, being 1-dimensional, is the subspace of $T_p(\mathscr{H}^2)$ consisting of the vectors which are horizontal.

The above example suggests that the characterization of tangent spaces in terms of velocity vectors of curves (cf. Lemma 7, p. 105) will need a slight modification. Namely, if $p \in \partial M$ and $\alpha\colon (-\varepsilon,\varepsilon) \to M$ has $\alpha(0) = p$, we will have that $\alpha(t) \in M$ either for all $t \in [0,\varepsilon)$, or for all $t \in (-\varepsilon, 0]$, but α must leave M if $\alpha'(0) \in T_pM \smallsetminus T_p(\partial M)$.

The difference $T_pM \smallsetminus T_p(\partial M)$ has two connected components, but in this setting we are actually able to distinguish them. Namely, let $(U; x^1, \ldots, x^n)$ be a boundary-chart centered at p. As $U \cap \partial M$ is described as $x^1 = \cdots = x^{n-1} = 0$, it follows that

$$\left\{ \frac{\partial}{\partial x^1}\bigg|_p, \ldots, \frac{\partial}{\partial x^{n-1}}\bigg|_p \right\} \text{ is a basis of the tangent space } T_p(\partial M).$$

This means that

$$v = \sum_{i=1}^n a^i \frac{\partial}{\partial x^n}\bigg|_p \in T_p(\partial M) \iff a^n = 0,$$

meaning that the connected components of $T_pM \smallsetminus T_p(\partial M)$ consist of all the $v \in T_pM$ for which $a^n > 0$, or for which $a^n < 0$, respectively. See Figure 70.

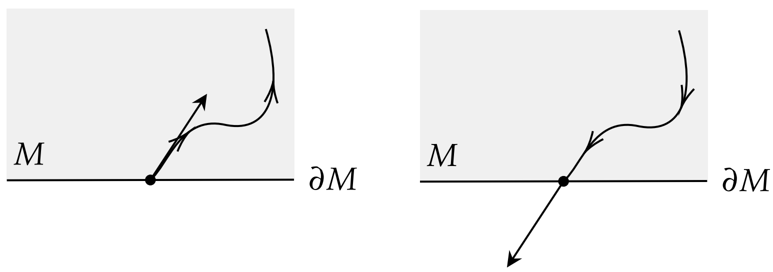


Figure 70: Inward- and outward-pointing vectors tangent to ∂M.

Before we turn this into a formal definition and declare that v is inward-pointing if $a^n > 0$ and outward-pointing if $a^n < 0$, we must verify that this condition is independent on the choice of boundary-chart. Namely, assume that $(V; y^1, \ldots, y^n)$ is another boundary-chart centered at p, and write

$$v = \sum_{i=1}^n b^i \frac{\partial}{\partial y^i}\bigg|_p.$$

Exercise 72 (p. 99) tells us that

$$b^n = \sum_{i=1}^{n} a^i \frac{\partial y^n}{\partial x^i}(p),$$

where $y^i = y^i(x^1, \ldots, x^n)$ are the components of the transition mapping between the boundary-charts. As $y^n(x^1, \ldots, x^{n-1}, 0) = 0$ for all $x^1, \ldots, x^{n-1}$, it follows that $(\partial y^n/\partial x^i)(p) = 0$ for each $i = 1, \ldots, n-1$, and hence $b^n = a^n(\partial y^n/\partial x^n)(p)$. At the same time, $y^n(0, \ldots, 0, x^n) \geq y^n(0, \ldots, 0, 0)$ for all $x^n \geq 0$ sufficiently small, with equality if and only if $x^n = 0$, so that $(\partial y^n/\partial x^n)(p) > 0$. We conclude that a^n and b^n have the same sign.

Definition 57

Let M be a smooth manifold with boundary, and $p \in \partial M$ be any point. A tangent vector $v \in T_pM \smallsetminus T_p(\partial M)$ is called **inward-pointing** (resp., **outward-pointing**) if for some (and hence every) boundary-chart $(U; x^1, \ldots, x^n)$ around p, we have that $\mathrm{d}x^n|_p(v) > 0$ (resp., $\mathrm{d}x^n|_p(v) < 0$).

Exercise 113

Let M be a smooth manifold with boundary, $p \in \partial M$, and $v \in T_pM \smallsetminus T_p(\partial M)$. Revisit the proof of Lemma 7 (p. 105) and take inspiration from Figure 70 to show that:

(a) v is inward-pointing if and only if there is a smooth curve $\alpha\colon [0, \varepsilon) \to M$ with $\alpha(0) = p$ and $\alpha'(0) = v$.

(b) v is outward-pointing if and only if there is a smooth curve $\alpha\colon (-\varepsilon, 0] \to M$ with $\alpha(0) = p$ and $\alpha'(0) = v$.

There is another elegant way of rephrasing all of this: a **boundary-definining** function is a smooth function $f\colon M \to [0, \infty)$ such that $f^{-1}(0) = \partial M$ and $\mathrm{d}f_p \neq 0$ for every $p \in \partial M$. For example, whenever $(U; x^1, \ldots, x^n)$ is a boundary-chart, x^n is a local boundary-defining function, but it is possible to prove that there is always a global defining-function. Then $v \in T_pM \smallsetminus T_p(\partial M)$ is inward-pointing (resp. outward-pointing) if and only if $\mathrm{d}f_p(v) > 0$ (resp., $\mathrm{d}f_p(v) < 0$).

5.4 Orientability

From this point onwards, all manifolds are allowed to have boundary.

Here, we start by recalling the notion of orientability for finite-dimensional real vector spaces. Namely, whenever V is such a vector space and we write $n = \dim V$, it holds that for any two ordered bases $\mathcal{B} = (e_1, \ldots, e_n)$ and $\widetilde{\mathcal{B}} = (\widetilde{e}_1, \ldots, \widetilde{e}_n)$ there is a transition matrix $A = [a^i_j]_{i,j=1}^n$ such that $\widetilde{e}_j = \sum_{i=1}^n a^i_j e_i$, for each $j = 1, \ldots, n$. The matrix A is necessarily nonsingular, and so we define an equivalence relation $\sim$ on the set of all bases for V by declaring that $\mathscr{B} \sim \widetilde{\mathscr{B}}$ if $\det A > 0$. There are exactly two equivalence classes for $\sim$, which are called **orientations for** V. Once one of them has been chosen, the bases belonging to it are called **positive**—all others are called **negative**.

The vector space $\mathbb{R}^n$ has a canonical orientation, determined by its standard basis. The determinant function (seen, like always, as an alternating multilinear function of its columns), helps us sort out all bases $(v_1, \ldots, v_n)$ for $\mathbb{R}^n$ into positive ones and negative ones, according to whether $\det(v_1, \ldots, v_n)$ is positive or negative, respectively. Parallelepipeds spanned by positive bases have positive "oriented volume", and this is something the theory of integration on manifolds has to take into account.

With the above being said, our immediate next objective is to consider orientations for all the tangent spaces T_pM of whatever manifold M is under consideration. This is not to be done haphazardly, but instead "smoothly" (or, at least "continuously"). Unfortunately, it is not always possible to do so. We will have good reason to restrict our attention to the manifolds where a consistent choice of orientations for the tangent spaces can be made.

Definition 58 (Orientability of manifolds)

Let M be a smooth manifold, with $n = \dim M$. We say that M is **orientable** if there exists a nowhere-vanishing top-degree differential form $\mu \in \Omega^n(M)$, i.e., if for every $p \in M$, the element $\mu_p \in \mathscr{A}^n(T_pM)$ is not the zero tensor. In this context, such μ is called a **volume form** (or, sometimes, an **orientation form**) and the pair (M, μ) is an **oriented manifold**.

The volume form μ in Definition 58 above plays the role of the determinant function in $\mathbb{R}^n$: for each $p \in M$, a basis $(v_1, \ldots, v_n)$ of T_pM is declared by μ to be positive if $\mu_p(v_1, \ldots, v_n) > 0$, and negative if $\mu_p(v_1, \ldots, v_n) < 0$. The fact that $\mu_p \neq 0$ implies that $\mu_p(v_1, \ldots, v_n) = 0$ can only happen when $v_1, \ldots, v_n$ are linearly dependent. The name "volume form" is also justified—μ is a device that allows us to compute oriented volumes of parallelepipeds in the tangent spaces to M (that is, infinitesimally). A chart $(U; x^1, \ldots, x^n)$ is then called positively-oriented if, for each $p \in U$, the basis $\partial/\partial x^1|_p, \ldots, \partial/\partial x^n|_p$ of T_pM is positive in the above sense, and an atlas for M is said to be positively-oriented if all of its charts are positively-oriented.

Exercise 114 (Transitions between positive charts have positive Jacobian)

With the above setting, show that if two positively-oriented charts (U, φ) and (V, ψ) have $U \cap V$, then $\psi \circ \varphi^{-1} \colon \varphi(U \cap V) \to \psi(U \cap V)$ satisfies the condition that $\det \mathrm{D}(\psi \circ \varphi^{-1})(x) > 0$ for every $x \in \varphi(U \cap V)$.

Example 109

The **standard volume form** for the manifold $M = \mathbb{R}^n$, in terms of its global coordinates $x^1, \ldots, x^n$, is $\mu = \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$. That is, for each $p \in M$ we have that $\mu_p = \det$, under the isomorphism $T_p(\mathbb{R}^n) \cong \mathbb{R}^n$.

Exercise 115

Show that the product of two orientable smooth manifolds is again orientable.

The result below allows us to obtain many more examples of orientable manifolds:

Proposition 47 (Orientability of transversely-framed submanifolds)

Let M be an orientable smooth manifold, and $S \subseteq M$ be an embedded k-dimensional submanifold. Assume that there are $n-k$ vector fields $\mathbf{X}_1, \ldots, \mathbf{X}_{n-k}$ tangent to M defined along S[a] and such that

$$T_pM = T_pS \oplus \mathbb{R}(\mathbf{X}_1)_p \oplus \cdots \oplus \mathbb{R}(\mathbf{X}_{n-k})_p, \quad \text{for all } p \in S. \tag{5.17}$$

Then, S is orientable as well.

[a] A vector field $\mathbf{Y}$ tangent to M along S assigns to each $p \in S$ a tangent vector $\mathbf{Y}_p \in T_pM$ which is not necessarily in T_pS, and $\mathbf{Y}$ is not necessarily defined outside of S either.

Proof: From a volume form $\mu \in \Omega^n(M)$, we define a volume form $\mu_S \in \Omega^k(S)$ as follows: for each point $p \in S$ and vectors $v_1, \ldots, v_k \in T_pS$, we set

$$(\mu_S)_p(v_1, \ldots, v_k) = \mu_p((\mathbf{X}_1)_p, \ldots, (\mathbf{X}_{n-k})_p, v_1, \ldots, v_k).$$

Smoothness of μ_S follows from the one of μ and of the vector fields $\mathbf{X}_1, \ldots, \mathbf{X}_{n-k}$. Then we necessarily have that $(\mu_S)_p \neq 0$: if $v_1, \ldots, v_k$ form a basis for T_pS, then $(\mu_S)_p(v_1, \ldots, v_k) \neq 0$ because $(\mathbf{X}_1)_p, \ldots, (\mathbf{X}_{n-k})_p, v_1, \ldots, v_k$ is a basis for T_pM (due to (5.17)) and $\mu_p \neq 0$. □

Example 110

Let $F: U \subseteq \mathbb{R}^n \to \mathbb{R}^{n-k}$ be a smooth function, and $c \in \mathbb{R}^{n-k}$ be a regular value of F, so that $F^{-1}(c)$ is an embedded k-dimensional submanifold of $\mathbb{R}^n$. Then, writing $F = (F^1, \ldots, F^{n-k})$, we have that

$$\mathbb{R}^n = \ker \mathrm{d}F_p \oplus \mathbb{R}\nabla F^1(p) \oplus \cdots \mathbb{R}\nabla F^{n-k}(p) \tag{5.18}$$

for each $p \in F^{-1}(c)$, where $\nabla F^i(p)$ denotes the Euclidean gradient of the component F^i, for each $i = 1, \ldots, n-k$. Indeed, $\nabla F^1(p), \ldots, \nabla F^{n-k}(p)$ are linearly independent as c is a regular value of F, and hence form a basis for the orthogonal complement of $\ker \mathrm{d}F_p$ (computed with the standard inner product of $\mathbb{R}^n$). Proposition 47 now says that $F^{-1}(c)$ is orientable.

In particular, spheres are orientable (Example 81, p. 126)

Remark. In the example above, the open subset $U \subseteq \mathbb{R}^n$ can be replaced with any orientable smooth manifold M. To make the same argument work, one must choose a

Riemannian metric g on M, that is, a smooth choice of inner products g_p in each tangent space T_pM. Replacing the gradients of each F^i with the vectors corresponding to $\mathrm{d}F^i_p \in T^*_pM$ under the isomorphisms $T_pM \to T^*_pM$ provided by g, a relation analogous to (5.18) holds.

Example 111

If M and N are smooth manifolds, with M orientable, and $F\colon M \to N$ is a smooth function, the graph $\mathrm{Gr}(F) = \{(p,q) \in M \times N : q = F(p)\}$ is orientable. Indeed, it is diffeomorphic to M itself under $M \ni p \mapsto (p, F(p)) \in \mathrm{Gr}(F)$, which is orientable.

If an orientable smooth manifold M has a nonempty boundary ∂M, then ∂M inherits an orientation from M in a rather natural way. For now, we take for granted the existence of an outward-pointing vector field $\mathbf{X}$ tangent to M along ∂M (see Proposition 50 ahead).

Proposition 48

If M is an orientable smooth manifold with boundary, $\mu \in \Omega^n(M)$ is a volume form, and $\mathbf{X}$ is an outward-pointing vector field tangent to M along ∂M, then $\mu_{\partial M} = \mu(\mathbf{X}, \cdot, \cdots, \cdot)$ defines a volume form on ∂M.

Proof: We directly apply Proposition 47 with $S = \partial M$, noting that the decomposition $T_pM = T_p(\partial M) \oplus \mathbb{R}\mathbf{X}_p$ holds for every $p \in \partial M$. □

Above, using an inward-pointing vector field along the boundary works equally well for inducing an orientation in ∂M—it will be the opposite orientation to the one provided by Proposition 48. It is a matter of convention, and we stick to the one that uses the orientation induced by an outward-pointing vector field.

It will also be important to understand when a given smooth mapping preserves orientation or not. If (M, μ) and (N, ν) are oriented n-dimensional manifolds and $F\colon M \to N$ is a smooth mapping, then $F^*\nu \in \Omega^n(M)$, and therefore $F^*\nu$ must be a function multiple of μ. We write $F^*\nu = J(F)\mu$, where $J(F)\colon M \to \mathbb{R}$ is called the **Jacobian** of F relative to μ and ν. The name "Jacobian" should not be surprising: choosing coordinates $(U; x^1, \ldots, x^n)$ and $(V; y^1, \ldots, y^n)$ on M and N for which the volume forms are expressed as $\mu = \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ and $\nu = \mathrm{d}y^1 \wedge \cdots \wedge \mathrm{d}y^n$—this is always possible[12]—a direct adaptation of Proposition 39 (p. 156) gives us that

$$J(F) = \det\left[\frac{\partial F^a}{\partial x^j}\right]_{a,j=1,\ldots,n},$$

where $F^a = y^a \circ F$ as usual. As a consequence of the Inverse Function Theorem, F is a local diffeomorphism if and only if $J(F)$ is nowhere-vanishing. In this case, assuming

[12]Let $(U; \hat{x}^1, x^2 \ldots, x^n)$ be any chart, and write $\mu = \phi\, \mathrm{d}\hat{x}^1 \wedge \mathrm{d}x^2 \cdots \wedge \mathrm{d}x^n$ for some suitable nowhere-vanishing function ϕ. If x^1 is any function such that $\partial x^1/\partial \hat{x}^1 = \phi$, then $\mu = \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \cdots \wedge \mathrm{d}x^n$ as required.

that M is connected, $J(F)$ must be either positive everywhere or negative everywhere. We are led to the definition below:

Definition 59

Let (M,μ) and (N,ν) be oriented manifolds, and $F\colon M \to N$ be a local diffeomorphism. We say that F is **orientation-preserving** if $J(F) > 0$, and **orientation-reversing** if $J(F) < 0$.

For mappings between open subsets of $\mathbb{R}^n$, there are no secrets: compute the Jacobian matrix, and take its determinant. Outside of $\mathbb{R}^n$, things may get more subtle.

Exercise 116

Consider the unit sphere $\mathbb{S}^n$, equipped with its standard orientation induced from closed unit ball in $\mathbb{R}^{n+1}$. Is the antipodal mapping $\tau\colon \mathbb{S}^n \to \mathbb{S}^n$ orientation-preserving or not? Does your answer depend on the parity of n?

In Exercise 115, we have seen that the product of two orientable manifolds is orientable. And what about quotients?

Proposition 49

Let M be an oriented smooth manifold, and Γ be a finite group of diffeomorphisms of M acting freely on M, cf. Exercise 68 (p. 91). Denote by $\pi\colon M \to M/\Gamma$ the quotient projection. Then M/Γ is orientable if and only if each $\gamma \in \Gamma$ is orientation-preserving, in which case π also becomes orientation-preserving.

We proceed.

5.5 From integration in $\mathbb{R}^n$ to integration on manifolds

In this section, we assume familiarity with Riemann integration in $\mathbb{R}^n$, and refer to [24, Chapter 3] or [23, Section 23.1] for more details. A bounded function is Riemann-integrable if and only if its set of discontinuities is of measure zero. In particular, compactly-supported continuous functions are always integrable.

We first generalize the notion of support. Namely, given a smooth manifold M and a differential form $\omega \in \Omega^k(M)$, the **support** of ω is defined to be the subset $\operatorname{supp}\omega = \overline{\{p \in M : \omega_p \neq 0\}}$ of M; here, the bar denotes closure. Then ω is said to be **compactly-supported** if its support is a compact subset of M, and we finally set $\Omega^k_c(M) = \{\omega \in \Omega^k(M) : \omega \text{ is compactly-supported}\}$. The support of a smooth function is included in this discussion, with $k = 0$. To integrate differential forms, we start with open subsets of $\mathbb{R}^n$.

Definition 60 (Integral of n-forms on $\mathbb{R}^n$)

If $U \subseteq \mathbb{R}^n$ is an open subset and $\omega \in \Omega^n_c(U)$ is a differential n-form, written as $\omega = f\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$, we define **the integral of ω over U** as

$$\int_U \omega = \int_U f(x^1, \ldots, x^n)\,\mathrm{d}x^1 \cdots \mathrm{d}x^n,$$

where in the right side we have a classical Riemann integral.

Note here that while we essentially replace $f(x^1, \ldots, x^n)\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ with the expression $f(x^1, \ldots, x^n)\,\mathrm{d}x^1 \cdots \mathrm{d}x^n$ before integrating, the order of the differentials $\mathrm{d}x^1, \ldots, \mathrm{d}x^n$ matters. For instance, if $\omega = f\,\mathrm{d}y \wedge \mathrm{d}x \in \Omega^2_c(\mathbb{R}^2)$, we have that

$$\int_{\mathbb{R}^2} \omega = \int_{\mathbb{R}^2} -f(x, y)\,\mathrm{d}x\,\mathrm{d}y,$$

as $\omega = -f\,\mathrm{d}x \wedge \mathrm{d}y$, where in the right side we have a classical Riemann integral.

In addition, the **change-of-variables** formula for multivariate integrals may be phrased with the language of derivatives as linear transformations: if $U, V \subseteq \mathbb{R}^n$ are open subsets, $f\colon V \to \mathbb{R}$ is continuous and compactly-supported, and $T\colon U \to V$ is a diffeomorphism, then

$$\int_V f(y)\,\mathrm{d}y = \int_U f(T(x))|\det DT(x)|\,\mathrm{d}x \tag{5.19}$$

holds, where $\mathrm{d}y$ and $\mathrm{d}x$ are shorthands for $\mathrm{d}y^1 \cdots \mathrm{d}y^n$ and $\mathrm{d}x^1 \cdots \mathrm{d}x^n$, of course. At the same time, we have that $T^*(\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n) = \det DT\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ (by Proposition 39), without the absolute-value signs. If we naively follow the usual idea of using charts to "transplant" the above notion of integration onto manifolds, we might run into consistency problems involving signs. Exercise 114 says that if we assume that M is orientable, fix an orientation, and work only with positively-oriented charts, things will be well-defined. Case in point:

Definition 61 (Integral of n-forms on chart domains)

Let M be an oriented n-dimensional smooth manifold, and $\omega \in \Omega^n(M)$. If (U, φ) is a positively-oriented chart for M, **the integral of ω over U** is defined to be

$$\int_U \omega = \int_{\varphi(U)} (\varphi^{-1})^*\omega,$$

where the integral on the right side is taken in the sense of Definition 60.

By Exercise 114 and formula (5.19), it follows that

$$\int_{\varphi(U)} (\varphi^{-1})^*\omega = \int_{\psi(U)} (\psi^{-1})^*\omega$$

whenever (U, ψ) is another positively-oriented chart with the same domain U (we simply take $T = \psi \circ \varphi^{-1}$). This makes $\int_U \omega$ independent on the choice of chart.

Generalizing this to the entire manifold M, which in general cannot be covered by a single chart, requires a different tool.

Definition 62 (Partition of unity)

Let X be a topological space, and $\{U_\alpha\}_{\alpha\in A}$ be an open cover of X. A **partition of unity** subordinate to $\{U_\alpha\}_{\alpha\in A}$ is a collection $\{\rho_\alpha\}_{\alpha\in A} \subseteq C^0(X)$ of non-negative functions such that:

(i) $0 \leq \rho_\alpha \leq 1$ for all $\alpha \in A$;

(ii) $\operatorname{supp} \rho_\alpha \subseteq U_\alpha$ for all $\alpha \in A$;

(iii) $\{\operatorname{supp} \rho_\alpha\}_{\alpha\in A}$ is locally finite and $\sum_{\alpha\in A} \rho_\alpha(x) = 1$ for all $x \in X$.

The collection of supports being locally finite means that every $x \in X$ has an open neighborhood which intersects only finitely many supports, and this condition ensures that $\sum_{\alpha\in A} \rho_\alpha(x)$ is always a finite sum—there are no convergence issues to worry about here.

Partitions of unity do not necessarily exist on arbitrary topological spaces. Here, we are only interested in smooth manifolds. The second-countability condition included in the definition of a topological manifold is finally used with full strength[13]:

Theorem 24

Let M be a smooth manifold, and $\{U_\alpha\}_{\alpha\in A}$ be an open cover of M. Then, there is a smooth partition of unity subordinate to $\{U_\alpha\}_{\alpha\in A}$.

See [15, Theorem 2.23] for a proof. We are now in position to justify why an outward-pointing vector field always exists along ∂M:

Proposition 50

Let M be a smooth manifold with boundary. Then, there exists an outward-pointing vector field $\mathbf{X}$ along ∂M.

Proof: Let $\{(U_\alpha, \varphi_\alpha)\}_{\alpha\in A}$ be an atlas of boundary-charts, covering ∂M, and let $\{\rho_\alpha\}_{\alpha\in A}$ be a partition of unity subordinate to $\{U_\alpha\}_{\alpha\in A}$. Assume that for each chart $(U_\alpha, \varphi_\alpha)$, the coordinate vector field $\partial/\partial x^n_\alpha$ is outward-pointing along $U_\alpha \cap \partial M$. Then

$$\mathbf{X} = \sum_{\alpha\in A} \rho_\alpha \frac{\partial}{\partial x^n_\alpha}$$

[13] Under the guise of *paracompactness*: a topological space X is called **paracompact** if every open cover has a locally finite open refinement. That is, if whenever $\{U_\alpha\}_{\alpha\in A}$ is an open cover of X, there is another open cover $\{V_\beta\}_{\beta\in B}$ with the two properties (i) for every $\beta \in B$ there is $\alpha \in A$ such that $V_\beta \subseteq U_\alpha$, and (ii) every $p \in X$ has an open neighborhood W for which $\{\beta \in B : V_\beta \cap W \neq \varnothing\}$ is finite.

defines an outward-pointing vector field along ∂M (the sum is always finite, and positive-coefficients linear combinations of outward-pointing vectors are again outward-pointing). $\square$

Back to the matter at hand, using again the idea that partitions of unity allow us to "patch" locally defined objects into global ones, we may now define integrals of differential forms over the entire manifold M:

Definition 63 (Integrals of differential forms over oriented manifolds)

Let M be an oriented smooth manifold, with $n = \dim M$, and $\omega \in \Omega^n_c(M)$. We define the **integral of** ω **over** M as

$$\int_M \omega = \sum_{\alpha \in A} \int_{U_\alpha} \rho_\alpha \omega, \tag{5.20}$$

where $\{(U_\alpha, \varphi_\alpha)\}_{\alpha\in A}$ is a positively-oriented atlas for M, $\{\rho_\alpha\}_{\alpha\in A}$ is a partition of unity subordinate to the open cover $\{U_\alpha\}_{\alpha\in A}$, and the integral of each $\rho_\alpha\omega$ over U_α is as in Definition 61.

Of course, to validate Definition 63, we must show that the choices of oriented atlas $\{(U_\alpha, \varphi_\alpha)\}_{\alpha\in A}$ and partition of unity $\{\rho_\alpha\}_{\alpha\in A}$ are immaterial. So, assume that $\{(V_\beta, \psi_\beta)\}_{\beta\in B}$ and $\{\chi_\beta\}_{\beta\in B}$ are different choices, and compute

$$\begin{aligned}\sum_{\alpha\in A}\int_{U_\alpha}\rho_\alpha\omega &\stackrel{(1)}{=} \sum_{\alpha\in A}\int_{U_\alpha}\rho_\alpha\left(\sum_{\beta\in B}\chi_\beta\omega\right) \stackrel{(2)}{=} \sum_{\alpha\in A}\sum_{\beta\in B}\int_{U_\alpha}\rho_\alpha\chi_\beta\omega \\ &\stackrel{(3)}{=} \sum_{\alpha\in A}\sum_{\beta\in B}\int_{U_\alpha\cap V_\beta}\rho_\alpha\chi_\beta\omega \stackrel{(4)}{=} \sum_{\beta\in B}\int_{V_\beta}\chi_\beta\omega,\end{aligned} \tag{5.21}$$

where each step is explained as follows:

(1) $\sum_{\beta\in B}\chi_\beta = 1$ and $\omega = 1\omega$;

(2) linearity of integrals;

(3) $\int_{U_\alpha}\rho_\alpha\chi_\beta\omega = \int_{U_\alpha\cap V_\beta}\rho_\alpha\chi_\beta\omega$, as $\operatorname{supp}(\rho_\alpha\chi_\beta) \subseteq U_\alpha\cap V_\beta$;

(4) "symmetry" (that is, we apply the previous steps in the reverse order and switching the roles of U_α and ρ_α with V_β and χ_β.

In practice, computing these integrals basically amounts to finding a parameterization F (that is, the inverse of a chart) which covers all of M (except perhaps for a set of measure zero), computing the pullback $F^*\omega$, and integrating the resulting expression over the domain of F (which is an open subset of some Euclidean space)—this latter integral becoming a classical Riemann integral, cf. Definition 60. Here is an example of how this works:

Example 112

Consider $\omega \in \Omega^2(\mathbb{S}^2)$ given by

$$\omega = \begin{cases} \dfrac{\mathrm{d}y \wedge \mathrm{d}z}{x}, & \text{if } x \neq 0 \\ \dfrac{\mathrm{d}z \wedge \mathrm{d}x}{y}, & \text{if } y \neq 0 \\ \dfrac{\mathrm{d}x \wedge \mathrm{d}y}{z}, & \text{if } z \neq 0 \end{cases}$$

First, note that ω is indeed well-defined, as differentiating $x^2 + y^2 + z^2 = 1$ leads to $x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z = 0$, and hence $x\,\mathrm{d}x \wedge \mathrm{d}y + z\,\mathrm{d}z \wedge \mathrm{d}y = 0$ finally implies that $(\mathrm{d}x \wedge \mathrm{d}y)/z = (\mathrm{d}y \wedge \mathrm{d}z)/x$ when $x, z \neq 0$; similar calculations shows that ω is well-defined on all other possible overlaps.

Next, we compute the integral $\int_{\mathbb{S}^2} \omega$. For this, we consider $F\colon [0,\pi] \times [0, 2\pi] \to \mathbb{S}^2$ given by $F(\varphi, \theta) = (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$, so that

$$F^*x = \sin\varphi\cos\theta, \quad F^*y = \sin\varphi\sin\theta, \quad \text{and} \quad F^*z = \cos\varphi.$$

Using Proposition 46, it follows that

$$\begin{aligned} F^*(\mathrm{d}x) &= \cos\varphi\cos\theta\,\mathrm{d}\varphi - \sin\varphi\sin\theta\,\mathrm{d}\theta \\ F^*(\mathrm{d}y) &= \cos\varphi\sin\theta\,\mathrm{d}\varphi + \sin\varphi\cos\theta\,\mathrm{d}\theta \\ F^*(\mathrm{d}z) &= -\sin\varphi\,\mathrm{d}\varphi. \end{aligned} \tag{5.22}$$

With this in place, we may show that

$$F^*\left(\frac{\mathrm{d}y \wedge \mathrm{d}z}{x}\right) = F^*\left(\frac{\mathrm{d}z \wedge \mathrm{d}x}{y}\right) = F^*\left(\frac{\mathrm{d}x \wedge \mathrm{d}y}{z}\right) = \sin\varphi\,\mathrm{d}\varphi \wedge \mathrm{d}\theta, \tag{5.23}$$

whenever all expressions make sense. In other words, $F^*\omega = \sin\varphi\,\mathrm{d}\varphi \wedge \mathrm{d}\theta$ in all cases, and so

$$\int_{\mathbb{S}^2} \omega = \int_0^{2\pi}\int_0^{\pi} \sin\varphi\,\mathrm{d}\varphi\,\mathrm{d}\theta = 4\pi.$$

This is valid because, up to a set of measure zero, F defines a chart for $\mathbb{S}^2$.

Exercise 117

Using (5.22), verify (5.23) in detail.

Try mimicking what was done above, in a different surface, to get a feeling for it.

Exercise 118

Consider a piece of paraboloid, $M = \{(x,y,z) \in \mathbb{R}^3 : z = x^2 + y^2 \text{ and } 0 < z < 1\}$.

(a) Show that $\omega \in \Omega^2(M)$ given by

$$\omega = \begin{cases} \dfrac{\mathrm{d}x \wedge \mathrm{d}z}{y}, & \text{if } y \neq 0, \\ \dfrac{\mathrm{d}z \wedge \mathrm{d}y}{x}, & \text{if } x \neq 0, \end{cases}$$

is well-defined.

(b) Compute $\displaystyle\int_M \omega$.

Hint: Use a parameterization $F\colon (0,1) \times (0,2\pi) \to M$.

5.6 The Stokes formula and applications

Theorem 25 (Stokes)

Let M be an oriented smooth manifold, and let $n = \dim M$. Then, we have that

$$\int_M \mathrm{d}\omega = \int_{\partial M} \omega \tag{5.24}$$

for every $\omega \in \Omega^n_c(M)$.

Remark. When M has no boundary, the right side of (5.24) is understood to vanish.

Proof: The proof consists of two main steps: reducing (5.24) to the case where $M = \mathbb{R}^n$ or $M = \mathscr{H}^n$, and then showing that (5.24) holds for $M = \mathbb{R}^n$ and $M = \mathscr{H}^n$.

So, assume that the theorem is valid in $\mathbb{R}^n$ and $\mathscr{H}^n$, and choose a positively-oriented atlas $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ such that each image $\varphi(U_\alpha)$ is diffeomorphic to either $\mathbb{R}^n$ or $\mathscr{H}^n$ (this is a direct adaptation of Exercise 53, p. 71). In particular, the theorem is valid for each U_α. Now, choosing a partition of unity $\{\rho_\alpha\}_{\alpha \in A}$ subordinate to the open cover $\{U_\alpha\}_{\alpha \in A}$, and noting that $\partial M \cap U_\alpha = \partial U_\alpha$, we compute:

$$\begin{aligned} \int_{\partial M} \omega &\overset{(1)}{=} \int_{\partial M} \sum_{\alpha \in A} \rho_\alpha \omega \overset{(2)}{=} \sum_{\alpha \in A} \int_{\partial M} \rho_\alpha \omega \overset{(3)}{=} \sum_{\alpha \in A} \int_{\partial M \cap U_\alpha} \rho_\alpha \omega \\ &\overset{(4)}{=} \sum_{\alpha \in A} \int_{\partial U_\alpha} \rho_\alpha \omega \overset{(5)}{=} \sum_{\alpha \in A} \int_{U_\alpha} \mathrm{d}(\rho_\alpha \omega) \overset{(6)}{=} \sum_{\alpha \in A} \int_M \mathrm{d}(\rho_\alpha \omega) \\ &\overset{(2)}{=} \int_M \sum_{\alpha \in A} \mathrm{d}(\rho_\alpha \omega) \overset{(7)}{=} \int_M \mathrm{d}\left(\sum_{\alpha \in A} \rho_\alpha \omega\right) \overset{(1)}{=} \int_M \mathrm{d}\omega, \end{aligned} \tag{5.25}$$

where each equality is explained as follows:

(1) $\sum_{\alpha\in A}\rho_\alpha = 1$ and $\omega = 1\omega$;

(2) linearity of integrals;

(3) $\operatorname{supp}(\rho_\alpha\omega|_{\partial M}) \subseteq \partial M \cap U_\alpha$;

(4) the previously mentioned equality $\partial M \cap U_\alpha = \partial U_\alpha$;

(5) the Stokes formula in U_α;

(6) $\operatorname{supp} \mathrm{d}(\rho_\alpha\omega) \subseteq U_\alpha$;

(7) linearity of the exterior derivative.

This concludes the first step.

Now, to illustrate the main ideas involved in the rest of the calculation, we prove the theorem for $M = \mathcal{H}^2$. Write $\omega = f\,\mathrm{d}x + g\,\mathrm{d}y$, and assume that the supports of f and g are contained in the interior of the rectangle $[-a,a]\times[0,a]$, for some $a>0$. As $\mathrm{d}\omega = (g_x - f_y)\,\mathrm{d}x\wedge\mathrm{d}y$, we have that

$$\begin{aligned}\int_{\mathcal{H}^2}\mathrm{d}\omega &= \int_0^{+\infty}\int_{-\infty}^{+\infty}(g_x - f_y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \int_0^{+\infty}\int_{-\infty}^{+\infty} g_x\,\mathrm{d}x\,\mathrm{d}y - \int_{-\infty}^{+\infty}\int_0^{+\infty} f_y\,\mathrm{d}y\,\mathrm{d}x\\ &= \int_0^a\left(\int_{-a}^a g_x\,\mathrm{d}x\right)\mathrm{d}y - \int_{-a}^a\left(\int_0^a f_y\,\mathrm{d}y\right)\mathrm{d}x\\ &= \int_0^a g(a,y) - g(-a,y)\,\mathrm{d}y - \int_{-a}^a f(x,a) - f(x,0)\,\mathrm{d}x.\end{aligned} \tag{5.26}$$

From our support assumption, we have that $g(a,y) = g(-a,y) = f(x,a) = 0$. If we were dealing with $\mathbb{R}^2$ instead of $\mathcal{H}^2$, considering the full rectangle $[-a,a]\times[-a,a]$, we would have $f(x,-a)$ instead of $f(x,0)$, with $f(x,-a) = 0$ as well, establishing (5.24) for $\mathbb{R}^2$. But as we consider $\mathcal{H}^2$, this last term survives and we obtain

$$\int_{\mathcal{H}^2}\mathrm{d}\omega = \int_{-a}^a f(x,0)\,\mathrm{d}x.$$

On the other hand, we have that $y = 0$ along $\partial\mathcal{H}^2$, so that ω restricted to $\partial\mathcal{H}^2$ reads simply as $\omega = f(x,0)\,\mathrm{d}x$. We now observe that $-\partial/\partial y$ is an outward vector field to $\mathcal{H}^2$ along $\partial\mathcal{H}^2$, so that the orientation form induced on $\partial\mathcal{H}^2$ by $\mathrm{d}x\wedge\mathrm{d}y \in \Omega^2(\mathcal{H}^2)$ is simply $(\mathrm{d}x\wedge\mathrm{d}y)(-\partial/\partial y,\cdot) = \mathrm{d}x$. For this reason, we have that

$$\int_{\partial\mathcal{H}^2}\omega = \int_{-\infty}^{\infty} f(x,0)\,\mathrm{d}x = \int_{-a}^a f(x,0)\,\mathrm{d}x.$$

This concludes the proof. □

Exercise 119

Give a direct proof of Stokes's theorem for $M = \mathscr{H}^n$, mimicking (5.26).

Stokes's Theorem is the ultimate generalization of the Fundamental Theorem of Calculus. To explain how this happens, recall the result and notation of Theorem 21.

Example 113 (The fundamental theorem of line integrals)

Given a vector field $\mathbf{F}\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ which is conservative, that is, for which there is a smooth function $\varphi\colon U \to \mathbb{R}$ such that $\mathbf{F} = \nabla\varphi$, it holds that

$$\int_C \langle \mathbf{F}, \mathbf{T}\rangle \, \mathrm{d}s = \varphi(q) - \varphi(p)$$

for any points $p, q \in U$ and curve $C \subseteq U$ starting at p and ending at q, where $\mathbf{T}$ is the (positively-oriented) unit tangent vector along C, and $\mathrm{d}s$ is the arclength element of C. Classically, the left side is computed as

$$\int_C \langle \mathbf{F}, \mathbf{T}\rangle \, \mathrm{d}s = \int_a^b \langle \mathbf{F}(\mathbf{r}(t)), \mathbf{r}'(t)\rangle \, \mathrm{d}t,$$

with the result being independent on the choice of parametrization $\mathbf{r}\colon [a,b] \to C$ with $\mathbf{r}(a) = p$ and $\mathbf{r}(b) = q$. In our setting, the manifold being considered is C, with manifold-boundary $\partial C = \{p, q\}$. The orientation in C (from p to q) induces the orientation in ∂C which assigns $+1$ to q and -1 to p. A volume-form for C is nothing more than $\mathrm{d}s$ itself, expressed as $\mathrm{d}s = \|\mathbf{r}'(t)\| \, \mathrm{d}t$ in terms of a parameterization $\mathbf{r}$, while $\mathbf{T}(t) = \mathbf{r}'(t)/\|\mathbf{r}'(t)\|$, with both $-\mathbf{T}(a)$ and $\mathbf{T}(b)$ outward-pointing, so that $\mathrm{d}s(-\mathbf{T}(a)) = -1$ and $\mathrm{d}s(\mathbf{T}(b)) = +1$. Finally, the 1-form $\langle \mathbf{F}, \mathbf{T}\rangle \, \mathrm{d}s$ on C is nothing more than $\mathrm{d}\varphi$, with

$$\int_C \mathrm{d}\varphi = \int_{\partial C} \varphi = \varphi(q) - \varphi(p)$$

in view of Stokes's theorem.

Example 114 (The Green-Stokes theorem)

Let $\Sigma \subseteq \mathbb{R}^3$ be a regular surface, with smooth boundary $C = \partial\Sigma$ and area form $\mathrm{d}\Sigma \in \Omega^2(M)$ (this notation is used classically, even though $\mathrm{d}\Sigma$ is by no means an exact form). The Green-Stokes theorem states that

$$\oint_C \langle \mathbf{F}, \mathbf{T}\rangle \, \mathrm{d}s = \iint_\Sigma \langle \operatorname{curl} \mathbf{F}, \mathbf{N}\rangle \, \mathrm{d}\Sigma$$

for every vector field $\mathbf{F}$ defined on some open neighborhood $U \subseteq \mathbb{R}^3$ of Σ, where the orientation of C is taken to be induced by the one of Σ, and $\mathbf{T}$ and $\mathrm{d}s$ are as in

Example 113 above. If Σ is a planar region $R \subseteq \mathbb{R}^2$ and $\mathbf{F} = (P, Q)$, this formula reads simply as

$$\oint_C P\,\mathrm{d}x + Q\,\mathrm{d}y = \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \mathrm{d}x\,\mathrm{d}y.$$

In any case, this is again a consequence of Stokes's theorem, as $\langle \mathbf{F}, \mathbf{T}\rangle\,\mathrm{d}s = \alpha_{\mathbf{F}}$ leads to $\mathrm{d}(\langle \mathbf{F}, \mathbf{T}\rangle\,\mathrm{d}s) = \mathrm{d}\alpha_{\mathbf{F}} = \beta_{\operatorname{curl}\mathbf{F}} = \langle \operatorname{curl}\mathbf{F}, \mathbf{N}\rangle\,\mathrm{d}\Sigma$.

Example 115 (Gauss's divergence theorem)

Let $V \subseteq \mathbb{R}^3$ be a compact region in space enclosed by a closed surface $\Sigma = \partial V$, and let $\mathbf{F}$ be a vector field defined on some open neighborhood $U \subseteq \mathbb{R}^3$ of V. Then Gauss's divergence theorem states that

$$\oiint_\Sigma \langle \mathbf{F}, \mathbf{N}\rangle\,\mathrm{d}\Sigma = \iiint_V \operatorname{div}\mathbf{F}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z,$$

where $\mathbf{N}$ and $\mathrm{d}\Sigma$ are as in Example 114, and the orientation of Σ is taken to be the one induced by the one of V. Indeed, as

$$\mathrm{d}(\langle \mathbf{F}, \mathbf{N}\rangle\,\mathrm{d}\Sigma) = \mathrm{d}\beta_{\mathbf{F}} = \gamma_{\operatorname{div}\mathbf{F}} = \operatorname{div}\mathbf{F}\,\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z,$$

we may simply apply Stokes's theorem again.

5.7 A teeny bit of de Rham cohomology

Some of the fundamental ideas in Mathematics revolve around the concept of an "invariant". Namely, given a class of objects to be considered (vector spaces, groups, topological spaces, manifolds, etc.), we assign to each object a simpler object or quantity (often a single number), in such a way that equivalent (isomorphic, homeomorphic, diffeomorphic) objects have the same associated quantity—hence the name "invariant". This means that objects having distinct invariants cannot be equivalent.

For example, to each finite-dimensional vector space we assign a non-negative integer—its dimension—and isomorphic spaces have the same dimension. Therefore, finite-dimensional vector spaces having different dimensions cannot be isomorphic. In this case, the dimension turns out to be a "complete" invariant: the converse holds and two finite-dimensional vector spaces having the same dimension must be isomorphic.

Finding useful invariants (let alone complete ones) is, in general, a very difficult task. The idea behind de Rham cohomology is to associate to each smooth manifold a finite sequence of vector spaces. And these vector spaces will ultimately arise from differential forms.

Definition 64 (Closed and exact forms)

Let M be a smooth manifold, and $k \geq 0$ be any integer. Then $\omega \in \Omega^k(M)$ is called:

- **closed**, if $\mathrm{d}\omega = 0$;
- **exact**, if $\omega = \mathrm{d}\eta$ for some $\eta \in \Omega^{k-1}(M)$.

We denote by $Z^k(M)$ and $B^k(M)$ the subspaces of $\Omega^k(M)$ consisting of all closed and exact k-forms, respectively.

As there are no differential forms of degree -1, we take $B^0(M)$ to be the zero vector space. At the other extreme, if $n = \dim M$, every n-form is closed for dimensional reasons, so that $Z^n(M) = \Omega^n(M)$. Similarly, if $k > n$, then $Z^k(M) = \Omega^k(M) = \{0\}$. The property $\mathrm{d}^2 = 0$ directly yields the inclusions $B^k(M) \subseteq Z^k(M)$, for each degree $k = 0, \ldots, n$. This allows us to write the following definition:

Definition 65 (de Rham cohomology)

Let M be a smooth manifold, and $k \geq 0$ be any integer. The k-th **de Rham cohomology space** of M is defined to be the quotient space $H^k_{\mathrm{dR}}(M) = Z^k(M)/B^k(M)$. The equivalence class of an element $\omega \in Z^k(M)$ in $H^k_{\mathrm{dR}}(M)$ is denoted by $[\omega]$.

Remark. The vector spaces $H^k_{\mathrm{dR}}(M)$ are often called the de Rham cohomology *groups* of M. The reason for this is that there are more general notions of "cohomology", which make sense for arbitrary topological spaces, but such "cohomology spaces" are not really vector spaces, only groups. The fact that $H^k_{\mathrm{dR}}(M)$ in particular has a vector space structure is somewhat special.

The definition of quotient vector spaces immediately tells us what each $H^k_{\mathrm{dR}}(M)$ is measuring: how badly closed k-forms fail to be exact. Or, alternatively, if we consider $\omega = \mathrm{d}\eta$ as an equation to be solved for η, the de Rham cohomology controls whether solutions exist or not, and how non-unique they can be. What makes it all interesting is that these obstructions to exactness must be of global nature, and relating to the topology of M. Locally, everything is trivial:

Theorem 26 (Poincaré Lemma)

Let M be a smooth manifold, $k \geq 1$ be an integer, and $\omega \in \Omega^k(M)$ be closed. Then, for every $p \in M$ there is an open neighborhood $U \subseteq M$ of p and $\eta \in \Omega^{k-1}(U)$ such that $\omega = \mathrm{d}\eta$ on U.

See [24, Theorem 4.11] for a proof. The zeroth cohomology is easy to understand, but it doesn't really tell us anything new.

Proposition 51

Let M be a smooth manifold. Then $H^0_{\rm dR}(M) \cong \mathbb{R}^m$, where $m \geq 1$ is the number of connected components of M.

Proof: Since $B^0(M) = \{0\}$, we have that $H^0_{\rm dR}(M) = Z^0(M)$ is the space of closed 0-forms. But a 0-form is simply a smooth function $f\colon M \to \mathbb{R}$, and closedness means that $\mathrm{d}f = 0$. This implies that f is locally constant, and the conclusion follows: sending $[f] \in H^0_{\rm dR}(M)$ to the vector in $\mathbb{R}^m$ consisting of the constant values of f on the m connected components of M defines an isomorphism (once an ordering of the components is fixed, of course). $\square$

Example 116 (The cohomology of the real line)

Consider the real line $M = \mathbb{R}$. By Proposition 51, $H^0_{\rm dR}(M) \cong \mathbb{R}$. It then remains to determine $H^1_{\rm dR}(\mathbb{R})$, but we claim that $H^1_{\rm dR}(\mathbb{R}) = \{0\}$. In other words, every 1-form on $\mathbb{R}$ is exact. Namely, if $\omega \in \Omega^1(\mathbb{R})$ is arbitrary, we may write it as $\omega = f\,\mathrm{d}x$ for some $f \in C^\infty(\mathbb{R})$, and define $F\colon \mathbb{R} \to \mathbb{R}$ by

$$F(x) = \int_0^x f(t)\,\mathrm{d}t.$$

Then F is also smooth and has $F'(x) = f(x)$, so that $\mathrm{d}F = \omega$.

The proof of the Poincaré Lemma consists in taking a chart centered at the given point p, constructing η in a starshaped neighborhood (cf. Definition 21, p. 57) of 0 in $\mathbb{R}^n$, and then pulling it back to the chart domain. When $M = \mathbb{R}^n$, such starshaped neighborhood of 0 can be taken to be the entire space $\mathbb{R}^n$. In other words, in the case where $M = \mathbb{R}^n$, the conclusion of the Poincaré Lemma is global, and we in fact have $H^0_{\rm dR}(\mathbb{R}^n) \cong \mathbb{R}$ and $H^k_{\rm dR}(\mathbb{R}^n) = \{0\}$ for every $k \geq 1$.

Example 117 (The cohomology of the circle)

Now, we consider the circle $M = \mathbb{S}^1$. As before, $H^0_{\rm dR}(\mathbb{S}^1) \cong \mathbb{R}$ and it remains to determine $H^1_{\rm dR}(\mathbb{S}^1)$.

This time, we claim that $H^1_{\rm dR}(\mathbb{S}^1) \cong \mathbb{R}$. To establish this, we regard integration over $\mathbb{S}^1$ as a linear functional

$$\int_{\mathbb{S}^1}\colon Z^1(\mathbb{S}^1) \to \mathbb{R}.$$

The desired conclusion will follow once we show that

$$\text{(i) } \int_{\mathbb{S}^1} \text{ is surjective,} \quad \text{and} \quad \text{(ii) has kernel equal to } B^1(\mathbb{S}^1). \tag{5.27}$$

Taking $\omega = -y\,\mathrm{d}x + x\,\mathrm{d}y \in \Omega^1(\mathbb{S}^1)$ and the parametrization $\gamma\colon (0, 2\pi) \to \mathbb{S}^1$

given by $\gamma(t) = (\cos t, \sin t)$, we may compute[a]

$$\int_{\mathbb{S}^1} \omega = \int_0^{2\pi} \omega_{\gamma(t)}(\gamma'(t))\,\mathrm{d}t = \int_0^{2\pi} (-\sin t)(-\sin t) + (\cos t)(\cos t)\,\mathrm{d}t = 2\pi$$

and conclude that (5.27-i) holds.

Stokes' theorem implies that $B^1(\mathbb{S}^1)$ is contained in the kernel of the integration functional (as $\mathbb{S}^1$ has no boundary). For the reverse inclusion, let $\alpha \in \Omega^1(\mathbb{S}^1)$ have $\int_{\mathbb{S}^1} \alpha = 0$, and write it as $\alpha = f\omega$ for some $f \in C^\infty(\mathbb{S}^1)$—this is possible because ω is nowhere-vanishing, and thus ω_p spans $T_p^*(\mathbb{S}^1)$ for each $p \in \mathbb{S}^1$. Identifying $\mathbb{S}^1$ with the quotient $\mathbb{R}/(2\pi\mathbb{Z})$, and letting $\pi\colon \mathbb{R} \to \mathbb{R}/(2\pi\mathbb{Z})$ denote the quotient projection, we have that $f \circ \pi \in C^\infty(\mathbb{R})$ is in fact a 2π-periodic function with

$$\int_0^{2\pi} f(\pi(t))\,\mathrm{d}t = 0. \tag{5.28}$$

We may once again define a smooth function $F\colon \mathbb{R} \to \mathbb{R}$ by

$$F(t) = \int_0^t f(\pi(s))\,\mathrm{d}s,$$

and then compute

$$\frac{\mathrm{d}}{\mathrm{d}t}(F(t+2\pi) - F(t)) = f(\pi(t+2\pi)) - f(\pi(t)) = 0.$$

This means that there is $c \in \mathbb{R}$ such that $F(t+2\pi) = F(t) + c$, for every $t \in \mathbb{R}$, but choosing $t = 0$ and using (5.28) we obtain that $c = 0$. It now follows that there is $\bar{F} \in C^\infty(\mathbb{S}^1)$ such that $\bar{F} \circ \pi = F$ (smoothness of $\bar{F}$ follows from the one of F via the characteristic property of π as a surjective submersion). As the final step, we argue that $\mathrm{d}\bar{F} = \alpha$. Namely, this follows from the easily-verified relation $\pi^*\omega = \mathrm{d}t$, which implies that

$$\pi^*(\mathrm{d}\bar{F}) = \mathrm{d}(\bar{F} \circ \pi) = \mathrm{d}F = F'(t)\,\mathrm{d}t = (f \circ \pi)(t)\pi^*\omega = \pi^*(f\omega) = \pi^*\alpha,$$

together with Proposition 45. Therefore, (5.27-ii) holds as well.

[a] The image of γ in $\mathbb{S}^1$ misses just a point, which has measure zero—the integral is not affected.

Exercise 120

In $\mathbb{R}^2 \smallsetminus \{0\}$, consider the 1-form

$$\omega = \frac{x\,\mathrm{d}y - y\,\mathrm{d}x}{x^2 + y^2}.$$

Show that $\mathrm{d}\omega = 0$, but $\int_{\mathbb{S}^1} \omega \neq 0$. Conclude that $[\omega] \neq 0$ in $H^1_{\mathrm{dR}}(\mathbb{R}^2 \smallsetminus \{0\})$.

As it turns out, we can still obtain more from the idea of regarding integration as a linear functional:

Proposition 52

Let M be a compact and oriented n-dimensional smooth manifold without boundary. The integration functional

$$\int_M : Z^n(M) \to \mathbb{R} \tag{5.29}$$

is surjective and has kernel equal to $B^n(M)$, therefore inducing an isomorphism $H^n_{\mathrm{dR}}(M) \cong \mathbb{R}$.

The majority of the argument is rather simple. As M is orientable, we may fix a volume form $\omega \in \Omega^n(M)$, while compactness of M ensures that $\int_M \omega$ is finite and nonzero, making (5.29) surjective. At the same time, Stokes's Theorem ensures that $B^n(M)$ is contained in the kernel of (5.29), as M is without boundary. The proof of the reverse inclusion—that is, that if an n-form integrates to zero it must be exact—is more subtle. You can see the detailed argument in [15, Theorem 17.31], as well as the companion result: if M is compact and non-orientable, then $H^n_{\mathrm{dR}}(M) = \{0\}$.

Example 118

Cohomology has very explicit meanings in the setting of classical vector calculus; the key is, of course, Theorem 21. Let $U \subseteq \mathbb{R}^3$ be open, and $\mathbf{X}$ be a vector field on U. Then:

(a) If $H^1_{\mathrm{dR}}(U) = \{0\}$, $\operatorname{curl} \mathbf{X} = 0$ implies that $\mathbf{X} = \nabla f$ for some function f.

(b) If $H^2_{\mathrm{dR}}(U) = \{0\}$, $\operatorname{div} \mathbf{X} = 0$ implies that $\mathbf{X} = \operatorname{curl} \mathbf{Y}$ for some vector field $\mathbf{Y}$.

We may now justify why the de Rham cohomology is a differential invariant, that is, why diffeomorphic manifolds have isomorphic de Rham cohomology spaces.

Proposition 53

Let M and N be smooth manifolds, and $F\colon M \to N$ be a smooth mapping. For each $k \geq 0$, the induced pullback operation $F^*\colon \Omega^k(N) \to \Omega^k(M)$ survives in the quotient as a linear mapping (also denoted by)

$$F^*\colon H^k_{\mathrm{dR}}(N) \to H^k_{\mathrm{dR}}(M), \text{ given by } F^*[\omega] \doteq [F^*\omega]. \tag{5.30}$$

This construction has the two properties below:

(a) If P is a third smooth manifold and $G\colon N \to P$ is a second smooth mapping, then we have that $G^* \circ F^* = (F \circ G)^*$ as linear mappings $H^k_{\mathrm{dR}}(P) \to H^k_{\mathrm{dR}}(M)$.

(b) $(\mathrm{Id}_M)^* = \mathrm{Id}_{H^k_{\mathrm{dR}}(M)}$.

In particular, if $F\colon M \to N$ is a diffeomorphism, then $F^*\colon H^k_{\mathrm{dR}}(N) \to H^k_{\mathrm{dR}}(M)$ is an isomorphism, with inverse $(F^*)^{-1} = (F^{-1})^*$.

Proof: What we must really prove here is that F^* in (5.30) is indeed well-defined. Once this is in place, items (a) and (b) are trivial consequences of items (iii) and (iv) in Proposition 44 (p. 169), while linearity of (5.30) is also clear. To show that (5.30) is well-defined, it suffices to show that $F^*(Z^k(N)) \subseteq Z^k(M)$ and $F^*(B^k(N)) \subseteq B^k(M)$ for each $k \geq 0$. This, in turn, follows directly from Proposition 46: if $\omega \in Z^k(N)$, then $F^*\omega \in Z^k(M)$ because $\mathrm{d}(F^*\omega) = F^*(\mathrm{d}\omega) = F^*(0) = 0$, while for any $\eta \in \Omega^{k-1}(N)$ we have that $F^*(\mathrm{d}\eta) = \mathrm{d}(F^*\eta) \in B^k(M)$. □

The invariance property of the de Rham cohomology is actually much more refined than what the previous result suggests. It is possible to show that manifolds which are **homotopy equivalent**[14] already have isomorphic de Rham cohomology spaces; see [15, Theorem 17.11] or [23, Chapter 27].

Another standard result in de Rham theory is the existence of **Mayer-Vietoris sequences**. The setup is the following: M is a smooth manifold, and $U, V \subseteq M$ are open subsets such that $M = U \cup V$. The goal is to compute the cohomology of M in terms of the cohomologies of U, V, and $U \cap V$. The inclusions

$$i\colon U \cap V \hookrightarrow U, \quad j\colon U \cap V \hookrightarrow V, \quad k\colon U \hookrightarrow M, \quad \text{and} \quad \ell\colon V \hookrightarrow M$$

induce, via Proposition 53, linear mappings

$$\begin{gathered} i^*\colon H^p_{\mathrm{dR}}(U) \to H^p_{\mathrm{dR}}(U \cap V), \quad j^*\colon H^p_{\mathrm{dR}}(V) \to H^p_{\mathrm{dR}}(U \cap V), \\ k^*\colon H^p_{\mathrm{dR}}(M) \to H^p_{\mathrm{dR}}(U), \quad \text{and} \quad \ell^*\colon H^p_{\mathrm{dR}}(M) \to H^p_{\mathrm{dR}}(V), \end{gathered}$$

for each $p \geq 0$. We may put them together as

$$H^p_{\mathrm{dR}}(M) \xrightarrow{\;k^* \oplus \ell^*\;} H^p_{\mathrm{dR}}(U) \oplus H^p_{\mathrm{dR}}(V) \xrightarrow{\;i^* - j^*\;} H^p_{\mathrm{dR}}(U \cap V). \tag{5.31}$$

There is one copy of (5.31) for each value of $p \geq 0$, but it turns out that they can be connected to each other:

Theorem 27 (Mayer-Vietoris)

Let M be a smooth manifold, and $U, V \subseteq M$ be open subsets such that $M = U \cup V$. There is a sequence of linear mappings

$$\cdots \to H^p_{\mathrm{dR}}(M) \to H^p_{\mathrm{dR}}(U) \oplus H^p_{\mathrm{dR}}(V) \to H^p_{\mathrm{dR}}(U \cap V) \to H^{p+1}_{\mathrm{dR}}(M) \to \cdots \tag{5.32}$$

such that the image of any mapping equals the kernel of the next one.

See [15, Theorem 17.20] or [23, Section 26.1] for a proof. The crucial point is the construction of the mappings $H^p_{\mathrm{dR}}(U \cap V) \to H^{p+1}_{\mathrm{dR}}(M)$. A sequence of vector spaces

[14] Two continuous mappings $f, g\colon X \to Y$ are called **homotopic** if there is a continuous mapping $H\colon X \times [0,1] \to Y$ such that $H(x,0) = f(x)$ and $H(x,1) = g(x)$ for every $x \in X$. A **homotopy equivalence** is a continuous mapping $F\colon X \to Y$ for which there is $G\colon Y \to X$ such that $F \circ G$ is homotopic to Id_Y, and $G \circ F$ is homotopic to Id_X. In other words, F is invertible "up to homotopy".

and linear mappings like (5.32), with the property that the image of a mapping equals the kernel of the next, is called a **long exact sequence**. In particular, a piece of a long exact sequence like

$$\cdots \to 0 \to W \to Z \to 0 \to \cdots,$$

where 0 stands for the trivial vector space, implies that the arrow from W to Z is an isomorphism. Indeed, if $f\colon W \to Z$ is the arrow in question, the piece $0 \to W \to Z$ says that $\ker f$ equals the image of the zero mapping, making f injective; at the same time $W \to Z \to 0$ says that $\operatorname{Im} f$ equals the kernel of the zero mapping, which is all of Z, making f surjective.

Many of the ideas present in the proof of Theorem 27 belong to a branch of mathematics called **homological algebra**. We will not pursue this here, and instead just use Theorem 27 to compute cohomologies of a few more concrete examples.

Example 119 (The cohomology of the sphere)

We claim that

$$H^0_{\rm dR}(\mathbb{S}^n) \cong H^n_{\rm dR}(\mathbb{S}^n) \cong \mathbb{R}, \quad \text{and} \quad H^k_{\rm dR}(\mathbb{S}^n) = \{0\} \text{ for } 0 < k < n.$$

The standard argument is by induction on n, applying Mayer-Vietoris. The base case $n = 1$ was already established back in Example 117. As for the induction step, assume that the cohomology of $\mathbb{S}^n$ has been computed, and consider the open sets $U = \mathbb{S}^{n+1} \smallsetminus \{N\}$ and $V = \mathbb{S}^{n+1} \smallsetminus \{S\}$, where N and S are the north and south poles of $\mathbb{S}^{n+1}$. We have that $U \cup V = \mathbb{S}^{n+1}$. Clearly U and V are diffeomorphic to $\mathbb{R}^n$ (via stereographic projection, and hence have trivial cohomology by Example 116), while $U \cap V$ equals the "equator" of $\mathbb{S}^{n+1}$, which is $\mathbb{S}^n$. The piece

$$\cdots \longrightarrow H^k_{\rm dR}(U) \oplus H^k_{\rm dR}(V) \longrightarrow H^k_{\rm dR}(U \cap V) \longrightarrow$$
$$H^{k+1}_{\rm dR}(\mathbb{S}^{n+1}) \longrightarrow H^{k+1}_{\rm dR}(U) \oplus H^{k+1}_{\rm dR}(V) \longrightarrow \cdots$$

of the Mayer-Vietoris sequence now reads as

$$\cdots \to 0 \to H^k_{\rm dR}(\mathbb{S}^n) \to H^{k+1}_{\rm dR}(\mathbb{S}^{n+1}) \to 0 \to \cdots,$$

and hence $H^{k+1}_{\rm dR}(\mathbb{S}^n) \cong H^{k+1}_{\rm dR}(\mathbb{S}^{n+1})$ for every $n \geq 1$ and $k = 1, \ldots, n$. The desired conclusion now follows.

There is one more very geometric example we can study using Theorem 27. Recall that the **genus** of a closed surface is informally defined as the "number of holes" it has. Here (and always) "closed" means compact and without boundary. For example, a closed surface of genus zero is a sphere, of genus one is a torus, and so on.

Example 120 (The cohomology of genus-g surfaces)

Consider a closed orientable surface Σ_g of genus $g \geq 0$, as in Figure 71.

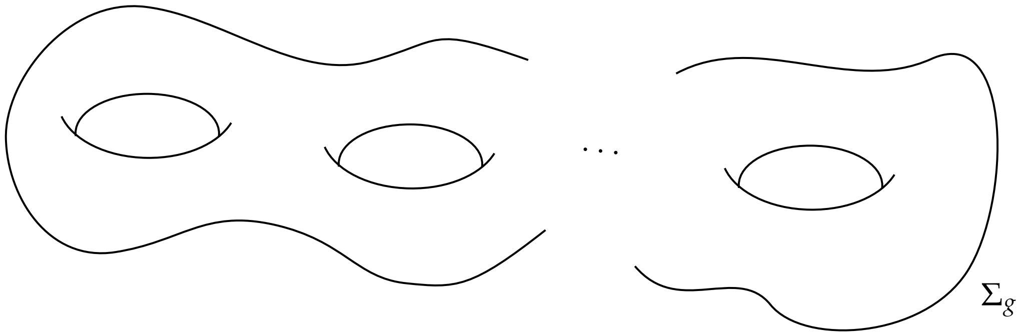


Figure 71: A closed surface of genus $g \geq 0$.

We claim that

$$H^0_{\mathrm{dR}}(\Sigma_g) = H^2_{\mathrm{dR}}(\Sigma_g) \cong \mathbb{R}, \quad \text{and} \quad H^1_{\mathrm{dR}}(\Sigma_g) \cong \mathbb{R}^{2g}.$$

Once again the argument is by induction, but this time on g. As $\Sigma_0 \cong \mathbb{S}^2$, the previous example takes care of the base case, and our goal reduces to showing that $H^1_{\mathrm{dR}}(\Sigma_{g+1}) \cong H^1_{\mathrm{dR}}(\Sigma_g) \oplus \mathbb{R}^2$. Proceeding, we take only one fact for granted: if M is any compact and oriented n-dimensional manifold, and $p \in M$, we have that $H^n_{\mathrm{dR}}(M \smallsetminus \{p\}) = \{0\}$. As $\Sigma_g = (\Sigma_g \smallsetminus \{p\}) \cup D$, where $p \in \Sigma_g$ is any point and $D \cong \mathbb{R}^2$ is a disk in Σ_g around p, so that $(\Sigma_g \smallsetminus \{p\}) \cap D$ is diffeomorphic to $\mathbb{S}^1 \times \mathbb{R}$, we obtain the Mayer-Vietoris sequence

$$0 \to \mathbb{R} \to \mathbb{R}^2 \to \mathbb{R} \to H^1_{\mathrm{dR}}(\Sigma_g) \to H^1_{\mathrm{dR}}(\Sigma_g \smallsetminus \{p\}) \oplus \{0\} \to \mathbb{R} \to \mathbb{R} \to 0.$$

Exactness of the long exact sequence implies that both arrows $\mathbb{R} \to H^1_{\mathrm{dR}}(\Sigma_g)$ and $H^1_{\mathrm{dR}}(\Sigma_g \smallsetminus \{p\}) \to \mathbb{R}$ are in fact the zero mapping, so that

$$0 \to H^1_{\mathrm{dR}}(\Sigma_g) \to H^1_{\mathrm{dR}}(\Sigma_g \smallsetminus \{p\}) \to 0$$

gives us that $H^1_{\mathrm{dR}}(\Sigma_g \smallsetminus \{p\}) \cong H^1_{\mathrm{dR}}(\Sigma_g)$.

With this in place, we set up a second Mayer-Vietoris sequence, noting that we may write $\Sigma_{g+1} = U \cup V$ for open subsets U and V which are diffeomorphic to $\Sigma_g \smallsetminus \{p\}$ and $\Sigma_1 \smallsetminus \{p\}$ (for some point $p \in \Sigma_{g+1}$), with $U \cap V$ diffeomorphic to

$\mathbb{S}^1 \times \mathbb{R}$. We obtain the long exact sequence

$$0 \longrightarrow \mathbb{R} \longrightarrow \mathbb{R}^2 \longrightarrow$$

$$\mathbb{R} \longrightarrow H^1_{\mathrm{dR}}(\Sigma_{g+1}) \longrightarrow H^1_{\mathrm{dR}}(\Sigma_g \smallsetminus \{p\}) \oplus H^1_{\mathrm{dR}}(\Sigma_1 \smallsetminus \{p\}) \longrightarrow$$

$$\mathbb{R} \longrightarrow \mathbb{R} \longrightarrow H^2_{\mathrm{dR}}(\Sigma_g \smallsetminus \{0\}) \oplus H^2_{\mathrm{dR}}(\Sigma_1 \smallsetminus \{p\}) \longrightarrow 0$$

which, similarly to what was done above, reduces to

$$0 \to H^1_{\mathrm{dR}}(\Sigma_{g+1}) \to H^1_{\mathrm{dR}}(\Sigma_g \smallsetminus \{p\}) \oplus \mathbb{R}^2 \to 0,$$

so that $H^1_{\mathrm{dR}}(\Sigma_{g+1}) \cong H^1_{\mathrm{dR}}(\Sigma_g) \oplus \mathbb{R}^2$, as required. (Here we have used the previously established relation $H^1_{\mathrm{dR}}(\Sigma_g \smallsetminus \{p\}) \cong H^1_{\mathrm{dR}}(\Sigma_g)$, as well as its particular case $H^1_{\mathrm{dR}}(\Sigma_1 \smallsetminus \{p\}) \cong H^1_{\mathrm{dR}}(\Sigma_1) = H^1_{\mathrm{dR}}(\mathbb{T}^2) = \mathbb{R}^2$, which can be justified ith a *third* application of Mayer-Vietoris, or using the Künneth formula below.)

The next natural question, as far as examples go, could be: how to compute the cohomology of a product of manifolds, in terms of the cohomologies of the factors? It is not as straightforward as it might seem, but there is also a result for that.

Theorem 28 (The Künneth formula)

Let M and N be smooth manifolds, and assume that one of them is compact. The cohomology of the product $M \times N$ is given by

$$H^k_{\mathrm{dR}}(M \times N) = \bigoplus_{i+j=k} H^i_{\mathrm{dR}}(M) \otimes H^j_{\mathrm{dR}}(N),$$

for every $k \geq 0$.

For a proof, see e.g. [2, pp. 47-50]. In practice, however, we don't really write out all of these tensor products explicitly. For any smooth n-dimensional manifold M, we may set $b_i(M) = \dim H^i_{\mathrm{dR}}(M)$ for each $i = 0, \ldots, n$; these are called the **Betti numbers** of M. We then use the Betti numbers as coefficients of a polynomial, $p_M(t) = \sum_{i=0}^n b_i(M) t^i$. It is called the **Poincaré polynomial of** M, and the Künneth formula reads simply as $p_{M \times N}(t) = p_M(t) p_N(t)$. Here are some explicit examples:

$$p_{\mathbb{R}^n}(t) = 1, \quad p_{\mathbb{S}^n}(t) = 1 + t^n, \quad p_{\mathbb{T}^n}(t) = (1+t)^n, \quad p_{\Sigma_g}(t) = 1 + 2gt + t^2.$$

The famous **Euler characteristic** of M (when compact) can be directly defined as the alternating sum $\chi(M) = \sum_{i=0}^n (-1)^i b_i(M)$, but in terms of the Poincaré polynomial, it becomes simply $\chi(M) = p_M(-1)$. So we have that $\chi(\mathbb{S}^n) = 1 + (-1)^n$, $\chi(\mathbb{T}^n) = 0$, and $\chi(\Sigma_g) = 2 - 2g$. We also have that $\chi(M \times N) = \chi(M)\chi(N)$. You can get a feeling for how the Künneth formula is used in the two exercises below.

Exercise 121

Show that if M is a smooth manifold with $n = \dim M$, then

$$H^k_{\rm dR}(M \times \mathbb{S}^1) = \begin{cases} H^0_{\rm dR}(M), & \text{if } k = 0, \\ H^k_{\rm dR}(M) \oplus H^{k-1}_{\rm dR}(M), & \text{if } k = 1, \ldots, n, \\ H^n_{\rm dR}(M), & \text{if } k = n+1. \end{cases}$$

And what about $H^k_{\rm dR}(M \times \mathbb{S}^m)$ for some fixed $m \geq 2$?

Exercise 122

Show that if M is a smooth manifold with $n = \dim M$, and Σ_g is a compact orientable surface of genus $g \geq 1$, as in Example 120, then

$$H^k_{\rm dR}(M \times \Sigma_g) = \begin{cases} H^0_{\rm dR}(M), & \text{if } k = 0, \\ H^1_{\rm dR}(M) \oplus H^0_{\rm dR}(M)^{\otimes 2g}, & \text{if } k = 1, \\ H^k_{\rm dR}(M) \oplus H^{k-1}_{\rm dR}(M)^{\otimes 2g} \oplus H^{k-2}_{\rm dR}(M), & \text{if } k = 2, \ldots, n, \\ H^n_{\rm dR}(M)^{\otimes 2g} \oplus H^{n-1}_{\rm dR}(M), & \text{if } k = n+1, \\ H^n_{\rm dR}(M), & \text{if } k = n+2. \end{cases}$$

One can simply write $H^k_{\rm dR}(M \times \Sigma_g) = H^k_{\rm dR}(M) \oplus H^{k-1}_{\rm dR}(M)^{\otimes 2g} \oplus H^{k-2}_{\rm dR}(M)$ without considering cases for k, with the understanding that all cohomology spaces of negative degree are trivial.

We conclude the course briefly visiting the situation dual to products: how to compute the cohomology of a quotient manifold? The next result below is also very well-known, but we present a complete proof (I could not find a reference for it stated in this manner and at this intermediate level of generality quickly enough).

Theorem 29 (Invariant cohomology versus cohomology of quotients)

Let M be a smooth manifold, and Γ be a finite group of diffeomorphisms of M acting freely on M, cf. Exercise 68 (p. 91). For each $k \geq 0$, let

$$\Omega^k(M)^\Gamma = \{\beta \in \Omega^k(M) : \gamma^*\beta = \beta \text{ for all } \gamma \in \Gamma\}$$

be the space of all **Γ-invariant** k-forms on M, and consider the spaces $Z^k(M)^\Gamma$ and $B^k(M)^\Gamma$ of all Γ-invariant closed and exact k-forms, respectively. Then we have that $B^k(M)^\Gamma \subseteq Z^k(M)^\Gamma$ and, defining the **k-th Γ-invariant cohomology space of** M as $H^k(M)^\Gamma = Z^k(M)^\Gamma / B^k(M)^\Gamma$, we have that

$$H^k_{\rm dR}(M/\Gamma) = H^k(M)^\Gamma,$$

for each $k \geq 0$. In addition, assigning to an invariant-cohomology class its corresponding de Rham cohomology class defines an injection $H^k(M)^\Gamma \hookrightarrow H^k_{\rm dR}(M)$.

Proof: The inclusions $B^k(M)^\Gamma \subseteq Z^k(M)^\Gamma$ follow, as usual, from $\mathrm{d}^2 = 0$ on Γ-invariant forms, together with the observation that $\mathrm{d}\beta \in \Omega^{k+1}(M)^\Gamma$ whenever $\beta \in \Omega^k(M)^\Gamma$. Indeed, by Proposition 46 we have $\gamma^*(\mathrm{d}\beta) = \mathrm{d}(\gamma^*\beta) = \mathrm{d}\beta$ for every $\gamma \in \Gamma$.

The first thing to establish here is that, if $\pi\colon M \to M/\Gamma$ denotes the quotient projection, then

$$\text{the image of } \pi^*\colon \Omega^k(M/\Gamma) \to \Omega^k(M) \text{ is precisely } \Omega^k(M)^\Gamma. \tag{5.33}$$

On one hand, whenever $\omega \in \Omega^k(M/\Gamma)$ we have that $\pi^*\omega \in \Omega^k(M)^\Gamma$, since for every $\gamma \in \Gamma$ we have—using item (iv) of Proposition 44—that $\gamma^*\pi^*\omega = (\pi \circ \gamma)^*\omega = \pi^*\omega$. For the reverse inclusion, if $\beta \in \Omega^k(M)^\Gamma$ is given, we define

$$\hat{\beta}_y(\hat{v}_1, \ldots, \hat{v}_k) = \beta_x(v_1, \ldots, v_k), \text{ for all } y \in M/\Gamma \text{ and } \hat{v}_1, \ldots, \hat{v}_k \in T_y(M/\Gamma), \tag{5.34}$$

where $x \in \pi^{-1}(y)$ and $v_1, \ldots, v_k \in T_xM$ such that $\mathrm{d}\pi_x(v_i) = \hat{v}_i$ for all $i = 1, \ldots, k$ are chosen at will. The definition of $\hat{\beta}$ is correct: if $x' \in \pi^{-1}(y)$ and $v'_1, \ldots, v'_k \in T_{x'}M$ are such that $\mathrm{d}\pi_{x'}(v'_i) = \hat{v}_i$ for all $i, \ldots, k$, there is $\gamma \in \Gamma$ such that $x' = \gamma(x)$, and thus $v'_i - \mathrm{d}\gamma_x(v_i) \in \ker \mathrm{d}\pi_{x'} = \{0\}$ leads to $v'_i = \mathrm{d}\gamma_x(v_i)$, for all $i = 1, \ldots, k$; then $\beta_x(v_1, \ldots, v_k) = \beta_{x'}(v'_1, \ldots, v'_k)$ follows from $\gamma^*\beta = \beta$. Then $\hat{\beta} \in \Omega^k(M/\Gamma)$ (it is smooth by the characteristic property of surjective submersions as π is a local diffeomorphism, cf. Exercise 80, p. 109) has $\pi^*\hat{\beta} = \beta$, by construction, proving (5.33).

As we know by Proposition 45 that π^* in (5.33) is injective, the desired conclusion (namely, that π^* induces an isomorphism at the cohomology level) will follow once we prove that

$$\text{(i) } \pi^*(Z^k(M/\Gamma)) = Z^k(M)^\Gamma \quad \text{and} \quad \text{(ii) } \pi^*(B^k(M/\Gamma)) = B^k(M)^\Gamma. \tag{5.35}$$

Consider (5.35-i): if $\omega \in Z^k(M/\Gamma)$, then $\pi^*\omega$ is already Γ-invariant due to (5.33), while $\mathrm{d}(\pi^*\omega) = \pi^*(\mathrm{d}\omega) = \pi^*(0) = 0$, so that $\pi^*\omega \in Z^k(M)^\Gamma$; if $\beta \in Z^k(M)^\Gamma$, (5.33) yields $\hat{\beta} \in \Omega^k(M/\Gamma)$ such that $\pi^*\hat{\beta} = \beta$, and taking exterior derivatives (again using Proposition 46) leads to $\pi^*(\mathrm{d}\hat{\beta}) = \mathrm{d}(\pi^*\hat{\beta}) = \mathrm{d}\beta = 0$, and hence $\mathrm{d}\hat{\beta} = 0$ (meaning that $\hat{\beta} \in Z^k(M/\Gamma)$) by injectivity of π^*.

Finally, we address (5.35-ii): if $\omega \in B^k(M/\Gamma)$, we may fix $\eta \in \Omega^{k-1}(M/\Gamma)$ such that $\omega = \mathrm{d}\eta$, and take pullbacks to obtain $\pi^*\omega = \mathrm{d}(\pi^*\eta)$; as $\pi^*\eta \in \Omega^{k-1}(M)^\Gamma$ by (5.33), we have that $\pi^*\omega \in B^k(M)^\Gamma$. The reverse inclusion is more subtle: if $\beta \in B^k(M)^\Gamma$ there is $\alpha \in \Omega^{k-1}(M)$ such that $\beta = \mathrm{d}\alpha$, *but α is not guaranteed to be Γ-invariant*. We instead consider

$$\text{the } \Gamma\text{-average } \bar{\alpha} \in \Omega^{k-1}(M)^\Gamma \text{ of } \alpha, \text{ defined as } \bar{\alpha} = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\gamma^*\alpha, \tag{5.36}$$

which is indeed Γ-invariant by construction: if $\mu \in \Gamma$ is arbitrary, then

$$\begin{aligned} \mu^*\bar{\alpha} = \mu^*\left(\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\gamma^*\alpha\right) &= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\mu^*\gamma^*\alpha \\ &= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}(\gamma\circ\mu)^*\alpha \overset{(\dagger)}{=} \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\gamma^*\alpha \\ &= \bar{\alpha}, \end{aligned} \tag{5.37}$$

where in (†) we use that $\Gamma \ni \gamma \mapsto \gamma \circ \mu \in \Gamma$ is a bijection. In addition, we have that

$$\begin{aligned} \mathrm{d}\bar{\alpha} &= \mathrm{d}\left(\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\gamma^*\alpha\right) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\mathrm{d}(\gamma^*\alpha) \\ &= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\gamma^*(\mathrm{d}\alpha) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\gamma^*\beta \\ &\overset{(\ddagger)}{=} \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\beta = \frac{1}{|\Gamma|}|\Gamma|\beta \\ &= \beta, \end{aligned} \tag{5.38}$$

where in (‡) we use Γ-invariance of β. Now, we may use (5.33) to fix $\hat{\alpha} \in \Omega^{k-1}(M/\Gamma)$ such that $\bar{\alpha} = \pi^*\hat{\alpha}$, and take exterior derivatives to obtain that $\beta = \pi^*(\mathrm{d}\hat{\alpha})$ with $\mathrm{d}\hat{\alpha} \in B^k(M/\Gamma)$, as required, establishing (5.35-ii).

That $H^k(M)^\Gamma \hookrightarrow H^k_{\mathrm{dR}}(M)$ is indeed an injection was already proved with (5.37) and (5.38): namely, if a Γ-invariant k-form has an antiderivative, then it has a Γ-invariant antiderivative. □

Remark. As discussed in the end of Section 4.2, one could reasonably wonder what Theorem 29 would look like if instead of Γ we had considered a positive-dimensional Lie group G acting freely and properly on M. However, there are two main changes. The first one occurs as immediately as in (5.33): the image of $\pi^*\colon \Omega^k(M/G) \to \Omega^k(M)$ is not $\Omega^k(M)^G$, but instead the subspace $\Omega^k(M)^G_{\mathrm{bas.}}$ consisting of the G-invariant k-forms which are also **basic**, that is, such that $\beta_x(v_1,\ldots,v_k) = 0$ whenever there is $i = 1,\ldots,k$ such that $v_i \in \ker \mathrm{d}\pi_x$ (when $G = \Gamma$ is zero-dimensional this kernel is trivial, so every form is basic). The need for this condition becomes apparent when proving that $\hat{\beta}$ in (5.34) is well-defined: we can still conclude that $v_i' - \mathrm{d}\gamma_x(v_i) \in \ker \mathrm{d}\pi_{x'}$, but the equality $\beta_x(v_1,\ldots,v_k) = \beta_{x'}(v_1',\ldots,v_k')$ now follows from rewriting the latter term as a certain telescopic sum and using basicness of β to eliminate all terms containing some such difference $v_i' - \mathrm{d}\gamma_x(v_i)$. The two conditions in (5.35) now obviously read $\pi^*(Z^k(M/G)) = Z^k(M)^G_{\mathrm{bas.}}$ and $\pi^*(B^k(M/G)) = B^k(M)^G_{\mathrm{bas.}}$, but while the proof of the former remains essentially the same as the one given above, the latter finally requires our second main change: the assumption that the Lie group G is compact. The reason for this is that if Γ is not finite, the sum used in (5.36) to define the average $\bar{\alpha}$ may not necessarily converge. However, every compact Lie group admits a unique bi-invariant probability measure μ_G, called its **Haar measure**, which is used as an replacement for the sum in (5.36); namely, we set $\bar{\alpha} = \int_G g^*\alpha\, \mathrm{d}\mu_G(g)$. We then have obvious analogues of (5.37) and (5.38), which allow us to conclude the proof. When $G = \Gamma$ is zero-dimensional and finite, we have that $\mu_\Gamma(A) = |A|/|\Gamma|$ is just the normalized counting measure.

A precise statement would be: *if M is a smooth manifold and G is a compact Lie group acting freely, properly, and effectively*[15] *on M, then $H^k_{\mathrm{dR}}(M/G) \cong H^k(M)^G_{\mathrm{bas.}}$, where $H^k(M)^G_{\mathrm{bas.}}$ is defined as the quotient $Z^k(M)^G_{\mathrm{bas.}}/B^k(M)^G_{\mathrm{bas.}}$.*

[15] So that we can uniquely identify each $g \in G$ with its corresponding diffeomorphism $g\colon M \to M$. As it turns out, this assumption is not really important here, but often made for psychological reasons.

Example 121 (The cohomology of $\mathbb{RP}^n$)

As an application of Theorem 29, we'll compute the cohomology of the real projective space $\mathbb{RP}^n$. We regard it as the quotient $\mathbb{S}^n/\mathbb{Z}_2$, where $\mathbb{Z}_2 = \{1, \tau\}$ for the antipodal mapping $\tau\colon \mathbb{S}^n \to \mathbb{S}^n$. As $H^k_{\rm dR}(\mathbb{RP}^n)$ can be seen as a subspace of $H^k_{\rm dR}(\mathbb{S}^n)$, which we have computed in Example 119, it follows that $H^k_{\rm dR}(\mathbb{RP}^n) = \{0\}$ for $k = 1, \ldots, n-1$. As $\mathbb{RP}^n$ is connected, we have that $H^0_{\rm dR}(\mathbb{RP}^n) = \mathbb{R}$. Finally, in view of Proposition 52 (and the companion result about the non-orientable case, mentioned after it), $H^n_{\rm dR}(\mathbb{RP}^n)$ can only be either $\{0\}$ or $\mathbb{R}$, depending on whether $\mathbb{RP}^n$ is orientable or not. This, by Proposition 49, depends on whether the antipodal mapping $\tau\colon \mathbb{S}^n \to \mathbb{S}^n$ is orientation-preserving or not. But you must have found in Exercise 116 that τ is orientation-preserving if and only if n is odd. In summary, we have that:

$$H^k_{\rm dR}(\mathbb{RP}^n) = \begin{cases} \mathbb{R}, & \text{if } k = 0, \\ \{0\}, & \text{if } k = 1, \ldots, n-1, \\ \mathbb{R}, & \text{if } k = n \text{ is odd}, \\ \{0\}, & \text{if } k = n \text{ is even}. \end{cases}$$

References

[1] Agaoka, Y.; *On a generalization of Cartan's lemma*. J. Algebra, 127(2):470-507, 1989.

[2] Bott, R.; Tu, L.; *Differential forms in algebraic topology.* Grad. Texts in Math., 82, Springer-Verlag, New York-Berlin, 1982. xiv+331 pp.

[3] Donaldson, S.; *Irrationality and the h-cobordism conjecture*, J. Diff. Geom. 26 (1987), 141–168.

[4] Donaldson, S.; *Polynomial invariants for smooth four manifolds*, Topology 29 (1990), 257–315.

[5] Freedman, M. H.; Taylor, L.; *A universal smoothing of four-space*, J. Diff. Geom. 24 (1986), 69–78.

[6] Freedman, M. W.; Gompf, R.; Morrison, S.; Walker, K.; *Man and machine thinking about the smooth 4-dimensional Poincaré conjecture*, Quantum Topology, Volume 1, Issue 2 (2010), pp. 171–208.

[7] Gompf, R.; *An infinite set of exotic* $\mathbb{R}^4$*'s*, J. Diff. Geom. 21 (1985), 283-300.

[8] Guillemin, V.; Pollack, A.; *Differential topology.* Reprint of the 1974 original. AMS Chelsea Publishing, Providence, RI, 2010. xviii+224 pp.

[9] Hamilton, M. J. D.; *Mathematical Gauge Theory. With applications to the standard model of particle physics*. Universitext, Springer, Cham, 2017. xviii+657 pp.

[10] Hirsch, M. W.; *Differential topology*, corrected reprint of the 1976 original, Grad. Texts in Math. 33, Springer, New York, 1994. x+222 pp.

[11] Isham, C.; *Modern Differential Geometry for Physicists*. Second edition. World Sci. Lecture Notes Phys., 61. World Scientific Publishing Co. Onc., River Edge, NJ, 1999. xiv+289 pp.

[12] Jänich, K.; *Vector analysis*. Translated from the second German (1993) edition by Leslie Kay. Undergrad. Texts Math. Springer-Verlag, New York, 2001. xiv+281 pp.

[13] Kervaire, M. A.; *A manifold which does not admit any differentiable structure*. Comment. Math. Helv. 34 (1): 257–270, 1960.

[14] Kervaire, M.; Milnor, J.; *Groups of homotopy spheres: I*, Ann. of Math. 77 (1963), 504–537.

[15] Lee, J. M.; *Introduction to Smooth Manifolds*. Second edition. Grad. Texts in Math. 218. Springer, New York, 2013. xvi+708 pp.

[16] Milnor, J. W. *Topology from the differentiable viewpoint.* Revised reprint of the 1965 original. Princeton Landmarks Math. Princeton University Press, Princeton, NJ, 1997. xii+64 pp.

[17] Milnor, J.; *On manifolds homeomorphic to the 7-sphere*. Ann. of Math. (2) 64 (1956), 399–405.

[18] Moise, E. E. *Affine structures in 3-manifolds, V: The triangulation theorem and Hauptvermutung*, Ann. of Math. (2) 56 (1952), 96–114, 1952.

[19] Munkres, J. R.; *Analysis on manifolds.* Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1991. xiv+366 pp.

[20] Munkres, J. R.; *Topology*. Second edition. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. xvi+537 pp.

[21] Radó, T.; *Über den Begriff der Riemannschen Fläche* , Acta Litt. Scient. Univ. Szegd 2, pp. 101-121, 1925.

[22] Stallings, J.; *The piecewise-linear structure of Euclidean space.* Proc. Cambridge Philos. Soc. 58 (1962), 481–488.

[23] Tu, L.; *Introduction to Manifolds*. Second edition. Universitext, Springer, New York, 2011. xviii+411 pp.

[24] Spivak, M.; *Calculus on Manifolds. A modern approach to classical theorems of advanced calculus.* W. A. Benjamin, Inc., New York-Amsterdam, 1965. xii+144 pp.

[25] Whitney, H.; *Differentiable manifolds.* Ann. of Math. (2) 37 (1936), no. 3, 645–680.

Index