Basic Geometry and Topology

Stephan Stolz

August 21, 2023

Contents

1 Pointset Topology

1.1 Open subsets of $\mathbb{R}^n$

Definition 1.1. For $x \in \mathbb{R}^n$ and $r > 0$, let

$$B_r(x) := \{y \in \mathbb{R}^n \mid \operatorname{dist}(x, y) < r\}$$

be the open ball $B_r(x)$ of radius r around x. Here

$$\operatorname{dist}(x, y) := ||x - y|| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

is the distance between the points x and y.

A subset $U \subset \mathbb{R}^n$ is *open* if for each point there is some $r > 0$ such that $B_r(x)$ is contained in U. Equivalently, U is open if and only if U is a union of open balls.

The point of this definition is that it makes it possible to give a very compact definition of continuity of maps $f\colon \mathbb{R}^m \to \mathbb{R}^n$ which is equivalent to the usual ϵ-δ definition.

Definition 1.2. A map $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is *continuous* if for every open subset $U \subset \mathbb{R}^n$ the preimage $f^{-1}(U)$ is open in $\mathbb{R}^m$. More generally, if $V \subset \mathbb{R}^m$, $W \subset \mathbb{R}^n$ are open subsets a map $f\colon V \to W$ is *continuous* if for every open subset $U \subset \mathbb{R}^n$ the preimage $f^{-1}(U)$ is open in $\mathbb{R}^m$.

Examples of continuous maps.

1. From calculus we know that the following maps $f\colon \mathbb{R} \supset V \to \mathbb{R}$ are continuous: polynomials, exponential functions, rational functions, trigonometric functions. Here $V \subset \mathbb{R}$ is the natural domain of these functions.
2. The maps $\mathbb{R}^2 \to \mathbb{R}$ given by $(x_1, x_2) \mapsto x_1 + x_2$ or $(x_1, x_2) \mapsto x_1 x_2$.
3. The *coordinate functions* $x^k\colon \mathbb{R}^m \to \mathbb{R}$ given by $(x_1, \dots, x_m) \mapsto x_k$, also known as *projection maps*.

Warning. The open set characterization of continuity is great for more abstract statements, like showing that the composition of continuous maps is continuous. However, checking that a given map f is continuous by verifying that $f^{-1}(U)$ is open for an open subset U of the codomain of f is usually cumbersome. A much better strategy is to recognize a given map as "built from simpler maps" that we already know to be continuous. The following three lemmas illustrate what we mean by "built from".

Lemma 1.3. *The composition of continuous maps is continuous.*

We leave the simple proof to the reader.

Lemma 1.4. *A map $f\colon V \to \mathbb{R}^n$ is continuous if and only if all its component maps $f_k\colon V \to \mathbb{R}$, $f_k := x^k \circ f$ are continuous (note that $f(x) = (f_1(x), \dots, f_n(x))$, which explains the terminology "component maps").*

The proof of this statement will follow from the much more general *continuity criterion for maps to a product*, which we will prove after introducing the product topology (see Lemma 1.19).

Lemma 1.5. *Let $f_1, f_2\colon \mathbb{R} \supset V \to \mathbb{R}$ be continuous maps. Then also $f_1 + f_2$ and $f_1 \cdot f_2$ are continuous.*

Proof. Let $f\colon V \to \mathbb{R}^2$ be the map with components maps f_1, f_2; i.e., $f(x) = (f_1(x), f_2(x))$. The map f is continuous since it component maps are continuous. The map $f_1 + f_2\colon V \to \mathbb{R}$ can be factored as

$$V \xrightarrow{f} \mathbb{R}^2 \xrightarrow{+} \mathbb{R}$$

and hence is continuous as the composition of continuous maps. Replacing the map $\mathbb{R}^2 \to \mathbb{R}$, $(x_1, x_2) \mapsto x_1 + x_2$ by the map $(x_1, x_2) \mapsto x_1 x_2$ similarly shows that $f_1 \cdot f_2$ is continuous. $\square$

Example 1.6. (More Examples of continuous maps.)

1. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial map, i.e.,

$$f(x) = \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} x_1^{i_1} \cdots x_n^{i_n} \qquad \text{for } x = (x_1,\dots,x_n) \in \mathbb{R}^n \text{ and coefficients } a_{i_1,\dots,i_n} \in \mathbb{R}$$

   We observe that f is a sum of functions, and each summand is a product of projection maps $x \mapsto x_k$ and the constant map $x \mapsto a_{i_1,\dots,i_n}$. Hence the continuity of the project maps and constant maps imply by Lemma 1.5 the continuity of each summand, which in turn implies the continuity of f.

2. Let $M_{n\times n}(\mathbb{R}) = \mathbb{R}^{n^2}$ be the vector space of $n \times n$ matrices. Then the map

$$M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R}) \longrightarrow M_{n\times n}(\mathbb{R}) \qquad (A,B) \mapsto AB$$

   given by matrix multiplication is continuous. To see this, it suffices by Lemma 1.4 to check that each component map is continuous. This is the case, since each matrix entry of AB is a polynomial and hence a continuous function of the matrix entries of A and B.

1.2 Topological spaces

The characterization 1.2 of continuous maps $f\colon \mathbb{R}^m \to \mathbb{R}^n$ in terms of open subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ suggests that we can define what we mean by a *continuous map* $f\colon X \to Y$ between sets X, Y, once we pick collections $\mathcal{T}_X$, $\mathcal{T}_Y$ of subsets of X resp. Y that we consider the "open subsets" of these sets. The next result summarizes the basic properties of open subsets of $\mathbb{R}^n$, which then motivates the restrictions that we wish to put on such collections $\mathcal{T}$.

Lemma 1.7. *Open subsets of $\mathbb{R}^n$ have the following properties.*

(i) $\mathbb{R}^n$ *and* $\emptyset$ *are open.*

(ii) *Any union of open sets is open.*

(iii) *The intersection of any finite number of open sets is open.*

Definition 1.8. A *topological space* is a set X together with a collection $\mathcal{T}$ of subsets of X, called *open sets* which are required to satisfy conditions (i), (ii) and (iii) of the lemma above. The collection $\mathcal{T}$ is called a *topology* on X. The sets in $\mathcal{T}$ are called the *open sets*, and their complements in X are called *closed sets*. A subset of X may be neither closed nor open, either closed or open, or both.

A map $f\colon X \to Y$ between topological spaces X, Y is *continuous* if the inverse image $f^{-1}(V)$ of every open subset $V \subset Y$ is an open subset of X.

It is easy to see that the composition of continuous maps is again continuous.

Example 1.9. (Examples of topological spaces.)

1. Let $\mathcal{T}$ be the collection of open subsets of $\mathbb{R}^n$ in the sense of Definition 1.1. Then $\mathcal{T}$ is a topology on $\mathbb{R}^n$, the *standard topology on* $\mathbb{R}^n$ or *metric topology on* $\mathbb{R}^n$ (since this topology is determined by the metric $\text{dist}(x, y) = ||x - y||$ on $\mathbb{R}^n$).

2. Let X be a set. Then $\mathcal{T} = \{\text{all subsets of } X\}$ is a topology, the *discrete topology.* We note that *any* map $f: X \to Y$ to a topological space Y is continuous. We will see later that the only continuous maps $\mathbb{R}^n \to X$ are the constant maps.

3. Let X be a set. Then $\mathcal{T} = \{\emptyset, X\}$ is a topology, the *indiscrete topology.*

Sometimes it is convenient to define a topology $\mathcal{T}$ on a set X by first describing a smaller collection $\mathcal{B}$ of subsets of X, and then defining $\mathcal{T}$ to be those subsets of X that can be written as *unions* of subsets belonging to $\mathcal{B}$. We've done this already when the topology on $\mathbb{R}^n$: Let $\mathcal{B}$ be the collection of all open balls $B_r(x) \subset \mathbb{R}^n$; we recall that $B_r(x) = \{y \in X \mid \text{dist}(x, y) < r\}$. The standard topology on $\mathbb{R}^n$ consists of those subsets U which are unions of subsets belonging to $\mathcal{B}$.

Lemma 1.10. *Let $\mathcal{B}$ be a collection of subsets of a set X satisfying the following conditions*

(i) *Every point $x \in X$ belongs to some subset $B \in \mathcal{B}$.*

(ii) *If $B_1, B_2 \in \mathcal{B}$, then for every $x \in B_1 \cap B_2$ there is some $B \in \mathcal{B}$ with $x \in B$ and $B \subset B_1 \cap B_2$.*

Then $\mathcal{T} := \{\text{unions of subsets belonging to } \mathcal{B}\}$ is a topology on X.

Definition 1.11. If the above conditions are satisfied, we call the collection $\mathcal{B}$ is called a *basis for the topology* $\mathcal{T}$. Conversely, X is a topological space, and $\mathcal{B}$ is a collection of open subsets of X such that each open subset is a union of subsets belonging to B, we say that $\mathcal{B}$ *generates the topology.*

We note that if $\mathcal{B}$ generates the topology of a topological space X, then $\mathcal{B}$ automatically satisfies conditions (i) & (ii) above, i.e., $\mathcal{B}$ is in fact a basis for the given topology.

1.2.1 Subspace topology

Definition 1.12. Let X be a topological space, and $A \subset X$ a subset. Then

$$\mathcal{T} = \{A \cap U \mid U \underset{open}{\subset} X\}$$

is a topology on A called the *subspace topology.*

We note that the inclusion map $i\colon A \hookrightarrow X$ is continuous for the subspace topology on A, since for every open subset $U \subset X$, $i^{-1}(U) = U \cap A$ is an open subset of A by construction of the subspace topology.

Example 1.13. (Examples of subspaces of $\mathbb{R}^n$) Here are examples of subspaces of $\mathbb{R}^n$ (i.e., subsets of $\mathbb{R}^n$ equipped with the subspace topology) we will be talking about during the semester:

1. The *n-disk* $D^n := \{x \in \mathbb{R}^n \mid |x| \le 1\} \subset \mathbb{R}^n$, and $D^n_r := \{x \in \mathbb{R}^n \mid |x| \le r\}$, the *$n$-disk of radius* $r > 0$.

2. The *n-sphere* $S^n := \{x \in \mathbb{R}^{n+1} \mid |x| = 1\} \subset \mathbb{R}^{n+1}$.

3. The *torus* $T = \{v \in \mathbb{R}^3 \mid \mathrm{dist}(v, K) = r\}$ for $0 < r < 1$. Here
$$K = \{(x, y, 0) \mid x^2 + y^2 = 1\} \subset \mathbb{R}^3$$
is the unit circle in the xy-plane, and $\mathrm{dist}(v, K) = \inf_{w \in K} \mathrm{dist}(v, w)$ is the distance between v and K.

4. The *general linear group*
$$\begin{aligned} GL_n(\mathbb{R}) &= \{\text{vector space isomorphisms } f\colon \mathbb{R}^n \to \mathbb{R}^n\} \\ &\longleftrightarrow \{(v_1, \dots, v_n) \mid v_i \in \mathbb{R}^n,\ \det(v_1, \dots, v_n) \neq 0\} \\ &= \{\text{invertible } n \times n\text{-matrices}\} \subset M_{n\times n}(\mathbb{R}) = \mathbb{R}^{n^2} \end{aligned}$$
Here we think of $(v_1, \dots, v_n)$ as an $n \times n$-matrix with column vectors v_i, and the bijection is the usual one in linear algebra that sends a linear map $f\colon \mathbb{R}^n \to \mathbb{R}^n$ to the matrix $(f(e_1), \dots, f(e_n))$ whose column vectors are the images of the standard basis elements $e_i \in \mathbb{R}^n$.

5. The *special linear group*
$$SL_n(\mathbb{R}) = \{(v_1, \dots, v_n) \mid v_i \in \mathbb{R}^n,\ \det(v_1, \dots, v_n) = 1\} \subset M_{n\times n}(\mathbb{R}) = \mathbb{R}^{n^2}$$

6. The *orthogonal group*
$$\begin{aligned} O(n) &= \{\text{linear isometries } f\colon \mathbb{R}^n \to \mathbb{R}^n\} \\ &= \{(v_1, \dots, v_n) \mid v_i \in \mathbb{R}^n,\ v_i\text{'s are orthonormal}\} \subset M_{n\times n}(\mathbb{R}) = \mathbb{R}^{n^2} \end{aligned}$$
We recall that a collection of vectors $v_i \in \mathbb{R}^n$ is *orthonormal* if $|v_i| = 1$ for all i, and v_i is perpendicular to v_j for $i \neq j$.

7. The *special orthogonal group*

$$SO(n) = \{(v_1, \dots, v_n) \in O(n) \mid \det(v_1, \dots, v_n) = 1\} \subset M_{n\times n}(\mathbb{R}) = \mathbb{R}^{n^2}$$

8. The *Stiefel manifold*

$$V_k(\mathbb{R}^n) = \{(v_1, \dots, v_k) \mid v_i \in \mathbb{R}^n,\ v_i\text{'s are orthonormal}\} \subset M_{n\times k}(\mathbb{R}) = \mathbb{R}^{nk}$$

Lemma 1.14. (Continuity criterion for maps to a subspace.) *Let X, Y be topological spaces and let B be a subset of Y equipped with the subspace topology. Then a map $f\colon X \to B$ is continuous if and only if the composition $X \xrightarrow{f} B \xrightarrow{i} Y$ is continuous.*

Proof. Homework □

Example 1.15. (Examples of continuous maps involving subspaces.)

1. The map $GL_n(\mathbb{R}) \to GL_n(\mathbb{R})$, $A \mapsto A^{-1}$ is continuous. Homework problem. Hint: by the above lemma, it suffices to prove continuity of the composition $GL_n(\mathbb{R}) \to GL_n(\mathbb{R}) \hookrightarrow M_{n\times n}(\mathbb{R})$, which in turn by Lemma 1.4 amounts to checking continuity of each matrix component of A^{-1} as a function of the matrix components of A.

2. Let G be one of the groups $SL_n(\mathbb{R})$, $O(n)$, $SO(n)$, equipped with the subspace topology as subsets of $M_{n\times n}(\mathbb{R})$. Then the map $G \to G$, $A \mapsto A^{-1}$ is continuous. To see that this map is continuous, we note it is the restriction of the continuous map $A \mapsto A^{-1}$ on $GL_n(\mathbb{R})$ to the subspace $G \subset GL_n(\mathbb{R})$ and use the following handy fact.

Lemma 1.16. *Let $f\colon X \to Y$ be a continuous map. If $A \subset X$, $B \subset Y$ are subspaces with $f(A) \subset B$, then the* restriction $f_{|A}\colon A \to B$ *is continuous (with respect to the subspace topology on X and Y.*

Proof. Consider the commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{f_{|A}} & B \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle j} \\ X & \xrightarrow{f} & Y \end{array}$$

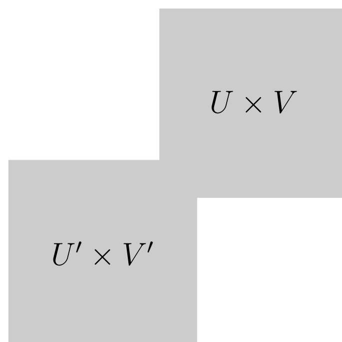
where i, j are the obvious inclusion maps. These inclusion maps are continuous w.r.t. the subspace topology on A, B by Lemma 1.14. The continuity of f and i implies the continuity of $f \circ i = j \circ f_{|A}$ which again by Lemma 1.14 implies the continuity of $f_{|A}$. □

1.2.2 Product topology

Definition 1.17. The *product topology* on the Cartesian product $X \times Y = \{(x, y) \mid x \in X,\ y \in Y\}$ of topological spaces X, Y is the topology generated by the subsets

$$\mathcal{B} = \{U \times V \mid U \underset{open}{\subset} X,\ V \underset{open}{\subset} Y\}$$

The collection $\mathcal{B}$ obviously satisfies property (1) of a basis (see Definition 1.11); property (2) holds since $(U \times V) \cap (U' \times V') = (U \cap U') \times (V \cap V')$. We note that the collection $\mathcal{B}$ is *not* a topology since the union of $U \times V$ and $U' \times V'$ is typically not a Cartesian product. For example, if $X = Y = \mathbb{R}$ and U, U', V, V' are open intervals the products $U \times V$ and $U' \times V'$ are (open) rectangles whose union might look like the shaded region in the figure below.



We note that the projection maps $p_1 \colon X \times Y \to X$ and $p_2 \colon X \times Y \to Y$ are continuous: if U is an open subset of X, then $p_1^{-1}(U) = U \times Y$ is a product of open subsets, i.e., it belongs to the collection B. In particular, it is an open subset of $X \times Y$ equipped with the product topology, and hence p_1 is continuous. The argument for p_2 is analogous.

There is obviously a plethora of examples of product spaces, e.g., the product of any two of the eight spaces of Example 1.13. Sometimes, the product topology on a product agrees with a topology described in a different way, for example:

Lemma 1.18. *The product topology on $\mathbb{R}^m \times \mathbb{R}^n$ (with each factor equipped with the metric topology) agrees with the metric topology on $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$.*

Proof: homework.

Other product spaces might be homeomorphic to topological spaces constructed completely differently. For example, we will see that the product $S^1 \times S^1$ is homeomorphic to the torus T of Example 1.13(3). To work with product spaces, it is very useful to have the following recognition principal for continuity of map to a product.

Lemma 1.19. (Continuity criterion for maps to a product.) *Let X, Y_1, Y_2 be topological spaces. Then a map $f \colon X \to Y_1 \times Y_2$ is continuous if and only if the compositions*

$$X \xrightarrow{f} Y_1 \times Y_2 \xrightarrow{p_i} Y_i$$

are continuous for $i = 1, 2$, where $p_i \colon Y_1 \times Y_2 \to Y_i$ is the projection map.

We note that the composition $p_i \circ f$ is the i-th component map of f. So according to the above lemma a map to a product is continuous if and only if all its component maps are continuous. This is a far reaching generalization of Lemma 1.4 which was about maps with target space $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$.

For the proof of Lemma 1.19, as well as in many other situations, it will be helpful to use the following simple result, the reader is charged with proving.

Lemma 1.20. *Let $f \colon X \to Y$ be a map be topological spaces. Suppose the topology on the codomain Y is generated by a basis $\mathcal{B}$. Then f is continuous if and only if $f^{-1}(U)$ is open in X for every $U \in \mathcal{B}$.*

Proof of Lemma 1.19. If $f \colon X \to Y_1 \times Y_2$ is continuous, then the component maps $f_i := p_i \circ f$ are continuous, since they are compositions of the continuous maps p_i and f. Conversely, assume that the component maps f_1, f_2 are continuous. To show that f is continuous it suffices by the previous lemma to show that $f^{-1}(U)$ is open where U belongs to the basis $\mathcal{B}$ that generated the product topology. In other words, U is a product $U = U_1 \times U_2$ of open subsets $U_1 \subset Y_1$, $U_2 \subset Y_2$. Then

$$f^{-1}(U) = f^{-1}(U_1 \times U_2) = f_1^{-1}(U_1) \cap f_2^{-1}(U_2) \subset X$$

is an open subset of X, since $f_i^{-1}(U_i)$ is open in X by the assumed continuity of f_i. $\square$

The following result is consequence of the Continuity criterion for maps to a product; its proof is a good illustration of how the criterion is used.

Lemma 1.21. *Let G be one of the groups $GL_n(\mathbb{R})$, $SL_n(\mathbb{R})$, $O(n)$, $SO(n)$, equipped with the subspace topology as subsets of $M_{n\times n}(\mathbb{R})$. Then G is a* topological group*, i.e., G is a topological space and a group, and the topology and the group structure are compatible in the sense that*

- *The multiplication map $G \times G \xrightarrow{\mu} G$ is continuous, and*
- *the map $G \to G$, $g \mapsto g^{-1}$ is continuous.*

Proof. We discussed continuity of the inverse map in Example 1.15. To prove continuity of the multiplication map μ, we consider the commutative diagram

$$\begin{array}{ccc} G \times G & \xrightarrow{\quad\mu\quad} & G \\ \downarrow{\scriptstyle i\times i} & & \downarrow{\scriptstyle i} \\ M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R}) & \xrightarrow{m} & M_{n\times n}(\mathbb{R}) \end{array}$$

where i is the inclusion map, and m is matrix multiplication which is continuous by Example 1.6. It might be tempting to argue that μ is the restriction of the continuous map m, and hence it is continuous by Lemma 1.16. However, that assumes that $G \times G$ is equipped with its *subspace topology* as a subset of $M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R})$, rather than as equipped with the *product topology.* Proving that these topologies in fact agree is one way to finish the proof.

Alternatively, using Lemma 1.19, we argue that the map $i \times i \colon G \times G \to M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R})$ is continuous since its component maps are: the first component map is the composition of the continuous maps

$$G \times G \xrightarrow{p_1} G \xrightarrow{i} M_{n\times n}(\mathbb{R})$$

and hence continuous; similarly for the second component map. Hence $m\circ(i\times i)$ is continuous, which equals $i \circ \mu$ by the commutativity of the diagram. It follows that μ is continuous by the criterion for continuity of a map to a subspace 1.14 □

1.2.3 Quotient topology.

Definition 1.22. Let X be a topological space and let $\sim$ be an equivalence relation on X. We denote by $X/\sim$ be the set of equivalence classes and by

$$p\colon X \to X/\sim \qquad x \mapsto [x]$$

be the projection map that sends a point $x \in X$ to its equivalence class $[x]$. The *quotient topology* on $X/\sim$ is given by the collection of subsets

$$\mathcal{T} = \{U \subset X/\sim \mid p^{-1}(U) \text{ is an open subset of } X\}.$$

The set $X/\sim$ equipped with the quotient topology is called the *quotient space.*

We note that the projection map $p\colon X \to X/\sim$ is continuous, since if U is an open subset of $X/\sim$ with respect to the quotient topology, then by definition of that topology $p^{-1}(U)$ is an open subset of X.

The quotient topology is often used to construct a topology on a set Y which is not a subset of some Euclidean space $\mathbb{R}^n$, or for which it is not clear how to construct a metric. If there is a surjective map

$$p\colon X \twoheadrightarrow Y$$

from a topological space X, then Y can be identified with the quotient space $X/\sim$, where the equivalence relation is given by $x \sim x'$ if and only if $p(x) = p(x')$. In particular, $Y = X/\sim$ can be equipped with the quotient topology. Here are important examples.

Example 1.23. (Examples of quotient spaces).

1. Let A be a subset of a topological space X. Define a equivalence relation $\sim$ on X by $x \sim y$ if $x = y$ or $x, y \in A$. We use the notation X/A for the quotient space $X/\sim$. A concrete example is provided by D^n/S^{n-1}, which is homeomorphic to the sphere S^n, as we will see later.

2. The *real projective space of dimension* n is the set

$$\mathbb{RP}^n := \{\text{1-dimensional subspaces of } \mathbb{R}^{n+1}\}.$$

The map

$$S^n \longrightarrow \mathbb{RP}^n \qquad \mathbb{R}^{n+1} \ni v \mapsto \text{subspace generated by } v$$

is surjective, leading to the identification

$$\mathbb{RP}^n = S^n/(v \sim \pm v),$$

and the quotient topology on $\mathbb{RP}^n$.

3. Similarly, working with complex vector spaces, we obtain a quotient topology on the the *complex projective space*

$$\mathbb{CP}^n := \{\text{1-dimensional subspaces of } \mathbb{C}^{n+1}\} = S^{2n+1}/(v \sim zv), \qquad z \in S^1$$

4. Generalizing, we can consider the *Grassmann manifold*

$$G_k(\mathbb{R}^{n+k}) := \{k\text{-dimensional subspaces of } \mathbb{R}^{n+k}\}.$$

There is a surjective map

$$V_k(\mathbb{R}^{n+k}) = \{(v_1, \dots, v_k) \mid v_i \in \mathbb{R}^{n+k},\ v_i\text{'s are orthonormal}\} \quad \twoheadrightarrow \quad G_k(\mathbb{R}^{n+k})$$

given by sending $(v_1, \dots, v_k) \in V_k(\mathbb{R}^{n+k})$ to the k-dimensional subspace of $\mathbb{R}^{n+k}$ spanned by the v_i's. Hence the subspace topology on the Stiefel manifold $V_k(\mathbb{R}^{n+k}) \subset \mathbb{R}^{(n+k)k}$ gives a quotient topology on the Grassmann manifold $G_k(\mathbb{R}^{n+k}) = V_k(\mathbb{R}^{n+k})/\sim$. The same construction works for the complex Grassmann manifold $G_k(\mathbb{C}^{n+k})$.

As the example 1.23(1) shows, a quotient space $Y = X/\sim$ might be homeomorphic to a topological space Z constructed in a different way. To establish the homeomorphism between Y and Z, we need to construct continuous maps

$$f\colon Y \longrightarrow Z \qquad g\colon Z \to Y$$

that are inverse to each other. The next lemma shows that it is easy to check continuity of the map f, the map *out of the quotient space.*

Lemma 1.24. (Continuity criterion for maps out of a quotient space). *Let X be a topological space, $p\colon X \to Y$ a surjective map, and let Y be equipped with the quotient topology. A map $f\colon Y \to Z$ to a topological space Z is continuous if and only if the composition*

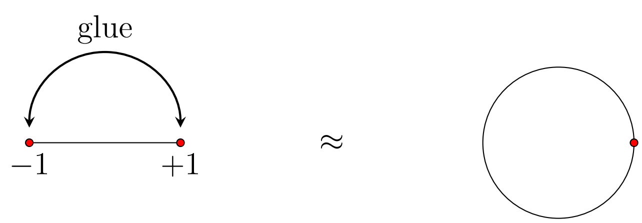
$$X \xrightarrow{p} Y \xrightarrow{f} Z$$

is continuous.

Proof: homework

As we will see in the next section, there are many situations where the continuity of the inverse map for a continuous bijection f is automatic. So in the examples below, and for the exercises in this section, we will defer checking the continuity of f^{-1} to that section.

Example 1.25. (1) We claim that the quotient space $[-1,+1]/\{\pm 1\}$ is homeomorphic to S^1 via the map $f\colon [-1,+1]/\{\pm 1\} \to S^1$ given by $[t] \mapsto e^{\pi i t}$. Geometrically speaking, the map f wraps the interval $[-1,+1]$ once around the circle. Here is a picture.



It is easy to check that the map f is a bijection. To see that f is continuous, consider the composition

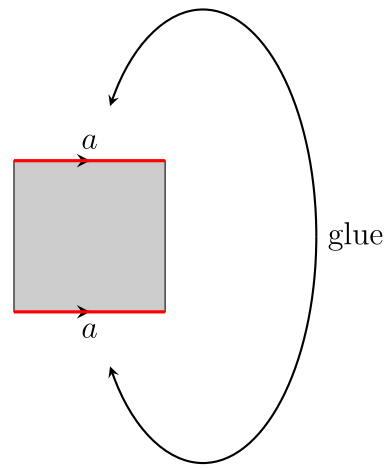
$$[-1,+1] \xrightarrow{p} [-1,+1]/\{\pm 1\} \xrightarrow{f} S^1 \xrightarrow{i} \mathbb{C} = \mathbb{R}^2,$$

where p is the projection map and i the inclusion map. This composition sends $t \in [-1,+1]$ to $e^{\pi i t} = (\cos \pi t, \sin \pi t) \in \mathbb{R}^2$. By Lemma 1.19 it is a continuous function, since its component functions $\sin \pi t$ and $\cos \pi t$ are continuous functions. By Lemma 1.24 the continuity of $i \circ f \circ p$ implies the continuity of $i \circ f$, which by Lemma 1.14 implies the continuity of f. As mentioned above, we'll postpone the proof of the continuity of the inverse map f^{-1} to the next section.

(2) More generally, D^n/S^{n-1} is homeomorphic to S^n. (proof: homework)

(3) Consider the quotient space of the square $[-1,+1] \times [-1,+1]$ given by identifying $(s,-1)$ with $(s,1)$ for all $s \in [-1,1]$. It can be visualized as a square whose top edge is

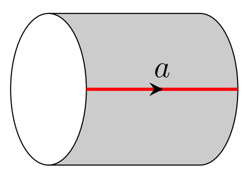
to be glued with its bottom edge. In the picture below we indicate that identification by labeling those two edges by the same letter.



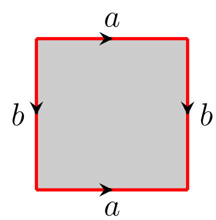
The quotient $([-1,+1]\times[-1,+1])/(s,-1)\sim(s,+1)$ is homeomorphic to the cylinder

$$C=\{(x,y,z)\in\mathbb{R}^3\mid x\in[-1,+1],y^2+z^2=1\}.$$

The proof is essentially the same as in (1). A homeomorphism from the quotient space to C is given by $f([s,t])=(s,\sin\pi t,\cos\pi t)$. The picture below shows the cylinder C with the image of the edge a indicated.



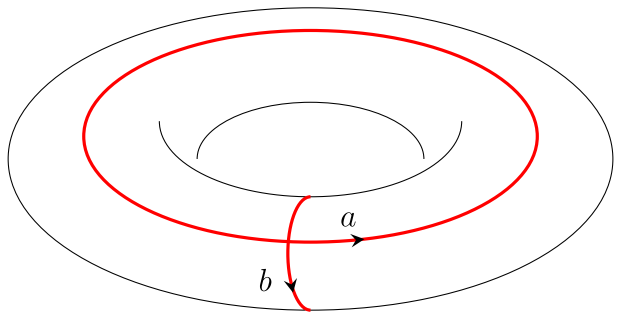
(4) Consider again the square, but this time using an equivalence relations that identifies more points than the one in the previous example. As before we identify $(s,-1)$ and $(s,1)$ for $s\in[-1,1]$, and in addition we identify $(-1,t)$ with $(1,t)$ for $t\in[-1,1]$. Here is the picture, where again corresponding points of edges labeled by the same letter are to be identified.



We claim that the quotient space is homeomorphic to the torus

$$T := \{x \in \mathbb{R}^3 \mid d(x, K) = d\},$$

where $K = \{(x_1, x_2, 0) \mid x_1^2 + x_2^2 = 1\}$ is the unit circle in the xy-plane and $0 < d < 1$ is a real number (see) via a homeomorphism that maps the edges of the square to the loops in T indicated in the following picture below.



Exercise: prove this by writing down an explicit map from the quotient space to T, and arguing that this map is a continuous bijection (as always in this section, we defer the proof of the continuity of the inverse to the next section).

(5) We claim that the quotient space $D^n/\sim$ with equivalence relation generated by $v \sim -v$ for $v \in S^{n-1} \subset D^n$ is homeomorphic to the real projective space $\mathbb{RP}^n$. More precisely, let $f\colon D^n \to S^n$ be the embedding of the n-disk as the upper hemisphere of S^n. Explicitly, $f(x)$ for $x = (x_1, \dots, x_n)$ is given by the formula

$$f(x_1, \dots, x_n) := (x_1, \dots, x_n, \sqrt{1 - (x_1^2 + \cdots + x_n^2)})$$

Lemma 1.26. *The map $\bar{f}\colon D^n/\sim \to \mathbb{RP}^n = S^n/\sim$ given by $[x] \mapsto [f(x)]$ is a continuous bijection.*

With more tools at our disposal in the next section we will argue that this map is in fact a homeomorphism.

Proof. To check that $\bar{f}$ is well-defined, we note that get identified in D^n are $x \sim -x$ for $x \in \partial D^n = S^{n-1}$. For such x, $f(x) = (x_1, \dots, x_n, 0)$ and $f(-x) = -(x_1, \dots, x_n, 0)$, showing that $\bar{f}$ is well-defined.

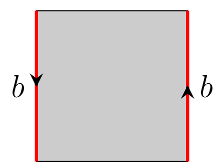
Next we argue that $\bar{f}$ is continuous. The map f is continuous since its components are continuous functions. By construction of $\bar{f}$ we have the commutative diagram

$$\begin{tikzcd} D^n \arrow[r, "f"] \arrow[d, "p_1"'] & S^n \arrow[d, "p_2"] \\ D^n/\sim \arrow[r, "\bar{f}"] & S^n/\sim \end{tikzcd}$$

where the vertical maps are the projection maps. Since f is continuous, so is the composition $p_2 \circ f = p_1 \circ \bar{f}$, and hence $\bar{f}$ (a map out of a quotient space is continuous if and only if its pre-composition with the projection map is).

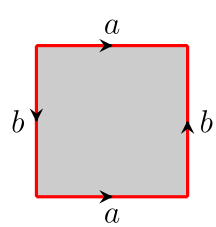
The map f provides a bijection between D^n and the upper hemisphere of S^n (including the equator); the inverse map is given by sending a point $(x_1, \dots, x_{n+1})$ in the upper hemisphere to $(x_1, \dots, x_n)$. Since every equivalence class in S^n can be represented by a point in the upper hemisphere, this implies that $\bar{f}$ is surjective. Since the only points in the upper hemisphere that are identified by the equivalence relation on S^n are antipodal points on the equator, this implies that $\bar{f}$ is injective. $\square$

(6) The quotient space $[-1,1] \times [-1,1]/ \sim$ with the equivalence relation generated by $(-1,t) \sim (1,-t)$ is represented graphically by the following picture.



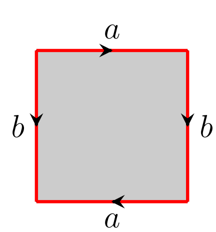
This topological space is called the *Möbius band*. It is homeomorphic to a subspace of $\mathbb{R}^3$ shown by the following picture

(7) The quotient space of the square by edge identifications given by the picture



is the *Klein bottle*. It is harder to visualize, since it is not homeomorphic to a subspace of $\mathbb{R}^3$ (which can be proved by the methods of algebraic topology).

(8) The quotient space of the square given by the picture



is homeomorphic to the real projective plane $\mathbb{RP}^2$. Exercise: prove this (hint: use the statement of example (5)). Like the Klein bottle, it is challenging to visualize the real projective plane, since it is not homeomorphic to a subspace of $\mathbb{R}^3$.

1.3 Properties of topological spaces

In the previous subsection we described a number of examples of topological spaces X, Y that we claimed to be homeomorphic. We typically constructed a bijection $f\colon X \to Y$ and argued that f is continuous. However, we did not finish the proof that f is a homeomorphism, since we defered the argument that the inverse map $f^{-1}\colon Y \to X$ is continuous. We note that not every continuous bijection is a homeomorphism.

For example, the map

$$f\colon [0,1) \longrightarrow S^1 \subset \mathbb{R}^2 = \mathbb{C} \qquad \text{given by} \qquad t \mapsto e^{2\pi i t} \tag{1.27}$$

is a bijection. It is the restriction of the map $\widetilde{f}\colon \mathbb{R} \to \mathbb{R}^2$ given by the same formula; $\widetilde{f}$ is continuous since its component functions $\cos 2\pi i t$ and $\sin 2\pi i t$ are continuos, and hence f is continuous (with the respect to the subspace topology on $[0,1) \subset \mathbb{R}$ and $S^1 \subset \mathbb{R}^2$). The inverse map $g\colon S^1 \to [0,1)$ is *not continuous*, since $[0,1/2) \subset [0,1)$ is open, but $g^{-1}([0,1/2)) = f([0,1/2))$ consists of the lower semicircle (the intersection of the lower open halfplane $\{(x,y) \in \mathbb{R}^2 \mid y < 0\}$ with S^1) and the point $(1,0)$) which we claim is not an open subset of S^1. To prove this, assume that $f([0,1/2))$ is in fact open in the subspace topology, i.e., $f([0,1/2)) = S^1 \cap U$ for some open subset $U \subset \mathbb{R}^2$. Since $(1,0) \in U$ and U is open, there is radius $r > 0$ such that the ball $B_r((1,0))$ is contained in U, and hence $S^1 \cap B_r((1,0)) \subset S^1 \cap U = f([0,1/2))$. This is the desired contradiction, since no point with positive y coordinate belongs to $f([0,1/2))$.

Fortunately, there are situations where the continuity of the inverse map is automatic as the following proposition shows.

Proposition 1.28. (Continuity criterion of the inverse of a continuous bijection). *Let $f\colon X \to Y$ be a continuous bijection. Then f is a homeomorphism provided X is compact and Y is Hausdorff.*

This result does not apply to the function (1.27) since the domain of the map is non-compact.

The goal of this section is to define these notions, prove the proposition above, and to give a tools to recognize that a topological space is compact and/or Hausdorff.

1.3.1 Hausdorff spaces

Definition 1.29. Let X be a topological space, $x_i \in X$, $i = 1, 2, \dots$ a sequence in X and $x \in X$. Then x *is a limit of the x_i's* if for any open subset $U \subset X$ containing x there is some N such that $x_i \in U$ for all $i \geq N$.

Caveat: If X is a topological space with the indiscrete topology 1.9, *every point* is the limit of every sequence. There is at most one limit of the x_i if the topological space has the following property:

Definition 1.30. A topological space X is *Hausdorff* if for every $x, y \in X$, $x \neq y$, there are disjoint open subsets $U, V \subset X$ with $x \in U$, $y \in V$.

Lemma 1.31. *The Euclidean space $\mathbb{R}^n$ is Hausdorff. More generally, any subspace $U \subset \mathbb{R}^n$ is Hausdorff.*

Proof. Let $x, y \in U$ with $x \neq y$. Then the balls $B_r(x)$, $B_r(y)$ are open subsets in $\mathbb{R}^n$ which are disjoint if we choose the radius r small enough; for example the choice $r := \mathrm{dist}(x, y)/2$ works. Then $B_r(x) \cap U$ and $B_r(y) \cap U$ are disjoint open neighborhoods of x resp. y in U, showing that U is Hausdorff. $\square$

Lemma 1.32. *Let X be a topological space and A a closed subspace of X. If $x_i \in A$ is a sequence with limit x, then $x \in A$.*

Proof. Assume $x \notin A$. Then x is a point in the open subset $X \setminus A$ and hence by the definition of limit, all but finitely many elements x_n must belong to $X \setminus A$, contradicting our assumptions. $\square$

1.3.2 Compact spaces

Definition 1.33. An *open cover* of a topological space X is a collection of open subsets of X whose union is X. If for every open cover of X there is a finite subcollection which also covers X, then X is called *compact*.

Some books (like Munkres' *Topology*) refer to open covers as *open coverings*, while newer books (and wikipedia) seem to prefer the above terminology, probably for the same reasons as me: to avoid confusion with *covering spaces*, a notion we'll introduce soon.

Example 1.34. (Example of a non-compact space.) The real line $\mathbb{R}$ with the metric topology is non-compact, since the collection of open intervals $(n-1, n+1) \subset \mathbb{R}$ for $n \in \mathbb{Z}$ form an open cover of $\mathbb{R}$, but it does not admit a finite subcover. Indeed, removing just any one interval $(k-1, k+1)$ from the cover, this is no longer a cover of $\mathbb{R}$, since the point $k \in \mathbb{R}$ is not contained in any interval $(n-1, n+1)$ for $n \neq k$.

While it is easy to show that a topological space X is non-compact (by finding an open cover without a finite subcover), showing that X is compact from the definition of compactness is hard: you need to ensure that *every open cover* has a finite subcover. That sounds like a lot of work... Fortunately, there is a very simple classical characterization of compact subspaces of Euclidean spaces, see Theorem 1.37.

Next we will prove some useful properties of compact spaces and maps between them, which will be the essential ingredients of the proof of Proposition 1.28.

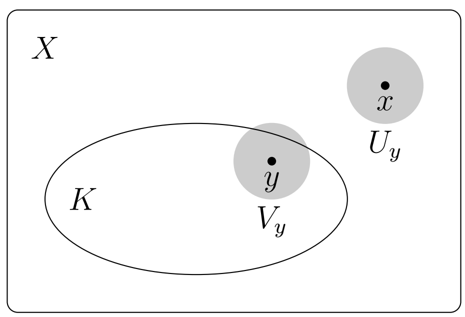
Lemma 1.35. *If $f: X \to Y$ is a continuous map and X is compact, then the image $f(X)$ is compact. In particular, if X is compact, then any quotient space $X/\sim$ is compact, since the projection map $X \to X/\sim$ is continuous with image $X/\sim$.*

Proof. To show that $f(X)$ is compact assume that $\{U_a\}$, $a \in A$ is an open cover of the subspace $f(X)$. Then each U_a is of the form $U_a = V_a \cap f(X)$ for some open subset $V_a \in Y$. Then $\{f^{-1}(V_a)\}$, $a \in A$ is an open cover of X. Since X is compact, there is a finite subset A' of A such that $\{f^{-1}(V_a)\}$, $a \in A'$ is a cover of X. This implies that $\{U_a\}$, $a \in A'$ is a finite cover of $f(X)$, and hence $f(X)$ is compact. □

Lemma 1.36. *1. If K is a closed subspace of a compact space X, then K is compact.*

2. If K is compact subspace of a Hausdorff space X, then K is closed.

Proof. The proof of part (1) is a homework problem. To prove (2), we need to show that $X \setminus K$ is open. So let $x \in X \setminus K$, and we aim to find an open neighborhood U of x which is contained in $U \setminus K$. Since X is Hausdorff, and $x \notin K$, for each $y \in K$ there are disjoint open neighborhoods V_y of y and U_y of x. This situation is illustrated in the following figure.



Then $V_y \cap K$ is an open subset of K, and the collection of subsets $\{V_y \cap K\}_{y \in K}$ is an open cover of K. The compactness of K guarantees that this contains a finite subcover, i.e., there are points $y_1, \dots, y_n \in K$ such that $\bigcup_{i=1,\dots,n} V_{y_i} \cap K = K$. In particular, $K \subset \bigcup_{i=1,\dots,n} V_{y_i}$. Then $U := \bigcap_{i=1,\dots,n} U_{y_i}$ is an open subset containing x; by construction,

$$U \cap \bigcup_{i=1,\dots,n} V_{y_i} = \emptyset \qquad \text{and hence} \qquad U \cap K = \emptyset,$$

which proves that U is an open subset in $U \setminus K$. □

Proof of Proposition 1.28. We need to show that the map $g: Y \to X$ inverse to f is continuous, i.e., that $g^{-1}(U) = f(U)$ is an open subset of Y for any open subset U of X. Equivalently (by passing to complements), it suffices to show that $g^{-1}(C) = f(C)$ is a closed subset of Y for any closed subset C of C.

Now the assumption that X is compact implies that the closed subset $C \subset X$ is compact by part (1) of Lemma 1.36 and hence $f(C) \subset Y$ is compact by Lemma 1.35. The assumption that Y is Hausdorff then implies by part (2) of Lemma 1.36 that $f(C)$ is closed. □

Now we want to apply Proposition 1.28 to show that the continuous bijections that we constructed in Example 1.23 and Lemma 1.26 are in fact homeomorphism. This requires that we are able to show that the domain of the map is compact, which is often done using the the following compactness criterion for subspaces of Euclidean space $\mathbb{R}^n$.

Theorem 1.37. (Heine-Borel Theorem) *A subspace $K \subset \mathbb{R}^n$ is compact if and only if K is a closed subset of $\mathbb{R}^n$ and bounded, i.e., there is some $R > 0$ such that K is contained in the ball $B_R(0)$ of radius R around the origin.*

With this tool in hand, we now revisit Example 1.25(1), (2) and (5):

Example 1.25(1) We have constructed a continuous bijection $f\colon [-1,+1]/\{\pm 1\} \longrightarrow S^1$. The domain of f is compact since $[-1,+1]$ is a closed and bounded subset of $\mathbb{R}$ and hence compact by the Heine-Borel Theorem. It follows that the quotient space $[-1,+1]/\{\pm 1\}$ is compact by Lemma 1.35. The codomain of f is the circle S^1 which is Hausdorff as a subspace of $\mathbb{R}^2$ by Lemma 1.31. Hence f is a homeomorphism by Proposition 1.28.

Example 1.25(2) The same argument as in the previous example shows that the continuous bijection $f\colon D^n/S^{n-1} \to S^n$ constructed in a homework problem is in fact a bijection.

Example 1.25(5) We have constructed a continuous bijection $f\colon D^n/\sim \longrightarrow \mathbb{RP}^n$. The domain is compact, since it is a quotient of the closed bounded subspace $R^n \subset \mathbb{R}^n$. So it remains to show that the codomain $\mathbb{RP}^n$ is Hausdorff. It might be tempting to argue that $\mathbb{RP}^n$ is Hausdorff, since it is a quotient of the Hausdorff space $S^n \subset \mathbb{R}^n$. Alas, Hausdorff is not a property inherited by quotient spaces as the example below shows. So a more detailed argument is needed.

Lemma 1.38. *The projective space $\mathbb{RP}^n$ is Hausdorff.*

Proof. Let $p\colon S^n \to \mathbb{RP}^n$ be the projection map. For $x \in S^n$ let $[x] = p(x) \in \mathbb{RP}^n$ be the equivalence class of x, consisting of the pair of antipodal points $\{x, -x\} \subset S^n$. If $[x] \neq [y] \in \mathbb{RP}^n$, then x, $-x$, y, $-y$ are four distinct points in S^n. Hence for sufficiently small r the four balls of radius r around these points are pairwise disjoint. In particular,

$$U := (B_r(x) \cup B_r(-x)) \cap S^n \qquad \text{and} \qquad V := (B_r(y) \cup B_r(-y)) \cap S^n$$

are disjoint open subsets of S^n. Then $p(U)$, $p(V)$ are disjoint open subsets of $\mathbb{RP}^n$ since $p^{-1}(p(U)) = U$ and $p^{-1}(p(V)) = V$. □

Example 1.39. (Example of a Hausdorff space a quotient of which is not Hausdorff). The interval $(-1, 1)$ is a subspace of $\mathbb{R}$ and so we can form the quotient space

$X := \mathbb{R}/(-1,1)$ where all points belonging to $(-1,1)$ are identified. We claim that X is not Hausdorff; more precisely, we claim that the points $[-1],[1] \in X$ do not have disjoint open neighborhoods $U \ni [-1]$, $V \ni [1]$. To prove this, assume that there are disjoint open neighborhoods. Then their preimages $p^{-1}(U)$, $p^{-1}(V)$ under the projection map $p\colon \mathbb{R} \to X$ are disjoint open subsets of $\mathbb{R}$ with $-1 \in p^{-1}(U)$ and $1 \in p^{-1}(V)$. Due to these being open subsets of $\mathbb{R}$, it follows that $p^{-1}(U)$ must contain some point $x \in (-1,1)$ and that $p^{-1}(V)$ must contain some point $y \in (-1,1)$. It follows that $U \ni p(x) = p(y) \in V$ contradicting the assumption that U and V are disjoint.

The proof of the Heine-Borel Theorem is based on the following two results.

Lemma 1.40. *A closed interval* $[a,b]$ *is compact.*

This lemma has a short proof that can be found in any pointset topology book, e.g., [Mu].

Theorem 1.41. *If* $X_1, \dots, X_n$ *are compact topological spaces, then their product* $X_1 \times \dots \times X_n$ *is compact.*

For a proof see e.g. [Mu, Ch. 3, Thm. 5.7]. The statement is true more generally for a product of *infinitely many* compact space (as discussed in [Mu, p. 113], the correct definition of the product topology for infinite products requires some care), and this result is called *Tychonoff's Theorem*, see [Mu, Ch. 5, Thm. 1.1].

Proof of the Heine-Borel Theorem. Let K be a compact subspace of $\mathbb{R}^n$. Then K is closed by Lemma 1.36(2). The collection $B_r(0) \cap K$, $r \in (0,\infty)$, is an open cover of K. By compactness, K is covered by a *finite* number of these balls; if R is the maximum of the radii of these finitely many balls, this implies $K \subset B_R(0)$, i.e., K is bounded.

Conversely, let $K \subset \mathbb{R}^n$ be closed and bounded, say $K \subset B_r(0)$. We note that $B_r(0)$ is contained in the n-fold product

$$P := [-r,r] \times \dots \times [-r,r] \subset \mathbb{R}^n$$

which is compact by Theorem 1.41. So K is a closed subset of P and hence compact by Lemma 1.36(1). □

Here is another interesting consequence of (the easier part of) the Heine-Borel Theorem.

Proposition 1.42. *If* $f\colon X \to \mathbb{R}$ *is a continuous function on a compact space* X*, then* f *has a maximum and a minimum.*

Proof. $K = f(X)$ is a compact subset of $\mathbb{R}$. Hence K is bounded, and thus K has an infimum $a := \inf K \in \mathbb{R}$ and a supremum $b := \sup K \in \mathbb{R}$. The infimum (resp. supremum) of K is the limit of a sequence of elements in K; since K is closed (by Lemma 1.36 (2)), the limit points a and b belong to K by Lemma 1.32. In other words, there are elements $x_{min}, x_{max} \in X$ with $f(x_{min}) = a \le f(x)$ for all $x \in X$ and $f(x_{max}) = b \ge f(x)$ for all $x \in X$. □

1.3.3 Connected spaces

Definition 1.43. A topological space X is *connected* if it can't be written as decomposed in the form $X = U \cup V$, where U, V are two non-empty disjoint open subsets of X.

For example, if a, b, c, d are real numbers with $a < b < c < d$, consider the subspace $X = (a,b) \amalg (c,d) \subset \mathbb{R}$. The topological space X is not connected, since $U = (a,b)$, $V = (c,d)$ are open disjoint subsets of X whose union is X. This remains true if we replace the open intervals by closed intervals. The space $X' = [a,b] \amalg [c,d]$ is not connected, since it is the disjoint union of the subsets $U' = [a,b]$, $V' = [c,d]$. We want to emphasize that while U' and V' are not open as subsets of $\mathbb{R}$, they are *open subsets of* X', since they can be written as

$$U' = (-\infty, c) \cap X' \qquad V' = (b, \infty) \cap X',$$

showing that they are open subsets for the subspace topology of $X' \subset \mathbb{R}$.

Lemma 1.44. *Any interval I in $\mathbb{R}$ (open, closed, half-open, bounded or not) is connected.*

Proof. Using proof by contradiction, let us assume that I has a decomposition $I = U \cup V$ as the union of two non-empty disjoint open subsets. Pick points $u \in U$ and $v \in V$, and let us assume $u < v$ without loss of generality. Then

$$[u,v] = U' \cup V' \qquad \text{with} \qquad U' := U \cap [u,v] \quad V' := U \cap [u,v]$$

is a decomposition of $[u,v]$ as the disjoint union of non-empty disjoint open subsets U', V' of $[u,v]$. We claim that the supremum $c := \sup U'$ belongs to both, U' and V', thus leading to the desired contradiction. Here is the argument.

- Assuming that c doesn't belong to U', for any $\epsilon > 0$, there must be some element of U' belonging to the interval $(c - \epsilon, c)$, allowing us to construct a sequence of elements $u_i \in U'$ converging to c. This implies $c \in U'$ by Lemma 1.32, since U' is a *closed* subspace of $[u,v]$ (its complement V' is open).
- By construction, every $x \in [u,v]$ with $x > c = \sup U'$ belongs to V'. So we can construct a sequence $v_i \in V'$ converging to c. Since V' is a closed subset of $[u,v]$, we conclude $c \in V'$.

□

Theorem 1.45. (Intermediate Value Theorem) *Let X be a connected topological space, and $f \colon X \to \mathbb{R}$ a continuous map. If elements $a, b \in \mathbb{R}$ belong to the image of f, then also any real number c between a and b belongs to the image of f.*

Proof. Assume that c is not in the image of f. Then $X = f^{-1}(-\infty, c) \cup f^{-1}(c, \infty)$ is a decomposion of X as a union of non-empty disjoint open subsets. □

There is another notion, closely related to the notion of connected topological space, which might be easier to think of geometrically.

Definition 1.46. A topological space X is *path connected* if for any points $x, y \in X$ there is a path connecting them. In other words, there is a continuous map $\gamma \colon [a, b] \to X$ from some interval to X with $\gamma(a) = x$, $\gamma(b) = y$.

Lemma 1.47. *Any path connected topological space is connected.*

Proof. Using proof by contradiction, let us assume that the topological space X is path connected, but not connected. So there is a decomposition $X = U \cup V$ of X as the union of non-empty open subsets $U, V \subset X$. The assumption that X is path connected allows us to find a path $\gamma \colon [a, b] \to X$ with $\gamma(a) \in U$ and $\gamma(b) \in V$. Then we obtain the decomposition

$$[a, b] = f^{-1}(U) \cup f^{-1}(V)$$

of the interval $[a, b]$ as the disjoint union of open subsets. These are non-empty since $a \in f^{-1}(U)$ and $b \in f^{-1}(V)$. This implies that $[a, b]$ is not connected, the desired contradiction. $\square$

For typical topological spaces we will consider, the properties "connected" and "path connected" are equivalent. But here is an example known as the *topologist's sine curve* which is connected, but not path connected, see [Mu, Example 7, p. 156]. It is the following subspace of $\mathbb{R}^2$:

$$X = \{(x, \sin \frac{1}{x}) \in \mathbb{R}^2 \mid 0 < x < 1\} \cup \{(0, y) \in \mathbb{R}^2 \mid -1 \le y \le 1\}.$$

References

[Mu] Munkres, James R. *Topology: a first course*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975. xvi+413 pp.

2 Topological manifolds

The purpose of this section is to provide interesting examples of topological spaces and homeomorphisms between them. There are many examples of "weird" topological spaces. There are non-Hausdorff spaces (they don't have well-defined limits) or the topologist's sine curve, which is connected, but not path connected. While there is a huge literature concerning pathological topological spaces, I must admit that I find those examples most interesting that "show up in nature". For example, topological spaces that appear as "configuration spaces" or "phase spaces" of physical systems. Often these are a particularly nice kind of topological space known as *manifold*.

There is much to say about manifolds. For example, you can find the text books *Introduction to topological manifolds* and *Introduction to smooth manifolds* on the reserved book shelf for this course. For this section, our focus is to discuss manifolds of dimension 2. Unlike higher dimensional manifolds, we can represent manifolds of dimension 2 by pictures, which greatly helps the intuition about these objects.

2.1 Definition and basic examples of manifolds

Definition 2.1. A *manifold of dimension n* or *n-manifold* is a topological space X which is locally homeomorphic to $\mathbb{R}^n$, that is, every point $x \in X$ has an open neighborhood U which is homeomorphic to an open subset V of $\mathbb{R}^n$. Moreover, it is useful and customary to require that X is Hausdorff (see Definition 1.30) and *second countable*, which means that the topology of X has a countable basis.

In most examples, the technical conditions of being Hausdorff and second countable are easy to check, since these properties are inherited by subspaces.

Homework 2.2. Show that a subspace of a Hausdorff space is Hausdorff. Show that a subspace of a second countable space is second countable.

Examples of manifolds.

1. Any open subset $U \subset \mathbb{R}^n$ is an n-manifold. The technical condition of being a second countable Hausdorff space is satisfied for U as a subspace of the second countable Hausdorff space $\mathbb{R}^n$; a countable basis for the topology on $\mathbb{R}^n$ is provided by the collection of balls $B_r(x)$, for which the radius r as well as all components of $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ are rational numbers.

2. The n-sphere $S^n := \{x \in \mathbb{R}^n \mid ||x|| = 1\}$ is an n-manifold. To prove this, let us look at the subsets

$$U_i^+ := \{(x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid x_i > 0\} \subset S^n$$
$$U_i^- := \{(x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid x_i < 0\} \subset S^n$$

   We want to argue that the map

$$\phi_i^\pm \colon U_i^\pm \longrightarrow \mathring{D}^n \qquad \text{given by} \qquad \phi_i^\pm(x_0, \dots, x_n) := (x_0, \dots, x_{i-1}, \widehat{x}_i, x_{i+1}, \dots, x_n)$$

   is a homeomorphism, where $\mathring{D}^n := \{(v_1, \dots, v_n) \in D^n \mid v_1^2 + \dots + v_n^2 < 1\}$ is the open n-disk. It is easy to verify that the map

$$\mathring{D}^n \longrightarrow U_i^\pm \qquad v = (v_1, \dots, v_n) \mapsto (v_1, \dots, v_i, \pm\sqrt{1 - ||v||^2}, v_{i+1}, \dots, v_n)$$

is in fact the inverse to $\phi_i^{\pm}$. Here $||v||^2 = v_1^2 + \cdots + v_n^2$ is norm squared of $v \in \mathring{D}^n$. Both maps, $\phi_i^{\pm}$ and its inverse, are continuous since all their components are continuous. This shows that $\phi_i^{\pm}$ is in fact a homeomorphism, and hence the n-sphere S^n is a manifold of dimension n.

Homework 2.3. Show that the product $X \times Y$ of manifold X of dimension m and a manifold Y of dimension n is a manifold of dimension $m+n$. Make sure to prove that $X \times Y$ is second countable and Hausdorff.

Homework 2.4. Show that the real projective space $\mathbb{RP}^n$ is manifold of dimension n. Make sure to prove that $\mathbb{RP}^n$ is second countable (we know it is Hausdorff by Lemma 1.38).

The following result is useful for showing that quotient spaces are second countable.

Lemma 2.5. *Let X a topological space and let $p\colon X \twoheadrightarrow X/\sim$ be the projection map onto a quotient space of X. If p is an open map (i.e., the images of open subsets $U \subset X$ under p are open in $X/\sim$), and X is second countable, then $X/\sim$ is second countable.*

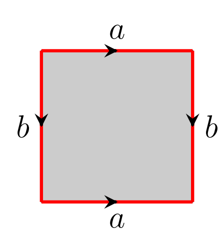
Proof of Lemma. Let $\mathcal{B}$ be countable basis for the topology of X. We claim that the collection $\mathcal{B}' := \{p(B) \subset Y \mid B \in \mathcal{B}\}$ is a basis for the topology of $X/\sim$. To prove this, let $U \subset X/\sim$ be open. Then $p^{-1}(U) \subset X$ is open, and hence $p^{-1}(U) = \bigcup_{a \in A} B_a$ is a union subsets $B_a \in \mathcal{B}$. Then

$$U = p(p^{-1}(U)) = p(\bigcup_{a \in A} B_a) = \bigcup_{a \in A} p(B_a)$$

is a union of subsets belonging to $\mathcal{B}'$. This shows that $\mathcal{B}'$ is a basis for the topology of $X/\sim$. □

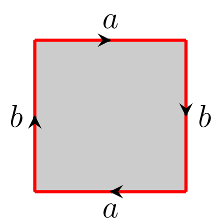
2.6. Examples of manifolds of dimension 2.

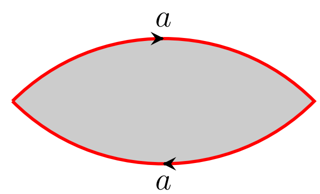
1. The 2-torus T can be described as subspace of $\mathbb{R}^3$, as the product $S^1 \times S^1$ and as the quotient of the square $[0,1] \times [0,1]$ by the identifying its edges as indicated in the following picture (see Example 1.23(4)):



From the description of the torus as the product $S^1 \times S^1$ and Lemma **??** it follows that the torus is a manifold of dimension 2.

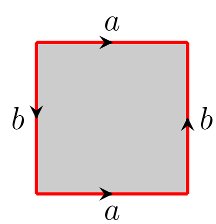
2. The real projective plane $\mathbb{RP}^2$. We recall from Example 1.23 (8) and Lemma 1.26 that $\mathbb{RP}^2$ is homeomorphic to the quotient spaces of the square resp. disk by identifying edges as indicated by the following pictures.





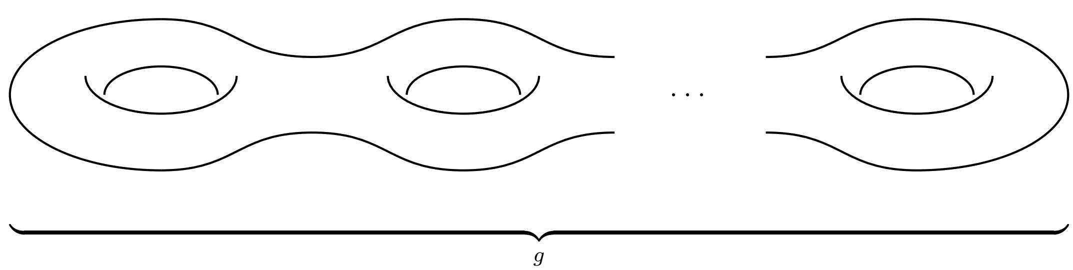
We prefer to draw the disk D^2 as a bigon here, since our goal is to describe *all* compact connected 2-manifolds as quotients of polygons by suitably identifying edges. We think of the bigon as a polygon with two vertices and two edges.

3. In Example 1.23(7) we defined the Klein bottle K as the quotient of the square with the identification of edges given by the following picture.



It is not hard to verify directly that K is a manifold of dimension 2 (draw open neighborhoods of a point in the interior of the square, on an edge of the square and of the one point of K represented by the vertices to convince yourself). Alternatively, we will see in Lemma **??** that the Klein bottle is homeomorphic to the connected sum $\mathbb{RP}^2 \# \mathbb{RP}^2$ of two copies of the projective plane $\mathbb{RP}^2$, which implies in particular that K is a 2-manifold.

4. The surface Σ_g of genus g is the subspace of $\mathbb{R}^3$ given by the following picture:



Here g is the number of "holes" of Σ_g. In particular Σ_1, the surface of genus 1, is the torus. By convention, the surface Σ_0 of genus 0 is the 2-sphere S^2. Since we have described the surface of genus g as a subspace of $\mathbb{R}^3$ given by a picture rather than a formula, it is impossible to give a precise argument that this subspace is locally homeomorphic to $\mathbb{R}^2$, but hopefully the picture makes this obvious at a heuristic level.

2.2 The connected sum construction

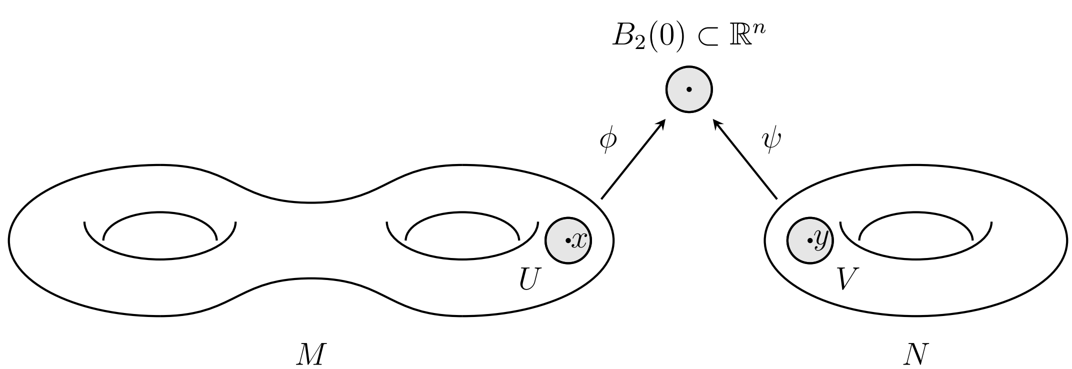
This construction produces a new manifold $M\#N$ of dimension n from two given manifolds M and N of dimension n. The manifold $M\#N$ is called the *connected sum* of M and N. The construction proceeds as follows. First we make some choices:

- We pick points $x \in M$ and $y \in N$.
- We pick a homeomorphism ϕ between an open neighborhood U of x and the open ball $B_2(0)$ of radius 2 around the origin $0 \in \mathbb{R}^n$. Similarly, we pick a homeomorphism $\psi\colon V \xrightarrow{\approx} B_2(0)$ where $V \subset N$ is an open neighborhood of $y \in N$.

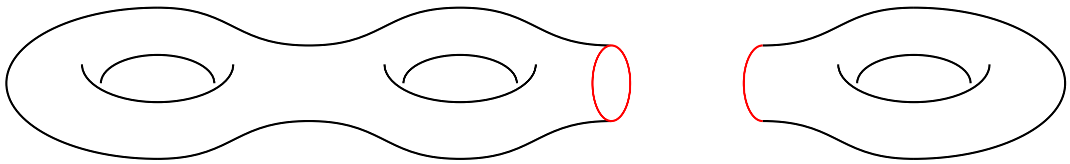
The existence of homeomorphisms ϕ, ψ with these properties follows from the assumption that M, N are manifolds of dimension n. This implies that there is an open neighborhood $U' \subset M$ of x and a homeomorphism ϕ' between U' and an open subset $V' \subset \mathbb{R}^n$. Composing ϕ by a translation in $\mathbb{R}^n$ we can assume that $\phi(x) = 0 \in \mathbb{R}^n$. Since V' is open, there is some $\epsilon > 0$ such that the open ball $B_\epsilon(0)$ of radius ϵ around $0 \in \mathbb{R}^n$ is contained in V'. Then restricting ϕ' to $U := (\phi')^{-1}(B_\epsilon(0)) \subset M$ gives a homeomorphism between U and $B_\epsilon(0)$. Then the composition

$$U \xrightarrow[\approx]{\phi'_U} B_\epsilon(0) \xrightarrow[\approx]{\text{multiplication by } 2/\epsilon} B_2(0)$$

is the desired homeomorphism ϕ between a neighborhood U of $x \in M$ and $B_2(0) \subset \mathbb{R}^n$. Analogously, we construct the homeomorphism ψ. Here is a picture illustrating the situation.



The next step is to remove the open disc $\phi^{-1}(B_1(0))$ from the manifold M and the open disc $\psi^{-1}(B_1(0))$ from the manifold N. The following picture shows the resulting topological spaces $M \backslash \phi^{-1}(B_1(0))$ and $N \backslash \psi^{-1}(B_1(0))$. Here the red circles mark the points corresponding to the sphere $S^{n-1} \subset B_2(0)$ via the homeomorphisms ϕ and ψ, respectively.



$M \setminus \phi^{-1}(B_1(0))$ $\qquad$ $N \setminus \psi^{-1}(B_1(0))$

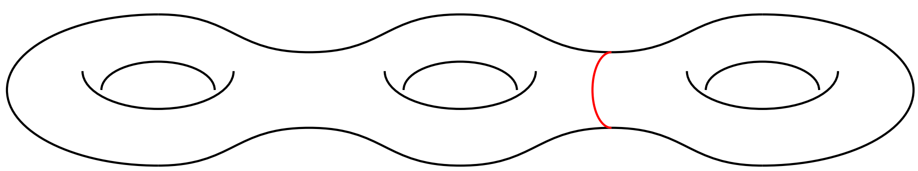
The final step is to pass to a quotient space of the union

$$M \setminus \phi^{-1}(B_1(0)) \quad \cup \quad N \setminus \psi^{-1}(B_1(0))$$

given by identifying points in $\phi^{-1}(S^{n-1})$ with their images under the homeomorphism

$$\phi^{-1}(S^{n-1}) \xrightarrow{\approx} \psi^{-1}(S^{n-1}) \qquad z \mapsto \psi^{-1}(\phi(z)).$$

The connected sum $M\#N$ is this quotient space. In terms of our pictures, the manifold $M\#N$ is obtained by gluing the two red circles, and is given by the following picture.



$M\#N$

Question: Is $M\#N$ independent of the choices made in its construction? A crucial ingredient of the construction of the connected sum $M\#N$ are the homeomorphisms $M \supset U \xrightarrow{\phi} B_2(0) \subset \mathbb{R}^n$ and $N \supset V \xrightarrow{\psi} B_2(0) \subset \mathbb{R}^n$. Since we remove in the first step of the construction the open disks $\phi^{-1}(B_1(0) \subset M$ and $\psi^{-1}(B_1(0) \subset N$, the set $M\#N$ will be different if we remove different disks.

Fact: Up to homeomorphism, the topological space $M\#N$ does not depend on these choices if "we are careful with orientations". Fortunately, for 2-dimensional manifolds, it is always independent of the choices.

Later this semester we will define what an orientation for a smooth manifold is (which is easier than defining an orientation for a topological manifold). We will restrict us to 2-manifolds, so orientations don't play a role, and we use the fact above for 2-manifolds without proof.

Example 2.7. (Examples of connected sums).

1. Our pictures above show that the connected sum $\Sigma_2 \# T$ of the surface of genus two and the torus is homeomorphic to the surface of genus 3. More generally, it is clear from drawing appropriate pictures that the connected sum $\Sigma_g \# \Sigma_{g'}$ is homeomorphic to $\Sigma_{g+g'}$ of genus $g + g'$. It follows that

$$\underbrace{T \# T \# \ldots \# T}_{g} \approx \Sigma_g.$$

   Strictly speaking, we have mathematically defined what we mean by a surface of genus g only for $g = 1$ (it is the torus T) and for $g = 0$ (it is the sphere S^2). For $g > 1$, we have only drawn a picture of what we mean by a surface of genus g, and hence we can prove the statement $\Sigma_g \# \Sigma_{g'} \approx \Sigma_{g+g'}$ only at that level of precision: by drawing pictures. From mathematical point of view, we can (and will) view the above homeomorphism now as the *definition* of the surface of genus g.

2. The connected sum $X_k := \underbrace{\mathbb{RP}^2 \# \ldots \# \mathbb{RP}^2}_{k}$ is a 2-manifold that, together with the surface of genus g, plays an important role in the Classification Theorem for compact connected 2-manifolds 2.8. Munkres refers to X_k as the *k-fold projective plane* [Mu2, Definition on p. 462].

2.3 Classification of compact connected 2-manifolds

Theorem 2.8. (Classification of compact connected 2-manifolds.) *Every compact connected manifold of dimension 2 is homeomorphic to exactly one of the following manifolds:*

- *The surface of genus g, denoted Σ_g which is the connected sum $\underbrace{T \# \ldots \# T}_{g}$ of g copies of the torus T, for $g > 0$, and the 2-sphere S^2 for $g = 0$.*
- *The connected sum $X_k = \underbrace{\mathbb{RP}^2 \# \ldots \# \mathbb{RP}^2}_{k}$ of k copies of the real projective plane $\mathbb{RP}^2$, $k \geq 1$;*

In this class, we won't give a complete proof of this classification result, but we will introduce the techniques used for the proof of this theorem (see e.g., [Mu2, Ch. 12]), and we prove partial results. Like any classification result, the classification of 2-manifolds involves two quite distinct aspects:

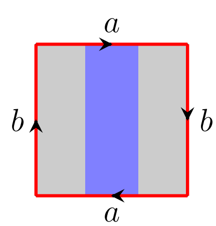
(1) the proof that the 2-manifolds Σ_0, Σ_1, Σ_2,..., X_1, X_2,... are pairwise non-homeomorphic.

(2) the proof that any compact connected 2-manifold Σ is homeomorphic to a manifold on this list.

To distinguish Σ_g, the surface of genus g, from X_k, the connected sum of k copies of $\mathbb{RP}^2$, the notion "orientable" can be used (alternatively, these can be distinguished by their abelianized fundamental group as we will show in section **??**). For a manifold M of any dimension n one can define what it means for M to be "orientable", but that requires tools not available to us in this course, namely homology groups. In section **??** we will define what an "orientation" is on a *smooth* n-manifold M (this will be needed to integrate differential forms of degree n over M), and M is "orientable" if and only if there is an orientation on M. He we give a definition of orientability for 2-manifold which is easy to state.

Definition 2.9. A 2-manifold Σ is *non-orientable* if it contains a Möbius band.

For example, the real projective plane $\mathbb{RP}^2$ is non-orientable, since representing $\mathbb{RP}^2$ as a square with boundary identifications (as in Example 2.6(2)), the blue rectangle inside the square becomes a Möbius band inside $\mathbb{RP}^2$ after the gluing the edges.



It follows that the connected sum $\mathbb{RP}^2\#\Sigma$ with any other surface also is non-orientable, since we can choose the disc to be removed from $\mathbb{RP}^2$ in order to form the connected sum to be located in the complement of the Möbius band in $\mathbb{RP}^2$. In particular, $X_k = \mathbb{RP}^2\#\dots\#\mathbb{RP}^2$ is non-orientable.

Heuristically, it is quite clear that Σ_g, the surface of genus g does not contain a Möbius band using the following argument. Think of Σ_g as being made from some thin material, and paint the outside of it with a different color than the inside. Then whenever you cut a band from Σ_g, it will have *two* colors, unlike the Möbius band, which can only be painted in one color. Needless to say, this isn't a *proof.* While this line of argument can be made into a proof, it seems better to invest time to develop powerful tools, like the fundamental group, as we will do in the next chapter, which can easily distinguish orientable from non-orientable 2-manifolds.

In the next section, we will introduce the Euler characteristic for compact 2-manifolds and will show that Σ_g is not homeomorphic to $\Sigma_{g'}$ for $g \neq g'$ and that Σ_k is not homeomorphic to $\Sigma_{k'}$ for $k \neq k'$, see 2.16.

In Examples 2.6 we showed that the torus, the Klein bottle and the projective plane $\mathbb{RP}^2$ can be described as quotients of a square (resp. a bigon in the case of $\mathbb{RP}^2$) by identifying edges with the same label. In section 2.5 we will show that *every* compact connected 2-manifold can be described as quotients of polygons with labeled edges by identifying edges with the same label.

2.4 The Euler characteristic of compact 2-manifolds

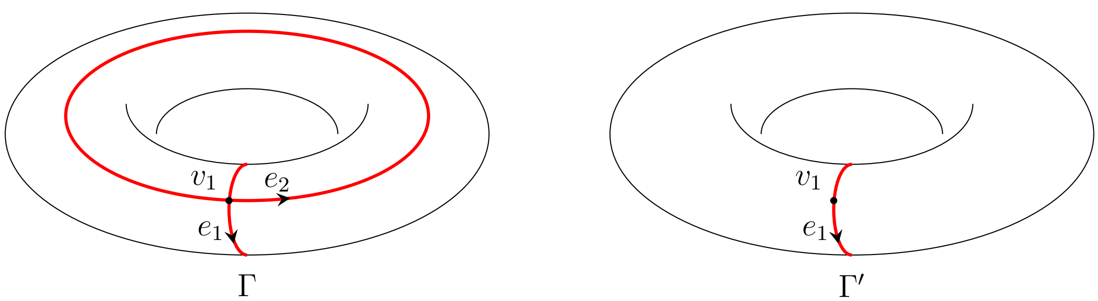
In this section we introduce the Euler characteristic of compact 2-manifolds. This invariant will allows us to show that some compact 2-manifolds are not homeomorphic. Our definition of the Euler characteristic is very geometric (and not particularly precise).

Definition 2.10. Let Σ be a compact 2-manifold. A *graph* Γ on Σ is a collection of finitely many points $v_1, \dots, v_k \in \Sigma$ (called *vertices*) and finitely many paths $e_i \colon [0,1] \to \Sigma$, $i = 1, \dots, \ell$ (called *edges*) such that

- the endpoints of e_i belong to the set of vertices $V := \{v_1, \dots, v_k\}$.
- the only intersection points of paths occur at their endpoints.

We call a graph Γ a *pattern of polygons* if the complement of all vertices and edges in Σ is a disjoint union of subspaces homeomorphic to open 2-disks.

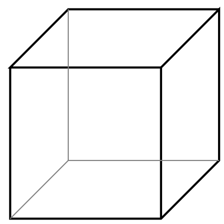
Example 2.11. Both pictures below show examples of graphs Γ, Γ' on the torus T. The complement of Γ in T is the open square, and hence Γ is a pattern of polygons. The complement of Γ' is a cylinder, and so Γ' is not a pattern of polygons.



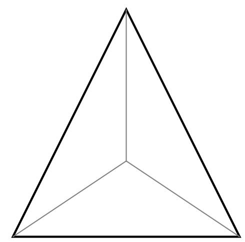
Let Γ be a pattern of polygons on a compact 2-manifold Σ.

$$\chi(\Sigma; \Gamma) := \#\{\text{vertices}\} - \#\{\text{edges}\} + \#\{\text{polygons}\}.$$

For example, the surface of a cube is homeomorphic to the sphere S^2. Via this homeomorphisms, the vertices, edges and faces of the cube can be interpreted as a pattern of polygons Γ_{cube} on S^2. More physically, think of the edges of the cube as a wireframe inside of a translucent sphere equipped with a light source at its center. Then the shadows of the edges give pattern of polygons (in this case quadrilaterals) on the sphere. Similarly, the tetrahedron can be interpreted as giving a pattern of polygons Γ_{tetra} on the sphere.



Γ_{cube}



Γ_{tetra}

We observe that

$$\chi(S^2; \Gamma_{\text{cube}}) = 8 - 12 + 6 = 2$$
$$\chi(S^2; \Gamma_{\text{cube}}) = 4 - 6 + 4 = 2$$

give the same number, independent whether we choose the pattern Γ_{cube} or Γ_{tetra} on S^2. This is in fact true generally:

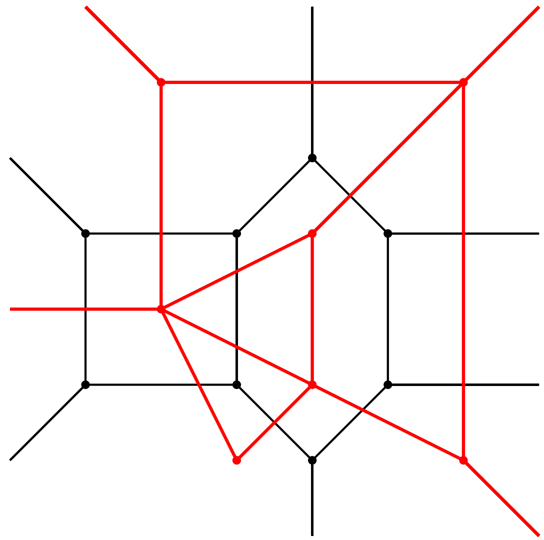
Lemma 2.12. *Let Γ, Γ' be two patterns of polygons on a compact 2-manifold Σ. Then $\chi(\Sigma; \Gamma) = \chi(\Sigma; \Gamma')$.*

Proof. **Step 1.** By moving the vertices and edges of the graph Γ' a little bit, we can assume that the vertex sets of Γ and Γ' are disjoint, and that that there are only finitely many intersection points between edges of Γ and edges of Γ'. We claim that then there is a pattern of polygons Γ'' which is a *refinement* of both, Γ and Γ'. This means that Γ'' can be obtained from Γ (resp. Γ') by inductively adding new vertices on the interior of existing edges, and adding new edges between two vertices of a polygon.

The graph Γ'' is constructed as follows:

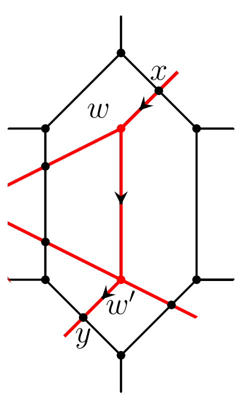
- The vertices of Γ'' are the vertices of Γ, the vertices of Γ' and all intersection points of edges of Γ and edges of Γ'.
- The edges of Γ'' are segments of edges of Γ or Γ' whose endpoints are vertices of Γ''.

The following picture shows (part of) the graph Γ on some surface Σ with black vertices and edges and (part of) the graph Γ' colored red.



The graph Γ'' is simply the graph you see when we you ignore the color (and indicate that every intersection point is a vertex by drawing a little dot). It is clear that the complement of the graph Γ'' in Σ is again a disjoint union of open balls, since each connected component of the complement is an (open) polygon obtained by subdividing a polygon of Γ by edges.

To show that Γ'' is a refinement of Γ we first add all intersection points of edges of Γ and Γ' as new vertices (which subdivide the existing edges of Γ). Before we can add vertices of Γ' we need to add new edges: if w is a vertex of Γ' in the interior of some polygon P of Γ (e.g., the top red vertex in the black hexagon in the center of the picture above), there is a path through w along red edges that starts at some intersection point x of a red edge with an edge of P and ends at an intersection point y of some red edge with an edge of P. The following picture shows the vertex w, and the path along red edge segments (indicated by arrows) starting at x and ending at y.



We add this path as a new edge to our graph. Then we can add the red vertices w and w' to our graph (thus subdividing our new edge). Finally, we add the three additional red edges that connect w resp. w' to intersection points on the boundary of P. Doing this for all polygons of Γ we see that Γ'' is a refinement of Γ.

Step 2. Let Γ_1 be a pattern of polygons on Σ, and let Γ_2 be obtained by adding a new vertex to the interior of an edge of a graph Γ_1. We claim that $\chi(\Sigma;\Gamma_2) = \chi(\Sigma;\Gamma_1)$. To prove this, let $V(\Gamma_i)$ be the number of vertices, $E(\Gamma_i)$ the number of edges and $F(\Gamma_i)$ the number of faces of Γ_i. We note that $V(\Gamma_2) = V(\Gamma_1)$, due to the additional vertex, and $E(\Gamma_2) = E(\Gamma_1)$, since the creation of the new vertex on an edge subdivides that edge in two edges. The number of faces is unchanged and hence

$$\chi(\Sigma;\Gamma_2) = V(\Gamma_2) - E(\Gamma_2) + F(\Gamma_2) = (V(\Gamma_1)+1) - (E(\Gamma_1)+1) + F(\Gamma_1) = \chi(\Sigma;\Gamma_1).$$

Step 3. Let Γ_1 be a pattern of polygons on Σ, and let Γ_2 be obtained by introducing a new edge which connects two vertices of some polygon in Γ_1. Then the number of edges and faces goes up by one while the number of vertices is unchanged. Hence again, $\chi(\Sigma,\Gamma_2) = \chi(\Sigma,\Gamma_1)$.

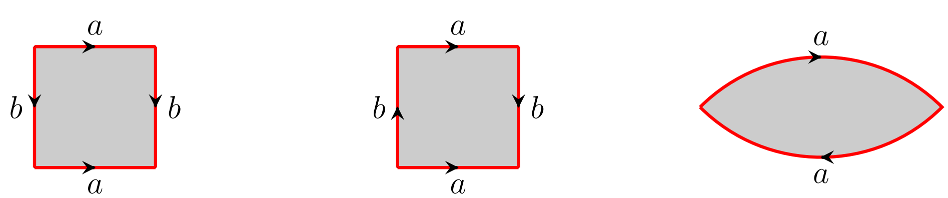
Steps 2 and 3 show that the alternating sum $\chi(\Sigma;\Gamma)$ doesn't change when we refine the graph Γ by adding vertices or edges. In particular, due to the existence of a common refinement Γ'' of graphs Γ and Γ' we conclude that

$$\chi(\Sigma,\Gamma) = \chi(\Sigma,\Gamma'') = \chi(\Sigma,\Gamma').$$

□

Definition 2.13. Let Σ be a compact 2-manifold. The *Euler characteristic of* Σ is defined to be the integer $\chi(\Sigma) := \chi(\Sigma;\Gamma)$.

To calculate the Euler characteristic of the torus T, the Klein bottle K and the real projective plane $\mathbb{RP}^2$ we use the fact that all three spaces can be described as quotients of polygons by identifying edges equipped with the same label.



T K $\mathbb{RP}^2$

The square from which the torus and the Klein bottle is built has four vertices, four edges and one face. However, we need to count vertices, edges and faces not for the square, but for the quotient space. The edges labeled a (resp. b) map to the *same* edge in the quotient under the projection map. Similarly, all four vertices (of the square) and the two vertices (of the bigon) map to the same vertex in the quotient. This shows that

$$\begin{aligned}\chi(T) =&1-2+1=0\\ \chi(K) =&1-2+1=0\\ \chi(\mathbb{RP}^2) =&1-1+1=-1\end{aligned}$$

We note that a homeomorphism $f\colon \Sigma \xrightarrow{\approx} \Sigma'$ between two compact 2-manifolds allows us to interpret a pattern of polygons Γ on Σ as a pattern of polygons on Σ'. This shows that the Euler characteristic of homeomorphic manifolds agrees. In others words, the Euler characteristic is an invariant that allows us to show that some 2-manifolds are not homeomorphic. In particular, our calculations above imply:

Corollary 2.14. *The compact* 2*-manifolds* S^2*,* T *and* $\mathbb{RP}^2$ *are pairwise non homeomorphic to each other.*

Lemma 2.15. *Let* Σ*,* Σ' *be compact* 2*-manifolds. Then* $\chi(\Sigma\#\Sigma') = \chi(\Sigma) + \chi(\Sigma') - 2$*.*

The proof is a homework problem.

Applying this inductively to the connected sums

$$\Sigma_g = \underbrace{T\#\dots\#T}_{g} \qquad \text{and} \qquad X_k = \underbrace{\mathbb{RP}^2\#\dots\#\mathbb{RP}^2}_{k}$$

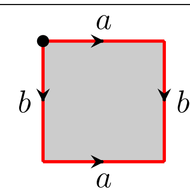
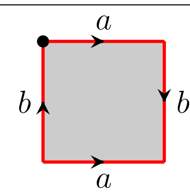
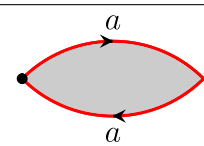
leads to the following result.

Corollary 2.16. $\chi(\Sigma_g) = 2-2g$ *and* $\chi(X_k) = 2-k$*. In particular,* Σ_g *is homeomorphic to* $\Sigma_{g'}$ *if and only if* $g = g'$*, and* $X_k \approx X_{k'}$ *if and only if* $k = k'$*.*

2.5 A combinatorial description of compact connected 2-manifolds

The Euler characteristic is an invariant which is very useful to show that two compact 2-manifolds are *not homeomorphic*. As advertised earlier, the goal of this section is to show that every compact connected 2-manifold is homeomorphic to the quotient of a polygon with labeled edges by identifying edges with the same label (see Prop. 2.25). Moreover, we will show that certain ways of relabeling the edges yields the same quotient space. This is the key technique used in the proof of the Classification Theorem 2.8. In this class we will only illustrate this technique to show that the Klein bottle K is homeomorphic to $\mathbb{RP}^2\#\mathbb{RP}^2$ (a homework problem) and that $T\#\mathbb{RP}^2$ is homeomorphic to $\mathbb{RP}^2\#\mathbb{RP}^2\#\mathbb{RP}^2$ (see Lemma **??**).

We recall that all three manifolds T, K and $\mathbb{RP}^2$ can be described as polygons (squares resp. bigons) with edge identifications as shown in the following table.

space	combinatorial picture	word
T = torus		$aba^{-1}b^{-1}$
K = Klein bottle		$aba^{-1}b$
$\mathbb{RP}^2$ = projective plane		aa

(2.17)

If we choose a distinguished vertex for these polygons, indicated by a black dot in the picture above, then the labeling of the edges by letters a, b and arrows can be encoded as follows. Going along the edges of the polygon clockwise, starting at the distinguished vertex, we write down for each edge

- the letter a if the edge has label a and the arrow of the edge points in the clockwise direction, or
- the letter a^{-1} if the edge has label a and the arrow of the edge points in the counter-clockwise direction.

Doing this in order for all of the edges of the polygon, we obtain a string of symbols, that is, a *word* whose *letters* are the edge labels and their inverses. The words obtained this way for our examples are shown in the third column of the table above. This process can be reversed by interpreting a word W consisting of letters $a, a^{-1}, b, b^{-1}, \dots$ as giving the edges of a polygon P a label and a direction. This in turn determines an equivalence relation $\sim_W$ on P according to which corresponding points on edges with the same label are identified, and hence a quotient space $\Sigma(W) := P/\sim_W$. Here is the formal definition.

Definition 2.18. Let L be a set whose elements we refer to as *labels*, typically $a, b, \dots \in L$. Let $W = x_1 x_2 \dots x_n$ be an n letter word with letters x_i belonging to the *alphabet*, which is the set consisting of the symbols ℓ and ℓ^{-1} for $\ell \in L$. So typically, our alphabet is the set $A = \{a, a^{-1}, b, b^{-1}, \dots\}$.

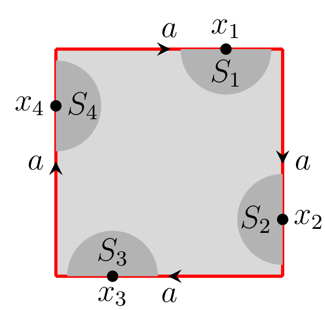
Let P_n be an n-gon (i.e., the polygon with n edges) with a distinguished vertex. Going around P_n clockwise, starting at the distinguished vertex, label the edges of P_n by

$x_1, x_2, \ldots, x_n$. More precisely, if $x_i = \ell$ or $x_i = \ell^{-1}$ label the i-th edge by the label ℓ and equip it with an arrow according to the convention explained above. Let $\sim_W$ be the equivalence relation on P_n which identifies any point on an edge labeled ℓ with the corresponding point on any other edge with the same label. Then the *topological space associated to* W, denoted $\Sigma(W)$ or $\Sigma(x_1x_2 \ldots x_n)$ is defined to be the quotient space $P_n/\sim_W$.

The word associated to a polygon with labeled oriented edges and a distinguished vertex depends on the choice of that vertex. For example, for the torus with the labeling as in (2.17) and the top left vertex as the distinguished vertex, the associated word is $aba^{-1}b^{-1}$. If we choose the top right vertex (resp. the bottom right vertex) as distinguished vertex, the associated word will be $ba^{-1}b^{-1}a$ (resp. $a^{-1}b^{-1}ab$). In other words, changing the distinguished vertex changes the associated word W by a cyclic permutation; it does *not change* the quotient space $\Sigma(W)$. Similarly, renaming the labels of the polygon (by a bijection of label sets) or interchanging for a label ℓ the letters ℓ and ℓ^{-1} in the word W does not change the quotient. For future reference, we state this remark as a lemma.

Lemma 2.19. *The quotient space $\Sigma(W) = P_n/\sim_W$ is homeomorphic to $\Sigma(W')$ if W' is obtained from W by by cyclically permuting the letters of W, renaming the labels involved via a bijection of label sets, or by interchanging the letters ℓ and ℓ' for a particular label ℓ.*

2.20. Warning. While all the spaces $\Sigma(W)$ mentioned as examples above were *manifolds*, this is not generally the case. For example $\Sigma(W) = \Sigma(aaaa)$ is *not* a manifold. To see this, consider a point x_1 in the interior of an edge of the square P_4. The equivalence class $[x_1] \in\in \Sigma(W) = P_4/\sim_W$ consists of the four points x_1, x_2, x_3, x_4, one on each edge as shown in the picture below. An open neighborhood of a point x_i consists of the dark semi-disk S_i containing x_i.

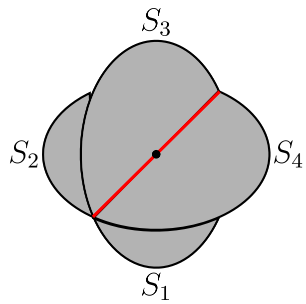
$\Sigma(aaaa) =$ 

It follows that an open neighborhood of $[x_1] \in \Sigma(W)$ has the form

$$(S_1 \cup S_2 \cup S_3 \cup S_4) / \sim,$$

where the equivalence relation is the restriction of $\sim_W$ to the union of these semi-circles. More geometrically, this is obtained by gluing these four semi-disks along there straight edge.

Here is a picture of that quotient space; the red line is the line where the semi-disks are glued together and the marked point is the point $[x_1] \in \Sigma(W)$.



Remark 2.21. It is easy to show by elementary means that $\Sigma(W)$ is a 2-manifold if each label occurs exactly twice in the word W. Conversely, the argument used in the example above can be expanded to show that if the word W contains a label ℓ only once or more that twice, then the quotient $\Sigma(W)$ is not a 2-manifold (we will be able to show this in the following chapter using the fundamental group).

Proposition 2.22. *Every compact connected* 2*-manifold* Σ *is homeomorphic to the quotient space* $\Sigma(W)$ *for a suitable word* Σ.

Proof. Let Γ be a pattern of polygons on Σ, i.e., a graph on Σ such that the complement of all edges is a disjoint union of 2-discs (see Def. 2.10). Orient all edges of Σ and label them by a label set L (mathematically speaking, choose a bijection between the set of edges and L). Cutting Σ along all edges of Γ then results in a disjoint union of polyhedra $P_{n_1}, \dots P_{n_k}$, whose edges are labeled and oriented (here k is the number of faces of Γ, and P_{n_i} is the number of edges of the i-th face). Then the quotient space of the disjoint union $P := P_{n_1} \amalg \cdots \amalg P_{n_k}$ given by identifying all the edges with the same label can then be identified with the original 2-manifold (taping the pieces $P_{n_1}, \dots, P_{n_k}$ together again is the inverse of the process of cutting Σ along all edges of the graph Γ). In other words,

$$\Sigma \approx (P_{n_1} \amalg \cdots \amalg P_{n_k}))/ \sim,$$

where the equivalence relation $\sim$ is given by identifying edges with the same label.

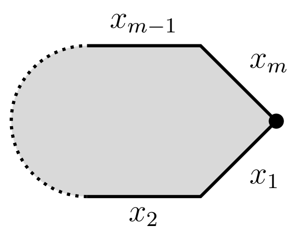
If there is only one polygon, then we are done. If $k > 1$, we claim that the first polygon must have an edge with a label which also occurs as a label of an edge of one of the other polygon. Otherwise, all edges of the first polygon would be glued only to other edges of the same polygon, and hence the quotient space is the disjoint union of the quotient of the first polygon and the quotient of the union of the other polygons. That would contradict our assumption that Σ is connected.

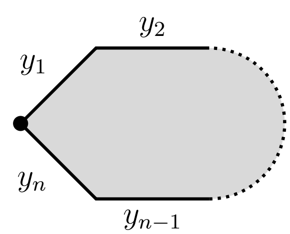
We note that cutting Σ along an edge of Γ labeled ℓ results in *two edges* of P labeled ℓ. So if ℓ is the label of an edge of the first polygon as well as the label of some other polygon, we can glue those two polygons together along the edge labeled ℓ, thus reducing the number of polygons by one. The label ℓ is no longer a label of any edge of the new set of polygons, but identifying all edges with the same label still yields the manifold Σ.

Repeating this process, we can reduce to *one* polygon with labeled oriented edges, whose quotient obtained by identifying edges with the same label is still homeomorphic to Σ. $\square$

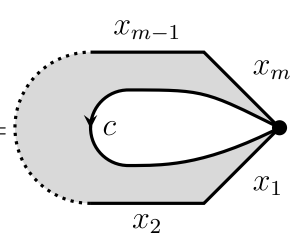
Proposition 2.23. *Let M, N be two compact connected 2-manifolds which are described combinatorially as $M = \Sigma(W_1)$, $N = \Sigma(W_2)$, where W_1 and W_2 are words from disjoint alphabets. Then the connected sum $M\#N$ is homeomorphic to $\Sigma(W_1W_2)$.*

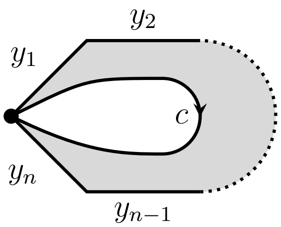
Proof. Let $W_1 = x_1 \dots x_m$ and $W_2 = y_1 \dots y_n$. Then M and N are described quotient spaces by the combinatorial pictures

$M = \Sigma(x_1 \dots x_m) =$ 

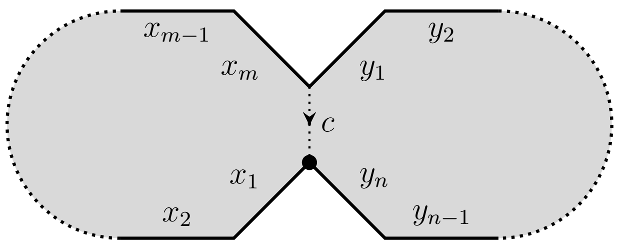
$N = \Sigma(y_1 \dots y_n) =$ 

Here the dot marks the distinguished vertex. Now we remove an open disk $\mathring{D}^2$ from M and N. In the pictures below, this is the disk enclosed by the curve labeled c. So after removing the open disk bounded by the curve c, this curve is the boundary of the resulting manifold with boundary.

$M \setminus \mathring{D}^2 =$ 

$N \setminus \mathring{D}^2 =$ 

Finally gluing these two spaces along the boundary circle c we obtain the connected sum $M\#N$, which looks as follows:

$M\#N =$ 

This shows that the connected sum $M\#N$ is homeomorphic to

$$\Sigma(x_1 \dots x_m y_1 \dots y_n) = \Sigma(W_1 W_2)$$

as claimed. $\square$

Corollary 2.24. *(1)* $\Sigma_g = \underbrace{T\# \dots \#T}_{g}$ *is homeomorphic to* $\Sigma(a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \dots a_g b_g a_g^{-1} b_g^{-1})$.

(2) $\underbrace{\mathbb{RP}^2\# \dots \#\mathbb{RP}^2}_{k}$ *is homeomorphic to* $\Sigma(a_1 a_1 a_2 a_2 \dots a_k a_k)$.

Proof. To prove part (1), we recall $T \approx \Sigma(aba^{-1}b^{-1})$. Then

$$\begin{aligned}\underbrace{T\# \dots \#T}_{g} &\approx \Sigma(a_1 b_1 a_1^{-1} b_1^{-1})\# \dots \#\Sigma(a_g b_g a_g^{-1} b_g^{-1}) \\ &\approx \Sigma(a_1 b_1 a_1^{-1} b_1^{-1} \dots a_g b_g a_g^{-1} b_g^{-1}),\end{aligned}$$

where the last homeomorphism follows from the proposition. Similarly, to prove part (2), we use that $\mathbb{RP}^2 \approx \Sigma(aa)$ and hence

$$\begin{aligned}\underbrace{\mathbb{RP}^2\# \dots \#\mathbb{RP}^2}_{k} &\approx \Sigma(a_1 a_1)\# \dots \#\Sigma(a_k a_k) \\ &\approx \Sigma(a_1 a_1 a_2 a_2 \dots a_k a_k)\end{aligned}$$

$\square$

Proposition 2.25. *Let W_1, W_2, W_3 be words, and let a be a letter which does not occur in these words. Then there are homeomorphisms*

$$\Sigma(W_1aW_2aW_3) \approx \Sigma(W_1aaW_2^{-1}W_3) \tag{2.26}$$
$$\Sigma(W_1aW_2aW_3) \approx \Sigma(W_1W_2^{-1}aaW_3) \tag{2.27}$$

Here W_2^{-1} is the inverse of the word $W_2 = x_1 \dots x_n$, given explicitly by $W_2^{-1} = x_n^{-1} \dots x_1^{-1}$ (as for products of elements of a group).

Proof. By part (2) of Lemma 2.19 there are homeomorphisms

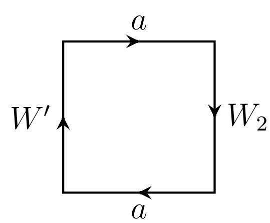
$$\Sigma(W_1aW_2aW_3) \approx \Sigma(aW_2aW') \qquad \text{and} \qquad \Sigma(W_1aaW_2^{-1}W_3) \approx \Sigma(aaW_2^{-1}W')$$

where $W' = W_3W_1$. Hence it suffice to produce homeomorphisms

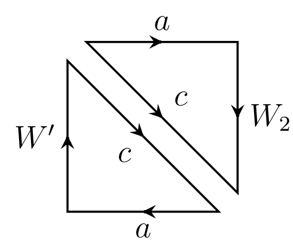
$$\Sigma(aW_2aW') \approx \Sigma(aaW_2^{-1}W') \tag{2.28}$$
$$\Sigma(aW_2aW') \approx \Sigma(W_2^{-1}aaW') \tag{2.29}$$

The homeomorphism (2.28) is given by the composition of the following homeomorphisms

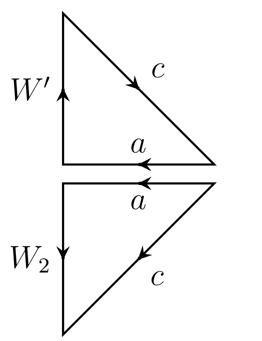
$\Sigma(aW_2aW') \approx$



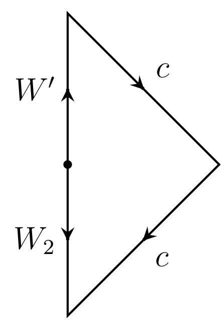
$\approx$



$\approx$



$\approx$



$$\approx \quad \Sigma(ccW_2^{-1}W') \quad \approx \quad \Sigma(aaW_2^{-1}W')$$

1. Here the first homeomorphism is an equality, by definition of the quotient space $\Sigma(aW_2aW')$ associated to the word aW_2aW';

2. The second homeomorphism arises by cutting the square along the diagonal. (Strictly speaking, this "square" is a polygon which may have many many more than four edges: the number of edges is the length of the word aW_2aW'. However, if we draw the edges corresponding to the words W' and W_2 vertically, and the two edges labeled a horizontally, then this polygon very much looks like a square, and so we prefer to use that terminology.) This results in two triangles (again, a slight abuse of language). We label the two new edges by the same label c (a new label distinct from all the other labels used so far) and the indicated direction. In Definition 2.18 we interpreted the labeling of the edges of one polygon as giving an equivalence relation on the polygon and hence an associated quotient space. Generalizing from one polygon with edge labeling to a disjoint union of polygons with edge labeling, we again interpret these pictures as giving us a quotient of the disjoint union of polygons by identifying all edges with the same label. Note that the order in which we glue the edges is irrelevant, and hence first gluing along the edge c gives back the previous quotient space.

3. The third homeomorphism is tautological, since the picture shows the same two polygons with the same edge labeling – we only moved the polygon drawn on the top right in the second picture to be below the other polygon (and we flipped it), so that the two edges labeled a in the two polygons are lined up.

4. The argument for the fourth homeomorphism is the same as for the second homeomorphism: first gluing along the edge labeled a, and then along the other edges gives the *same quotient* as identifying all edges with the same label simultaneously.

5. The fifth homeomorphism holds by definition of $\Sigma(ccW_2^{-1}W')$.

6. The sixth homeomorphism holds, since we may rename edges without changing the quotient they describe (see Lemma 2.19).

The homeomorphism (2.29) is constructed completely analogously by a sequence of pictures. The difference comes from using the *other diagonal* to cut the square in the first picture (going from the top right to the bottom left corner). $\square$

Proof of Lemma **??***(2).* The desired homeomorphism is given by the following composition of homeomorphism. The numbers below these homeomorphism indicate the reference to the appropriate Lemma/Proposition/Definition.

$$\begin{aligned}
T\#\mathbb{RP}^2 &\underset{(2.17)}{\approx} \Sigma(aba^{-1}b^{-1})\#\Sigma(cc) \underset{(2.25)}{\approx} \Sigma(aba^{-1}b^{-1}cc) \underset{(2.27)}{\approx} \Sigma(abcbac) \\
&\underset{(2.26)}{\approx} \Sigma(abbc^{-1}ac) \underset{(2.19)(2)}{\approx} \Sigma(bbc^{-1}aca) \underset{(2.27)}{\approx} \Sigma(bbc^{-1}c^{-1}aa) \\
&\underset{(2.25)}{\approx} \Sigma(bb)\#\Sigma(c^{-1}c^{-1})\#\Sigma(aa) \underset{(2.17)}{\approx} \mathbb{RP}^2\#\mathbb{RP}^2\#\mathbb{RP}^2
\end{aligned}$$

□

Outline of the constructive part of the Classification Theorem 2.8. Here by "constructive part" we mean the statement that every compact connected 2-manifold is homeomorphic to either Σ_g or X_k.

1. Show that every compact surface Σ admits a pattern of polygons Γ. Usually, this is stated as the stronger statement that every compact surface can be *triangulated*, meaning that it admits a pattern of triangles. Labeling all edges of Γ with a different letter and an arrow, and then cutting Σ along all edges gives a disjoint union of labeled polygons. By construction, Σ is the homeomorphic to the quotient space of this disjoint union by gluing along the pair of edges with the same label (see [Mu2, Thm. 78.1]).

2. The number of polygons involved can be reduced by one by gluing pairs of edges with the same label belonging to different polygons. Inductively, this shows that Σ can be obtained by edge identifications of *one* polygon (see [Mu2, Thm. 78.2]).

3. Use moves of the type described in Lemma 2.19 or Proposition 2.25 to show that the labeling of the edges of the polygon can be modified without changing the homeomorphism type of the quotient space to obtain the standard labeling for the surface of genus g or the k-fold projective space X_k (see [Mu2, Thm. 77.5]).

□

References

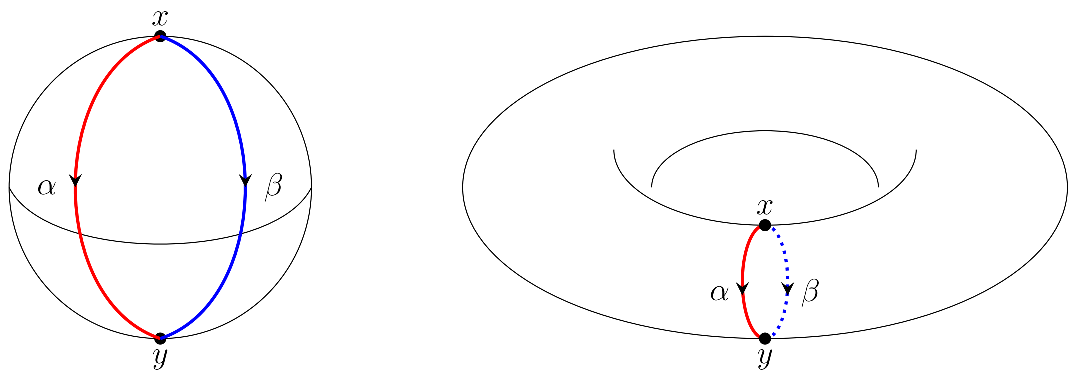
[Mu2] Munkres, James R. Topology. Second edition. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. xvi+537 pp.

3 The fundamental group

In this section we define the fundamental group of a topological space X. This is an invariant which can be used to distinguish topological spaces. For example, we will see that all the compact connected manifolds can be distinguished by their fundamental group. This is done using the Seifert van Kampen Theorem, a powerful tool to calculate fundamental groups.

3.1 The definition of the fundamental group

The basic idea of the fundamental group is that paths in different topological spaces might have different behavior. For example, the picture below shows two paths α, β with the same starting point and end point in the sphere and the torus. The difference between the two situations is that for the sphere the path α can be deformed to give the path β, while the paths on the torus cannot be deformed into one another.



Definition 3.1. A *path* in a topological space X is a continuous map $\gamma\colon [0,1] \to X$. The point $\gamma(0) \in X$ is the *starting point*, the point $\gamma(1) \in X$ is the *endpoint* of the path γ. With a slight abuse of language, both point $\gamma(1)$ and $\gamma(0)$ might be referred to as *endpoints* of the path γ. If $\gamma(0) = x$ and $\gamma(1) = y$, we say that γ is a path from x to y.

Let γ, δ be two paths in X from x to y. These paths are *homotopic relative endpoints* or *path homotopic* or simply *homotopic* if for every $t \in [0,1]$ there is a path γ_t from x to y such that

- $\gamma_0 = \gamma$ and $\gamma_1 = \delta$;
- The map $H\colon [0,1] \times [0,1] \to X$, $(s,t) \mapsto \gamma_t(s)$ is continuous. This condition expresses the idea that the family of paths γ_t depends continuously on the parameter t.

The map H is called a *homotopy from* γ *to* δ, and we write $\gamma \sim \delta$ to say that γ is homotopic to δ. It is easy to show that *homotopic* is an equivalence relation (we leave the proof to the reader). We use the notation $[\gamma]$ for the homotopy class of a path γ.

Let $U \subset \mathbb{R}^n$ be a *convex* subset, i.e., for any points $x, y \in U$ the straight line segment between x and y is contained in U. Explicitly, the straight line segment is the set

$$\{(1-t)x + ty \in \mathbb{R}^n \mid 0 \leq t \leq 1\}.$$

Examples of convex subspaces of $\mathbb{R}^n$:

- $\mathbb{R}^n$;

- an open ball $B_r(x)$ of radius r around some point $x \in \mathbb{R}^n$;
- a closed ball $D_r(x) := \{y \in \mathbb{R}^n \mid ||y - x|| \le r\}$ of radius r around some point $x \in \mathbb{R}^n$;

The punctured space $\mathbb{R}^n \setminus \{v\}$ is not convex, since for any nonzero $w \in \mathbb{R}^n$ the straight line segment between $x = v + w$ and $y = v - w$ contains the point v.

Lemma 3.2. *Let U be a convex subset of $\mathbb{R}^n$, and let α, β be paths in U with the same endpoints (i.e., $\alpha(0) = \beta(0)$ and $\alpha(1) = \beta(1)$). Then α and β are homotopic (relative endpoints). An explicit homotopy, called* linear homotopy *is given by the formula*

$$H\colon [0,1] \times [0,1] \longrightarrow U \qquad \textit{is given by} \qquad H(s,t) := (1-t)\gamma(s) + t\delta(s).$$

We note that for fixed $s \in [0,1]$ the path $t \mapsto H(s,t) = (1-t)\gamma(s) + t\delta(s)$ is the straight line path from $\gamma(s)$ to $\delta(s)$.

Definition 3.3. Let $\alpha, \beta\colon I \to X$, be paths in a topological space X. If $\alpha(1) = \beta(0)$, i.e., if the endpoint of α matches the starting point of β, then we can form a new path $\alpha * \beta$ called the *concatenation of α and β* by first following the path α and then following the path β. Explicitly, the path

$$\alpha * \beta\colon I \to X \qquad \text{is given by} \qquad (\alpha * \beta)(s) = \begin{cases} \alpha(2s) & 0 \le s \le \frac{1}{2} \\ \beta(2s-1) & \frac{1}{2} \le s \le 1 \end{cases}$$

It has starting point $\alpha(0)$ and endpoint $\beta(1)$.

Let α, β and γ be paths in X with $\alpha(1) = \beta(0)$ and $\beta(1) = \gamma(0)$, then we can form the concatenated paths $\alpha * \beta$ and $\beta * \gamma$. Since the endpoint of $\alpha * \beta$ is $\beta(1)$, it can further be concatenated with γ, forming the path $(\alpha * \beta) * \gamma$. Similarly, we can form the concatenation $\alpha * (\beta * \gamma)$. We want to point out that these paths are typically *not* equal to each other:

- $\alpha * (\beta * \gamma)(s)$ is a point on the path α for $0 \le s \le 1/2$, on the path β for $1/2 \le s \le 3/4$ and on the path γ for $3/4 \le s \le 1$, while
- $((\alpha * \beta) * \gamma)(s)$ on the path α for $0 \le s \le 1/4$, on the path β for $1/4 \le s \le 1/2$, and on the path γ for $1/2 \le s \le 1$.

Lemma 3.4. *Concatenation is associative up to homotopy, that is, if α, β, γ are paths in X with $\alpha(1) = \beta(0)$ and $\beta(1) = \gamma(0)$, then the paths*

$$\alpha * (\beta * \gamma) \qquad \textit{and} \qquad (\alpha * \beta) * \gamma \qquad\qquad \textit{are homotopic};$$

in other words, $[\alpha * (\beta * \gamma)] = [(\alpha * \beta) * \gamma]$.

The associativity of the concatenation of paths up to homotopy suggests that we might be able to construct a *group* associated to a topological space X by taking the elements of this group to be homotopy classes of paths in X. The problem with this is that paths can only be concatenated it the endpoint of the first path matches the starting point of the second path, while *any* two elements of a group can be multiplied with each other. There are two ways to deal with this issue:

- We pick a point $x_0 \in X$ and only consider paths that start and end at x_0; this is what we will do in the definition below of the fundamental group of a topological space X.
- We give up the idea of constructing a *group*, but instead construct a *groupoid* which is called the *fundamental groupoid* of X.

Definition 3.5. Let X be a topological space and let $x_0 \in X$ be a point of X, usually referred to as *base point*. Such a pair (X, x_0) is called a *pointed topological space*. A *based loop* in (X, x_0) is a path $\gamma\colon I \to X$ with $\gamma(0) = x_0 = \gamma(1)$. Let

$$\pi_1(X, x_0) := \{\text{based loops in } (X, x_0)\}/\text{homotopy}.$$

Proposition 3.6. *The set $\pi_1(X, x_0)$ is a group, the* fundamental group of (X, x_0), *with*

- *multiplication given by concatenation of based loops, i.e., $[\alpha] \cdot [\beta] := [\alpha * \beta]$ for based loops α, β;*
- *the identity element of $\pi_1(X, x_0)$ is given by the homotopy class of the constant path c_{x_0} (i.e., $c_{x_0}(s) = x_0$ for all $s \in I$);*
- *the inverse of an element $[\gamma] \in \pi_1(X, x_0)$ is given by $[\bar{\gamma}]$, where $\bar{\gamma}\colon I \to X$ is the path γ run backwards, i.e., $\bar{\gamma}(s) = \gamma(1 - s)$.*

The proof of this statement is pretty straightforward. The associativity of the product is a consequence of Lemma 3.4, which is a more general since it is a statement for composable paths rather than just based loops. Similarly, the claim that the homotopy class of constant path c_{x_0} is the identity element of $\pi_1(X, x_0)$ is a consequence of the first two homotopies of the following lemma, while the last two homotopies imply that $[\bar{\gamma}]$ is the inverse to $[\gamma]$ for a based loop in (X, x_0). Again, it will be useful for us to state these homotopies for paths, rather than just based loops.

Lemma 3.7. *Let $\gamma\colon I \to X$ be a path in a topological space X. Let $\bar{\gamma}\colon I \to X$ be the path defined by $\bar{\gamma}(s) := \gamma(1 - s)$ and let $c_x\colon I \to X$ be the constant path at a point $x \in X$. Then there are homotopies*

$$\gamma * c_{\gamma(1)} \sim \gamma \qquad c_{\gamma(0)} * \gamma \sim \gamma \qquad \gamma * \bar{\gamma} \sim c_{\gamma(0)} \qquad \bar{\gamma} * \gamma \sim c_{\gamma(1)}$$

Example 3.8. Let X be a convex subset of $\mathbb{R}^n$, and $x_0 \in X$. Then by Lemma 3.2 any based loop in (X, x_0) is homotopic to the constant loop c_{x_0}. Hence the fundamental group $\pi_1(X, x_0)$ is trivial.

Lemma 3.9. *Let X be a topological space and let β be a path from x_0 to x_1. Then the map*

$$\Phi_\beta \colon \pi_1(X, x_0) \longrightarrow \pi_1(X, x_1) \qquad [\gamma] \mapsto [\bar{\beta} * \gamma * \beta]$$

is an isomorphism of groups. In particular, the isomorphism class of the fundamental group $\pi(X, x_0)$ of a path connected space does not depend on the choice of the base point $x_0 \in X$.

3.2 Fundamental group of S^1 and the winding number

Theorem 3.10. *The fundamental group of the circle $\pi_1(S^1, 1)$ is isomorphic to $\mathbb{Z}$. The isomorphism $\Psi \colon \mathbb{Z} \to \pi_1(S^1, 1)$ is given by mapping $n \in \mathbb{Z}$ to the element $[\omega_n] \in \pi_1(S^1, 1)$ given by the based loop $\omega_n \colon I \to S^1$ defined by $\omega_n(s) = e^{2\pi i n s}$.*

Homework 3.11. Show that Ψ is a homomorpism.

We will proof this result in the next section, using the homotopy lifting property for the covering map $p \colon \mathbb{R} \to S^1$, $t \mapsto e^{2\pi i t}$ which we will prove when we talk about covering spaces.

Let $W \colon \pi_1(S^1, 1) \to \mathbb{Z}$ be the isomorphism inverse to Ψ. It associates to each based loop $\gamma \colon (I, \partial I) \to (S^1, 1)$ an integer $W(\gamma) \in \mathbb{Z}$ called the *winding number* of γ. Heuristically, it counts how often the loop γ "winds around" the circle. For example, by definition of W as the inverse of Ψ, the winding number $W(\omega_n)$ is equal to n, which matches the geometric intuition that the loop ω_n winds counterclockwise around the circle n times (for negative n this means that it winds around clockwise $|n|$ times). In the next section we will provide a much more direct definition of the winding number of based loops in S^1 3.13.

The goal of this section is to define a winding number $W(\gamma)$ for maps $\gamma \colon S^1 \to \mathbb{C} \setminus \{w\}$, i.e., for loops in the punctured plane $\mathbb{C} \setminus \{w\}$, and to use this to prove the fundamental theorem of algebra. A basic property of the winding number $W(\gamma)$ is that it only depends on the homotopy class of γ.

Definition 3.12. (Homotopy class of a map) Let X, Y be topological spaces and let $f, g \colon X \to Y$ be continuous maps. A *homotopy* from f to g is a continuous map

$$H \colon X \times I \longrightarrow Y$$

such that H restricts to f on $X \times \{0\}$ and to g on $X \times \{1\}$, i.e., $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in X$. If there is a homotopy from f to g, we say f *is homotopic to* g, and write $f \sim g$. It is not hard to show that this is an equivalence relation, we write $[f]$ for the homotopy class of f and $[X, Y]$ for the set of homotopy classes of maps $X \to Y$.

Let $w \in \mathbb{C}$ be a fixed element of the complex plane. Then there is a map

$$f\colon \mathbb{C} \setminus \{w\} \longrightarrow S^1 \qquad \text{defined by} \qquad f(z) := \frac{z-w}{|z-w|}.$$

If $\gamma\colon S^1 \to \mathbb{C} \setminus \{w\}$ is a loop in the punctured plane $\mathbb{C} \setminus \{w\}$, then $f \circ \gamma$ is a loop in S^1. In general, $f(\gamma(1)) \neq 1 \in S^1$, i.e., $f \circ \gamma$ is not a loop based at $1 \in S^1$, and hence does not represent an element of the fundamental group $\pi_1(S^1, 1)$. This can be fixed by multiplying the loop $f \circ \gamma$ by $f(\gamma(1))^{-1} \in S^1$. More precisely, let

$$\Phi_\gamma\colon S^1 \to S^1 \qquad \text{be the based loop given by} \qquad \Phi_\gamma(z) = f(\gamma(z)) \cdot f(\gamma(1))^{-1}.$$

It is easy to check that a homotopy H from γ to γ' induces a homotopy Φ_H from Φ_γ to $\Phi_{\gamma'}$. Explicitly,

$$\Phi_H\colon S^1 \times I \to S^1 \qquad \text{is given by} \qquad \Phi_H(z,t) = f(H(z,t)) \cdot f(H(1,t))^{-1}.$$

This shows that there is a well-defined map

$$\Phi\colon [S^1, \mathbb{C} \setminus \{w\}] \longrightarrow \pi_1(S^1, 1) \qquad \text{given by} \qquad \Phi([\gamma]) = [\Phi_\gamma].$$

Definition 3.13. Let $w \in \mathbb{C}$, and let $\gamma\colon S^1 \to \mathbb{C} \setminus \{w\}$ be a loop. Then $W(\gamma, w) \in \mathbb{Z}$, the *winding number of* γ *around the point* w is the winding number $W(\Phi_\gamma)$ of the associated based loop in S^1. The discussion above shows that the winding number $W(\gamma, w)$ depends only on the homotopy class of γ.

Example 3.14. For $n \in \mathbb{Z}$ the map $\omega_n\colon S^1 \to \mathbb{C} \setminus \{0\}, z \mapsto z^n$ has winding number $W(\gamma, 0) = n$. To see this, we observe that ω_n is already a based loop in S^1 and hence $\Phi_{\omega_n} = \omega_n$ (as maps from S^1 to S^1). Moreover, via the standard homeomorphism $I/\partial I \cong S^1$ given by $t \mapsto e^{2\pi i t}$, the based loop $\omega_n\colon S^1 \to S^1$ corresponds to the based loop $\omega_n\colon I \to S^1$, $\omega_n(s) = e^{2\pi i n s}$ mentioned in Theorem 3.10, which has winding number n.

Theorem 3.15. The Fundamental Theorem of Algebra. *Let*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$$

be a polynomial of degree $n > 0$. *Then* p *has a zero, that is, there is some* $z \in \mathbb{C}$ *such that* $p(z) = 0$.

Proof. Aiming for a proof by contradiction, we assume that $p(z)$ belongs to $\mathbb{C}^\times := \mathbb{C} \setminus \{0\}$ for all $z \in \mathbb{C}$. This allows us to talk about the winding number $W(\gamma, 0) \in \mathbb{Z}$ of the closed curve

$$\gamma\colon S^1 \longrightarrow \mathbb{C}^\times \qquad \text{given by} \qquad z \mapsto p(z)$$

around $0 \in \mathbb{C}$. Since all the closed loops considered in this proof are in $\mathbb{C}^\times$, we will just write $W(\gamma)$ instead of $W(\gamma, 0)$. We will calculate $W(\gamma)$ in two different ways, the first one resulting in $W(\gamma) = 0$, the other one resulting in $W(\gamma) = n$. This is the desired contradiction.

The loop γ is obtained by restricting the polynomial $p(z)$ to the unit circle. Restricting $p(z)$ instead to the circle of radius r, we obtain the loop

$$\gamma_r \colon S^1 \longrightarrow \mathbb{C}^\times \qquad \text{defined by} \qquad z \mapsto p(rz).$$

We note that the loop γ_r is homotopic to $\gamma = \gamma_1$. A homotopy is given by

$$H \colon S^1 \times I \longrightarrow \mathbb{C}^\times \qquad \text{given by} \qquad H(z,t) = p((tr + (t-1))z).$$

This implies that $W(\gamma) = W(\gamma_r)$ for all r, including $r = 0$.

First calculation. The loop γ_0 is the constant loop, and hence $W(\gamma_0, 0) = 0$. It follows that $W(\gamma) = 0$.

Second calculation. Instead of shrinking r to 0, we will now consider the loop γ_r for large radius r. Writing $\gamma_r(z)$ in the form

$$\gamma_r(z) = a_n r^n z^n + a_{n-1} r^{n-1} z^{n-1} + \cdots + a_0 = r^n \left(a_n z^n + \frac{a_{n-1}}{r} z^n + \cdots + \frac{a_0}{r^n} \right),$$

we see that the term $b_r(z) := \frac{a_{n-1}}{r} z^n + \cdots + \frac{a_0}{r^n}$ converges to 0 uniformly for $z \in S^1$. In particular, for sufficiently large r we have

$$|b_r(z)| < |a_n z^n| \qquad \text{for all } z \in S^1.$$

It follows that $a_n z^n + t b_r(z)$ belongs to $\mathbb{C}^\times$ for all $z \in S^1$ and $t \in [0,1]$. Then

$$H \colon S^1 \times I \longrightarrow \mathbb{C}^\times \qquad \text{defined by} \qquad H(z,t) := r^n \left(a_n z^n + t b_r(z) \right)$$

is a homotopy between $\gamma_r(z)$ and the loop $\gamma'(z) := r^n a_n z^n$. The loop γ' in turn is homotopic to the loop γ_n, $\gamma_n(z) := z^n$ of Example 3.14. A homotopy $H' \colon S^1 \times I \to \mathbb{C}^\times$ is given by choosing a path δ in $\mathbb{C}^\times$ with $\delta(0) = r^n a_n$, $\delta = 1$, and defining

$$H'(z,t) := \delta(t) z^n.$$

Since homotopic loops have the same winding number, it follows that

$$W(\gamma_r) = W(\gamma') = W(\omega_n) = n.$$

$\square$

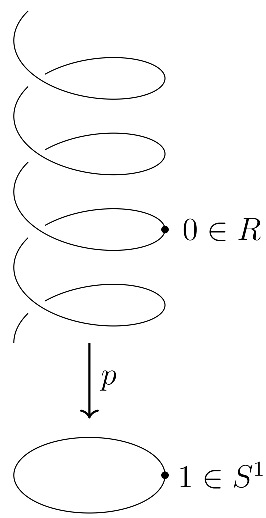
3.3 The covering map $p\colon \mathbb{R} \to S^1$

The goal of this section is to prove Theorem 3.10, i.e., to show that the fundamental group $\pi_1(S^1, 1)$ is isomorphic to $\mathbb{Z}$. It relies on a direct definition of the winding number $W(\gamma) \in \mathbb{Z}$ of a based loop $\gamma\colon (I, \partial I) \to (S^1, 1)$.

We first give a heuristic discussion of the definition of $W(\gamma)$ before spelling out the results on covering spaces it is based upon. Consider the surjective map

$$p\colon \mathbb{R} \longrightarrow S^1 \qquad \text{given by} \qquad p(t) = e^{2\pi i t}.$$

We picture this map as the vertical projection of a vertical spiral onto a circle in the xy-plane as in the picture



As indicated in the picture, we choose $0 \in \mathbb{R}$ as the base point of the real line and $p(0) = 1 \in S^1$ as the base point of the circle. Let $\gamma\colon (I, \partial I) \to (S^1, 1)$ be a based loop, i.e., a path $\gamma\colon I \to S^1$ which starts and ends at $1 \in S^1$.

As the picture suggests, let us think $\mathbb{R}$ as a circular stairway located above the circular track represented by S^1. Let us interpret $\gamma(s) \in S^1$ as the location at time $s \in [0, 1]$ of person taking a leisurely walk along the circular track which starts and ends an hour later at $1 \in S^1$. Suppose that this person is the owner of a dog who at the same time runs around on the spiral staircase above and that $\widetilde{\gamma}(s) \in \mathbb{R}$ is the dog's location at time $s \in [0, 1]$. Suppose that the dog starts his outing on the staircase at $0 \in \mathbb{R}$, and that this faithful dog tries to be as close as it can be to its owner, in other words, always vertically above its owner.

Question. At which location is the dog at the end of the outing?

Since the dog is always vertically above its owner, i.e., $p(\widetilde{\gamma}(s)) = \gamma(s)$, the dog's ending location $\widetilde{\gamma}(1)$ must belong to $p^{-1}(\gamma(1)) = p^{-1}(1) = \mathbb{Z} \subset \mathbb{R}$. We observe that the dog's ending location does depend on the path γ its owner takes. E.g., if the owner walks counterclockwise once around the track, the dog ends up at $1 \in \mathbb{R}$. If the owner walks around twice, the dog ends up at $2 \in \mathbb{R}$. With a clockwise walk once around the track of the owner, the dog ends up at $-1 \in \mathbb{R}$. In other words, the ending location of the dog is the winding number $W(\gamma)$ of the owner's loop γ around the track. For this reason, the invariant $W(\gamma)$ could also be referred to as the "faithful dog invariant".

Intuitively, it is obvious that the path γ the owner takes uniquely determines the path $\widetilde{\gamma}$ of the dog. Mathematically, the path $\widetilde{\gamma}$ with the property $p \circ \widetilde{\gamma} = \gamma$ is called a "lift" of γ, and its existence has to be proved. Also, it is intuitively clear that the lift $\widetilde{\gamma}$ is unique, once we choose a particular starting point $\widetilde{\gamma}(0) \in p^{-1}(\gamma(0))$. Again that needs to be proved.

It turns out that this "path-lifting" does not just work for the specific map $p\colon \mathbb{R} \to S^1$, but for any *covering map* (see definition below). Later this semester we will make extensive use of path-lifting for covering spaces, and for that reason we will define covering spaces now and state their path lifting properties, but will refer the proof of their path-lifting properties to the section on covering spaces.

Definition 3.16. A continuous map $p\colon \widetilde{X} \to X$ is a *covering map* if p is surjective, and if for each $x \in X$ there is an open neighborhood U with the property that

- $p^{-1}(U)$ is the disjoint union of open subsets $U_a \subset \widetilde{X}$, $a \in A$, and
- for every $i \in I$ the restriction $p_{|U_a}\colon U_a \to U$ is a homeomorphism.

Any open subset $U \subset X$ with this property is called *evenly covered.* If $p\colon \widetilde{X} \to X$ is a covering map, the space $\widetilde{X}$ is called a *covering space* of X.

Example 3.17. The map $p\colon \mathbb{R} \to S^1$, $p(t) = e^{2\pi i t}$ is a covering map, since for any point $e^{2\pi i t} \in S^1$, an open neighborhood of U of $e^{2\pi i t}$ is given by the image of the interval $(t-\epsilon, t+\epsilon)$ under p (this is an open map, as is easy to check in the same way we verified that $S^n \to \mathbb{RP}^n$ is an open map). Then

$$p^{-1}(U) = \bigcup_{n \in \mathbb{Z}} (n + t - \epsilon, n + t + \epsilon),$$

and these are *disjoint* open intervals for ϵ small enough (any $\epsilon < 1/2$ works). It is easy to show that p restricted to each interval $(n + t - \epsilon, n + t + \epsilon)$ is a homeomorphism to U. In other words, U is evenly covered and hence $p\colon \mathbb{R} \to S^1$ is a covering map.

Lemma 3.18. (Path-lifting for covering maps) *Let $p\colon \widetilde{X} \to X$ be a covering map. Let $\gamma\colon I \to X$ be path, and let $\widetilde{x}_0 \in \widetilde{X}$ with $p(\widetilde{x}_0) = \gamma(0)$. Then there is a unique path $\widetilde{\gamma}\colon I \to \widetilde{X}$ with $\widetilde{\gamma}(0) = \widetilde{x}_0$ which is a* lift *of γ, i.e., $p \circ \widetilde{\gamma} = \gamma$.*

We remark that this statement can be formulated in the form of a commutative diagram:

$$\begin{tikzcd}
\{0\} \arrow[r, "\widetilde{x}_0"] \arrow[d, hook] & \widetilde{X} \arrow[d, "p"] \\
I \arrow[r, "\gamma"] \arrow[ur, dashed, "\exists!\ \widetilde{\gamma}"] & X
\end{tikzcd}$$

- The solid arrows of the commutative outer square represent the *assumptions* of the lemma: the data given in the lemma consists of the maps p, γ, the inclusion map $\{0\} \hookrightarrow I$, and the point $\widetilde{x}_0 \in \widetilde{X}$, which can be interpreted as a map from the one-point space $\{0\}$ to $\widetilde{X}$. The assumption $p(\widetilde{x}_0) = \gamma(0)$ of the lemma is equivalent to commutativity of the square.

- The dashed arrow represents the *statement* of the lemma, the existence of the map $\widetilde{\gamma}$ and its uniqueness. Note that commutativitity of the upper triangle expresses the requirement $\widetilde{\gamma}(0) = \widetilde{x}_0$, while the commutativity of the lower triangle expresses the requirement that $\widetilde{\gamma}$ is a lift of γ.

Definition 3.19. Let $\gamma\colon (I, \partial I) \to (S^1, 1)$ be based loop, and let $\widetilde{\gamma}\colon I \to \mathbb{R}$ the unique lift of γ with $\widetilde{\gamma}(0) = 0$. Then $W(\gamma) := \widetilde{\gamma}(1) \in \mathbb{Z}$ is the *winding number of* γ.

Lemma 3.20. *The winding number $W(\gamma)$ of a based loop $\gamma\colon (I, \partial I) \to (S^1, 1)$ depends only on the homotopy class $[\gamma] \in \pi_1(S^1, 1)$.*

The proof of this result uses the following generalization of the path-lifting property of covering maps.

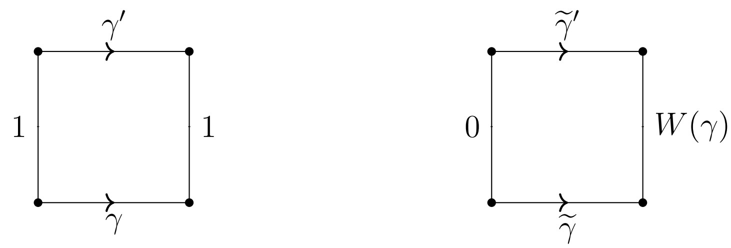
Proposition 3.21. *Let $p\colon \widetilde{X} \to X$ be a covering map, let $H\colon Y \times I \to X$ be a homotopy from $f\colon Y \to X$ to $g\colon Y \to X$, and let $\widetilde{f}\colon Y \to \widetilde{X}$ be a lift of f, i.e., $p \circ \widetilde{f} = f$. Then there is a unique homotopy $\widetilde{H}\colon Y \times I \to \widetilde{X}$ which is a lift of H (i.e., $p \circ \widetilde{H} = H$) whose restriction to $Y \times \{0\}$ is the map $\widetilde{f}$.*

As the path-lifting property this homotopy lifting property of covering maps can be expressed as a commutative diagram:

$$\begin{tikzcd}
Y \times \{0\} \arrow[r, "\widetilde{f}"] \arrow[d, hook] & \widetilde{X} \arrow[d, "p"] \\
Y \times I \arrow[r, "H"] \arrow[ur, dashed, "\exists!\ \widetilde{H}"] & X
\end{tikzcd}$$

Proof of Lemma 3.20. Let γ, γ' be based loops in S^1 for the basepoint $x_0 := 1 \in S^1$, and let $\widetilde{\gamma}\colon I \to \mathbb{R}$ be a lift of γ with $\widetilde{\gamma}(0) = \widetilde{x}_0 := 0 \in \mathbb{R}$. Let $H\colon I \times I \to S^1$ be a path homotopy from γ to γ', and let $\widetilde{H}\colon I \times I \to \mathbb{R}$ be the unique lift of H whose restriction to $I \times \{0\}$ is $\widetilde{\gamma}$.

The following pictures show what we know or can deduce about the maps H, $\widetilde{H}$. The pictures show the square $I \times I$, the domain of H and $\widetilde{H}$, and indicate where the four boundary edges map to by labeling them by the corresponding path in S^1 resp. $\mathbb{R}$.



$H\colon I \times I \to S^1$ $\qquad\qquad$ $\widetilde{H}\colon I \times I \to \mathbb{R}$

The left picture shows the path homotopy H from γ to γ', so H restricted to the bottom (resp. top) edge is the path γ (resp. γ'). Since H is a path homotopy, the endpoints stay fixed during the homotopy, i.e., H maps both vertical edges to the basepoint $1 \in S^1$.

The homotopy $\widetilde{H}$ is a lift of H, which restricts to $\widetilde{\gamma}$ on its bottom edge. In particular, it maps the left bottom vertex to $\widetilde{\gamma}(0) = 0$ and the right bottom vertex to $\widetilde{\gamma}(1) = W(\gamma) \in \mathbb{Z}$ (by definition of the winding number, Def. 3.13). Unlike H, the lift $\widetilde{H}$ is a priori not a path homotopy, but only a homotopy, i.e., there is no built-in requirement that $\widetilde{H}$ keeps the endpoints fixed (note that we obtain the lifted homotopy by appealing to Proposition **??**, which deals with homotopies for maps $Y \to X$, where it doesn't make sense to talk about "endpoints").

Fortunately, we can *prove* that the lifted homotopy $\widetilde{H}$ is in fact a path homotopy: restricting $\widetilde{H}$ to the vertical edges gives paths in $\mathbb{R}$ which are lifts of *constant paths* in S^1. Hence by uniqueness of path lifting (Lemma **??**) the restriction of $\widetilde{H}$ to the vertical edges must be constant paths in $\mathbb{R}$, and hence $\widetilde{H}$ is in fact a path homotopy. The left edge must map to 0, since the bottom left vertex maps to $\widetilde{\gamma}(0) = 0$, while the right edge must map to $W(\gamma)$ since the bottom right edge maps to $\widetilde{\gamma}(1) = W(\gamma)$.

The path $\widetilde{\gamma}'$ obtained by restricting $\widetilde{H}$ to the top edge is a lift of γ' with starting point 0 and endpoint $W(\gamma)$. Since $\widetilde{\gamma}'$ has starting point 0 and is a lift of γ', its endpoint $\widetilde{\gamma}'(1)$ is definition the winding number of γ'. Hence $W(\gamma') = W(\gamma)$ as claimed. $\square$

Theorem 3.22. *The homomorphism $W\colon \pi_1(S^1, 1) \to \mathbb{Z}$ that maps $[\gamma] \in \pi_1(S^1, 1)$ to the winding number $W(\gamma)$ is an isomorphism. It is inverse to the map Φ of Theorem 3.10.*

Proof. To show that W is a group homomorphism, let γ, γ' be based loops in $(S^1, 1)$, and let $\widetilde{\gamma}, \widetilde{\gamma}'\colon I \to \mathbb{R}$ be their lifts for the covering map $p\colon \mathbb{R} \to S^1$, $t \mapsto e^{2\pi i t}$ with starting point $0 \in \mathbb{R}$. We need to determine the lift of the concatenated loop $\gamma * \gamma'$. It is tempting to say that that is $\widetilde{\gamma} * \widetilde{\gamma}'$, but this concatenation not be defined since the endpoint of the first path, i.e., $\widetilde{\gamma}(1) = W(\gamma)$, in general is not equal to the starting point of the second path $\widetilde{\gamma}'(0) = 0$. We can fix this problem by replacing $\widetilde{\gamma}'$ by the path $\widehat{\gamma}'\colon I \to \mathbb{R}$ defined by

$$\widehat{\gamma}'(s) := W(\gamma) + \widetilde{\gamma}'(s).$$

The path $\widehat{\gamma}'$ also is a lift of γ', but its starting point is $\widehat{\gamma}'(0) = W(\gamma)$, and hence the concatenation $\widetilde{\gamma} * \widehat{\gamma}'$ is a continuous path which is a lift of $\gamma * \gamma'$ with starting point 0 and endpoint

$$(\widetilde{\gamma} * \widehat{\gamma}')(1) = \widehat{\gamma}'(1) = W(\gamma) + \widetilde{\gamma}'(1) = W(\gamma) + W(\gamma').$$

This shows that the winding number of $\gamma * \gamma'$ is $W(\gamma) + W(\gamma')$.

To show that W is surjective, let $n \in \mathbb{Z}$, and let $\omega_n \colon I \to S^1$ be the based loop defined by $\omega_n(s) = e^{2\pi ins}$. Then $\widetilde{\omega} \colon I \to \mathbb{R}$ defined by $\widetilde{\omega}(s) = ns$ is a lift of ω for the covering map $p \colon \mathbb{R} \to S^1$, $t \mapsto e^{2\pi it}$ (i.e., $p \circ \widetilde{\omega}_n = \omega$). Moreover, $\widetilde{\omega}_n(0) = 0$ and hence by the definition of the winding number $W(\omega) = \widetilde{\omega}(1) = n$.

Once we verify that W is injective, this also shows that W is an inverse to Φ which was defined by mapping $n \in \mathbb{Z}$ to $[\omega_n] \in \pi_1(S^1, 1)$.

To show that W is injective, let γ, γ' be based loops in $(S^1, 1)$ and assume that $W(\gamma) = W(\gamma')$. Let $\widetilde{\gamma}, \widetilde{\gamma}' \colon I \to \mathbb{R}$ be lifts of γ resp. γ' with starting point $0 \in \mathbb{R}$. Then $\widetilde{\gamma}(1) = W(\gamma)$, $\widetilde{\gamma}'(1) = W(\gamma')$ and hence the starting points and end points of these two paths in $\mathbb{R}$ agree. Let $\widetilde{H} \colon I \times I \to \mathbb{R}$ be the linear homotopy between these two paths, i.e.,

$$H(s,t) = (1-t)\widetilde{\gamma}(s) + t\widetilde{\gamma}'(s).$$

This is a path homotopy, i.e., $H(0,t)$ and $H(1,t)$ are independent of t. Then $p \circ H$ a path homotopy between $p \circ \widetilde{\gamma} = \gamma$ and $p \circ \widetilde{\gamma}' = \gamma'$, showing that $[\gamma] = [\gamma'] \in \pi_1(S^1, 1)$. □

4 A little bit of category theory

So far, we've calculated the fundamental groups for very few spaces: for convex subspaces of $\mathbb{R}^n$, for the circle S^1 and products of the circle. The main technique for calculating the fundamental group of more complicated spaces X is to write X as a union of open subspaces X_1 and X_2 such that the fundamental groups of X_1, X_2 and the intersection $X_1 \cap X_2$ are already known (and $X_1 \cap X_2$ is path-connected). The Seifert van Kampen Theorem then gives a formula for the fundamental of X in terms of the fundamental groups of X_1, X_2, $X_1 \cap X_2$ and the group homomorphisms

$$\begin{tikzcd} \pi_1(X_1 \cap X_2) \arrow[r, "(j_1)_*"] \arrow[d, "(j_2)_*"'] & \pi_1(X_1) \\ \pi_1(X_2) & \end{tikzcd}$$

induced by the inclusion maps of $X_1 \cap X_2$ into X_1 resp. X_2. More precisely, according to the Seifert van Kampen Theorem, the fundamental group $\pi_1(X, x_0)$ is given by the *pushout* of the above diagram in the category of groups. Since the space X is the pushout of the corresponding diagram

$$\begin{array}{ccc} X_1 \cap X_2 & \xrightarrow{j_1} & X_1 \\ {\scriptstyle j_2}\downarrow & & \\ X_2 & & \end{array}$$

in the category of topological spaces, the Seifert van Kampen Theorem can be stated conceptually by saying that "the fundamental group functor preserves pushouts".

The goal of this chapter is to define "pushouts" and to calculate pushouts in the category of topological space and the category of groups. In the first section we introduce the language of categories and functors. In the second section, we first the more basic notions 'categorical product" and "categorical coproduct" before defining "pushouts".

4.1 Functors and categories

In the previous sections we have discussed how we can associate to any pointed topological space (X, x_0) a group $\pi_1(X, x_0)$ (the fundamental group, Definition 3.5) and how to associate to a base point preserving map $f\colon (X, x_0) \to (Y, y_0)$ between pointed topological spaces a group homomorphism $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$ (the induced homomorphism on fundamental groups, Definition **??**). In other words, this construction associates to one kind of mathematical object (a pointed topological space) a different kind of mathematical object (a group), and to appropriate maps between the first kind of objects (basepoint preserving continuous maps) appropriate maps between the second kind of objects (group homomorphisms).

Such a construction is called a *functor between categories*. The goal of this section is to provide a quick introduction to categories and functors. Even if you haven't seen the formal definition of a category, it is likely that you already know many examples of categories. So it seems appropriate to mention some mathematical objects and appropriate maps between them that will then motivate the definition of a category.

When studying various mathematical objects, we usually also talk about the appropriate kind of maps between these objects. The following table lists some well known examples.

mathematical objects	appropriate maps
sets	maps
groups	group homomorphisms
vector spaces	linear maps
topological spaces	continuous maps

What is the structure that is common to all of these four types of mathematical objects and the maps between them? There isn't too much there, but we observe that composing "appropriate maps" leads again to "appropriate maps" (assuming the domain/source of one map matches the codomain/target of the other map), and that there is an "identity map" for every object. The following definition captures this structure, which is called a *category*. The four kinds of mathematical objects and the maps between them are then examples of categories.

Remark 4.1. Let X, Y, Z be sets and let $g\colon X \to Y$ and $f\colon Y \to Z$ be maps. Then there are *two* usual ways to write the composition, namely as

$$g \circ f \qquad \text{or as} \qquad X \xrightarrow{f} Y \xrightarrow{g} Z$$

Using the first way to write compositions, it is natural to think of composition as the map given by $(g, f) \mapsto g \circ f$. Writing $\mathrm{Maps}(X, Y)$ for the set of maps from X to Y, this is the map

$$\begin{array}{rcl} \mathrm{Maps}(Y, Z) \times \mathrm{Maps}(X, Y) & \longrightarrow & \mathrm{Maps}(X, Z) \\ (g, f) & \mapsto & g \circ f \end{array}$$

However, thinking about it the second way, it is more natural to think of composition as the map

$$\begin{array}{rcl} \mathrm{Maps}(X, Y) \times \mathrm{Maps}(Y, Z) & \longrightarrow & \mathrm{Maps}(X, Z) \\ (X \xrightarrow{f} Y, Y \xrightarrow{g} Z) & \mapsto & X \xrightarrow{f} Y \xrightarrow{g} Z \end{array}$$

Both ways have their advantages and disadvantages; to me, the second one seems more elegant, but alas, the first way is probably too deeply entrenched in mathematics to be thrown out. The sad effect is that there is there is no general consensus of how to write compositions in categories. I will follow the first convention.

Definition 4.2. A *category* $\mathcal{C}$ consists of the following data:

- A class of objects, denoted $\mathrm{ob}(\mathcal{C})$; the elements of $\mathrm{ob}(\mathcal{C})$ are called the *objects* of the category $\mathcal{C}$.
- For each pair of objects $X, Y \in \mathrm{ob}(\mathcal{C})$ a set $\mathrm{mor}_{\mathcal{C}}(X, Y)$. The elements of $\mathrm{mor}_{\mathcal{C}}(X, Y)$ are called *morphisms in $\mathcal{C}$ from X to Y* or morphisms with domain (or source) X and codomain (or target) Y. Alternative notations for this set include $\mathrm{mor}(X, Y)$ or $\mathcal{C}(X, Y)$.
- For objects $X, Y, Z \in \mathcal{C}$ there is a composition map

$$\circ\colon \mathcal{C}(Y, Z) \times \mathcal{C}(X, Y) \longrightarrow \mathcal{C}(X, Z) \qquad (g, f) \mapsto g \circ f.$$

- For each object $X \in \mathrm{ob}(\mathcal{C})$ a morphism $\mathrm{id}_X \in \mathcal{C}(X, X)$ called *identity morphism.*

These data are subject to the following requirements:

associativity For morphisms $f \in \mathcal{C}(U, X)$, $g \in \mathcal{C}(X, Y)$, $h \in \mathcal{C}(Y, Z)$ we have

$$(h \circ g) \circ f = h \circ (g \circ f) \in \mathcal{C}(U, Z).$$

identity property For $f \in \mathcal{C}(X, Y)$ we have $f \circ \mathrm{id}_X = f = \mathrm{id}_Y \circ f \ \in \mathcal{C}(X, Y)$.

Remark 4.3. For a morphism $f \in \mathcal{C}(X, Y)$ we often write $X \xrightarrow{f} Y$ to indicate the domain and codomain of f. For $f \in \mathcal{C}(Y, Z)$ and $g \in \mathcal{C}(X, Y)$ we often write $X \xrightarrow{g} Y \xrightarrow{f} Z$ for the composition $f \circ g$.

Definition 4.4. Let $\mathcal{C}$ be a category. A morphism $f \in \mathcal{C}(X, Y)$ is called an *isomorphism* if there exists a morphism $g \in \mathcal{C}(Y, X)$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. The category $\mathcal{C}$ is a *groupoid* if every morphism is an isomorphism.

We can now recast our motivating examples of sets, groups, vector spaces and topological spaces as categories.

Example 4.5.

category $\mathcal{C}$	objects	morphisms	isomorphisms
Set	sets	maps	bijections
Grp	groups	group homomorphisms	group isomorphisms
Vect	vector spaces	linear maps	linear isomorphisms
Top	topological spaces	continuous maps	homeomorphisms

Our previous examples of category might suggest that morphisms are always maps of sets compatible with additional structure these sets might have. In the following examples of categories, this is not the case.

Example 4.6. (Examples of categories whose morphisms are not maps between sets).

- To any group G we can associate a category $\mathcal{C}$ as follows. The category $\mathcal{C}$ has one object denoted $*$, and the set of morphisms $\mathcal{C}(*, *)$ from $*$ to $*$ is the set G of group elements. The composition map

  $$\circ \colon \mathcal{C}(*, *) \times \mathcal{C}(*, *) \longrightarrow \mathcal{C}(*, *)$$

  is given by the map $m \colon G \times G \to G$ that describes multiplication of elements of G. The identity morphism id_* is defined to be the identity element $1 \in G$ of the group

G. Associativity and the identity property hold for the category $\mathcal{C}$, since the group multiplication is associative and $1 \in G$ is the identity element of the group G.

We note that every morphism $g \in \mathcal{C}(*, *)$ is an isomorphism (its inverse is given by the group element g^{-1}), and hence $\mathcal{C}$ is a groupoid.

- To any topological space X we can associate a groupoid $\Pi_1(X)$, called the *fundamental groupoid of* X. As the name suggests, this is a generalization of the fundamental group of X. The objects of $\Pi_1(X)$ are the points of X. For x, y, the set of morphisms $\mathrm{mor}(x, y)$ is defined to be

$$\mathrm{mor}(x, y) := \{\text{paths } \gamma\colon I \to X \text{ with } \gamma(1) = x,\ \gamma(0) = y\}/\text{homotopy}.$$

For $x, y, z \in X$, the composition in this category is induced by concatenation of paths:

$$\circ\colon \mathrm{mor}(y, z) \times \mathrm{mor}(x, y) \longrightarrow \mathrm{mor}(x, z) \qquad \text{is given by} \qquad ([\alpha], [\beta]) \mapsto [\alpha * \beta]$$

We recall that for any pointed topological space (X, x_0) we have defined its fundamental group $\pi_1(X, x_0)$, and for any pointed map $f\colon (X, x_0) \to (Y, y_0)$ there is the group homomorphism

$$f_*\colon \pi_1(X, x_0) \longrightarrow \pi_1(Y, y_0) \qquad \text{defined by} \qquad [\gamma] \mapsto [f \circ \gamma]$$

for a based loop $\gamma\colon (I, \partial I) \to (X, x_0)$. This is compatible with compositions in the sense that for pointed maps $(X, x_0) \xrightarrow{f} (Y, y_0) \xrightarrow{g} (Z, z_0)$

$$(g \circ f)_* = g_* \circ f_*$$

it is also compatible with identities, i.e., if id_X is the identity map of a pointed space (X, x_0), the induced map $(\mathrm{id}_X)_*$ is the identity automorphism on the fundamental group $\pi_1(X, x_0)$.

This is an example of a *functor*, in this case a functor from the category Top_* of pointed topological spaces and basepoint preserving maps to the category Grp of groups. This is our motivating example for the definition of functor.

Definition 4.7. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A *functor* $F\colon \mathcal{C} \to \mathcal{D}$

- associates to every object X of $\mathcal{C}$ an object $F(X)$ of $\mathcal{D}$;
- associates to every morphism $f \in \mathcal{C}(X, Y)$ a morphism $F(f) \in \mathcal{D}(F(X), F(Y))$,

subject to the following compatibility requirements with:

compositions: $F(g \circ f) = F(g) \circ F(f)$ for $f \in \mathcal{C}(X, Y)$ and $g \in \mathcal{C}(Y, Z)$;

identities: $F(\mathrm{id}_X) = id_{F(X)}$ for any object X of $\mathcal{C}$.

Example 4.8. (Forgetful functors) Frequently, the objects of a category $\mathcal{C}$ consist of sets with suitable additional structure (i.e., a group structure, the structure of vector space over a field K, a topological space), and morphisms that consists of those maps of the underlying sets that are compatible with the additional structure. Then we obtain a *forgetful functor* $F\colon \mathcal{C} \to \mathsf{Set}$ by mapping each object of $\mathcal{C}$ to its underlying set, and mapping each morphism to the underlying map of sets. This way we obtain forgetful functors

1. $\mathsf{Grp} \longrightarrow \mathsf{Set}$ (by forgetting the group structure);
2. $\mathsf{Vect}_K \longrightarrow \mathsf{Set}$ (by forgetting the vector space structure);
3. $\mathsf{Top} \longrightarrow \mathsf{Set}$ (by forgetting the topology);

Other forgetful functors forget just part of the structure. For example, there are forgetful functors

4. $\mathsf{Vect}_K \longrightarrow \mathsf{Grp}$ (by forgetting the scalar multiplication in a K-vector space V, but remembering the abelian group structure);
5. $\mathsf{Top}_* \longrightarrow \mathsf{Top}$ (by forgetting the basepoint).

Example 4.9. (Free functors)

(1) The functor $\mathsf{Set} \to \mathsf{Vect}_K$ that sends a set S to the K-vector space

$$KS := \{\text{(formal) finite linear combinations } \sum_{s\in S} k_s s,\ k_s \in K\}.$$

Calling a linear combination $\sum_{s\in S} k_s s$ finite means that only finitely many of the coefficients $k_s \in K$ are non-zero. Addition of these linear combinations and multplication by some $k \in K$ gives KS the structure of a vector space. Note that a basis of KS is given by the elements of S (by regarding $s \in S$ as a particular simple linear combination of elements in S). In particular, the vector space KS could be called the free K-module generated by S.

A map $f\colon S \to T$ of sets determines a linear map

$$f_*\colon KS \longrightarrow KT \qquad \text{defined by} \qquad f_*(\sum_{s\in S} k_s s) := \sum_{s\in S} k_s f(s).$$

Note that $f(s)$ is in element of T, and so $\sum_{s\in S} k_s f(s)$ is indeed a finite linear combination of elements of T.

(2) A functor $\mathsf{Set} \to \mathsf{Ab}$ to the category Ab of abelian groups and group homomorphisms can be defined very similarly, by mapping a set S to

$$\mathbb{Z}S := \{\text{(formal) finite linear combinations } \sum_{s\in S} n_s s,\ n_s \in \mathbb{Z}\}.$$

$\mathbb{Z}S$ is called the *free abelian group generated by* S. A map $f\colon S \to T$ of sets determines a group homomorphism $f_*\colon \mathbb{Z}S \to \mathbb{Z}T$ as in the example above.

(3) There is a functor $\mathsf{Set} \to \mathsf{Grp}$ which maps a set S to

$$F(S) := \{\text{words whose letters are } s^{\pm 1} \text{ for } s \in S\}/\sim$$

Here a word obtained by deleting a string of the form ss^{-1} or $s^{-1}s$ from a word W is identified with W, and the equivalence relation $\sim$ is generated by this. Concatenation of words gives a well-defined multiplication on these equivalence classes. This gives $F(S)$ a group structure, with the identity element given by the empty word, and the inverse given by replacing each letter $s^{\pm 1}$ by $s^{\mp 1}$ and writing the letters in reverse order. The group $F(S)$ is called the *free group generated by* S.

A map $f\colon S \to T$ of sets induces a homomorphism $f_*\colon F(S) \to F(T)$ which maps (an equivalence class of) a word $s_1^{\epsilon_1} \dots s_k^{\epsilon_k}$, $\epsilon_i \in \{\pm 1\}$ to $f(s_1)^{\epsilon_1} \dots f(s_k)^{\epsilon_k}$.

4.2 Product, coproducts and pushouts

4.2.1 Products

Let X_1, X_2 be sets and let $X := X_1 \times X_2$ be their Cartesian product. The product X is related to its factors X_1, X_2 via the projection maps

$$X_1 \xleftarrow{p_1} X \xrightarrow{p_2} X_2.$$

We note that map f from a set Y to the product $X = X_1 \times X_2$ are easy to describe: a pair of maps $f_1\colon Y \to X_1$, $f_2\colon Y \to X_2$ uniquely determines a map $f\colon Y \to X$ whose component maps are f_1, f_2. This statement can be concisely expressed by the commutative diagram

$$\begin{tikzcd}
 & X_1 \\
Y \arrow[ur, "f_1"] \arrow[r, dashed, "\exists! f"] \arrow[dr, "f_2"'] & X_1 \times X_2 \arrow[u, "p_1"'] \arrow[d, "p_2"] \\
 & X_2
\end{tikzcd}$$

We note the the commutativity of the top triangle, i.e., $f_1 = p_1 \circ f$ expresses the requirement that the first component of f is f_1, while the commutativity of the bottom triangle forces the second component of f to be f_2.

The benefit of describing a property in terms of a commutative diagram is that the same statement can be made *in any category.* This motivates the following definition.

Definition 4.10. Let X_1, X_2 be objects in a category $\mathcal{C}$. An object X in $\mathcal{C}$ is called a *categorical product* (often denoted $X_1 \times X_2$) if there are morphism $p_1 \colon X \to X_1$ and $p_2 \colon X \to X_2$ such that the diagram $X_1 \xleftarrow{p_1} X \xrightarrow{p_2} X_2$ has the property expressed by the commutative diagram

$$\begin{tikzcd} & X_1 \\ Y \arrow[ur, "f_1"] \arrow[r, dashed, "\exists! f"] \arrow[dr, "f_2"'] & X \arrow[u, "p_1"'] \arrow[d, "p_2"] \\ & X_2 \end{tikzcd} \tag{4.11}$$

Expressed in words, this is: for any pair of morphisms $f_1 \colon Y \to X_1$, $f_2 \colon Y \to X_2$, there is a unique morphism $f \colon Y \to X$ making the diagram commutative.

The universal property (4.11) implies that the map

$$\mathcal{C}(Y, X) \longrightarrow \mathcal{C}(Y, X_1) \times \mathcal{C}(Y, X_2) \qquad f \mapsto (p_1 \circ f, p_2 \circ f) \tag{4.12}$$

is a bijection. In particular, it is easy to understand maps whose codomain is a categorical product, i.e., maps to a categorical product.

A categorical product of two objects X_1, X_2 in a category $\mathcal{C}$, i.e., an object X with the property expressed by diagram (4.11), might or might not exist. Here is a an example showing that products might not exist in a category $\mathcal{C}$.

Example 4.13. Non-existence of categorical products. where For let $\mathcal{C}$ be the category whose objects are the sets of cardinality $\leq n$ and whose morphisms are maps between these sets. We claim that the categorical product of two sets X_1, X_2 exists if and only if $|X_1||X_2| \leq n$, where $|X| \in \mathbb{N}_0$ denotes the cardinality of a finite set X.

The number $|X_1||X_2|$ is the cardinality of the Cartesian product $X_1 \times X_2$ of these sets, which we know is a categorical product of these in the category Set. If $|X_1||X_2| \leq n$, then $X_1 \times X_2$ belongs to the subcategory $\mathcal{C}$, and is in particular also a categorical product of X_1 and X_2 in $\mathcal{C}$.

To show that for $|X_1||X_2| > n$ no categorical product X in $\mathcal{C}$ exists, suppose that X is a categorical product of X_1 and X_2. Then we have the bijection (4.12). For finite sets X, Y, the morphism set $\mathrm{mor}_{\mathsf{Set}}(Y, X) = \mathrm{map}(Y, X)$, the set of maps from Y to X has cardinality $|X|^{|Y|}$. Hence the bijection (4.12) implies

$$|X|^{|Y|} = |X_1|^{|Y|} \cdot |X_2|^{|Y|} = (|X_1||X_2|)^{|Y|},$$

and hence $|X| = |X_1||X_2|$ for $|Y| = 1$. This shows that there is no categorical product X of X_1 and X_2 in the subcategory $\mathcal{C}$ for $|X_1||X_2| > n$.

Even if a categorical product exists, it is in general not unique; in fact, any object isomorphic to X will also have this property. However, X is *unique up to isomorphisms* in the following sense.

Lemma 4.14. *Let X_1, X_2 be objects of a category $\mathcal{C}$ and let X and X' be categorical products of X_1 and X_2. Then X and X' are isomorphic.*

Proof. Let $X_1 \xleftarrow{p_1} X \xrightarrow{p_2} X_2$ and $X_1' \xleftarrow{p_1'} X' \xrightarrow{p_2'} X_2'$ be the two diagrams that satisfy the universal property (4.11) (the maps p_i, p_i' exists due to the assumption that X, X' are categorical products). Then consider the diagram

$$\begin{tikzcd}
& X_1 & \\
X' \arrow[ur, "p_1'"] \arrow[dr, "p_2'"'] \arrow[rr, dashed, shift right, "f"'] & & X \arrow[ul, "p_1"'] \arrow[dl, "p_2"] \arrow[ll, dashed, shift right, "g"'] \\
& X_2 &
\end{tikzcd}$$

The morphism f exists by the property (4.11) for X (applied to $Y = X'$), g exists by the property (4.11) for X' (applied to $Y = X$), and the composition $f \circ g$ is the identity morphism id_X by the uniqueness statement of the property (4.11) for X (applied to $Y = X$). Similarly, $g \circ f = \mathrm{id}_{X'}$, and hence X, X' are isomorphic. $\square$

Proposition 4.15. *Let $\mathcal{C}$ be the category* Set, Grp, Vect, Top *or* Top_*. *Then the categorical product of objects X_1 and X_2 is given by the usual Cartesian product $X_1 \times X_2$ equipped with the usual projection maps $p_i \colon X_1 \times X_2 \to X_i$ for $i = 1, 2$.*

Proof. For $\mathcal{C} = \mathsf{Set}$ we have already checked that the Cartesian product of sets X_1, X_2 has the universal property (4.11) of the categorical product – this was our motivating example.

To check that the Cartesian product $X_1 \times X_2$ of two topological spaces X_1, X_2 has the property (4.11), let $f_1 \colon Y \to X_1$ and $f_2 \colon Y \to X_2$ be continuous maps. Then by the universal property of $X_1 \times X_2$ as sets, there exists a unique map of sets $f \colon Y \to X_1 \times X_2$. So it only remains to show that f is continuous. This is the case, since a map $f \colon Y \to X_1 \times X_2$ to the

Cartesian product of topological spaces is continuous if and only if its component maps f_1, f_2 are continuous.

The same line of arguments works for the category of groups or vector spaces, since a map $f\colon Y \to X_1 \times X_2$ to a Cartesian product of groups (resp. vector spaces) is a group homomorphism (resp. a linear map) if and only if its components maps f_1, f_2 are. Similarly, for Top_*, it suffices to observe that f is basepoint preserving if and only if f_1 and f_2 are.

$\square$

4.2.2 Coproducts

Before characterizing the coproduct of objects X_1, X_2 in a category $\mathcal{C}$ by a universal property in Definition 4.17, we discuss the disjoint union of sets as a motivating example.

Definition 4.16. Let X_1, X_2 be sets. The *disjoint union* of X_1 and X_2, denoted $X_1 \amalg X_2$ is defined to be the set

$$X_1 \amalg X_2 := \{(1,x) \mid x \in X_1\} \cup \{(2,x) \mid x \in X_2\} \subset \{1,2\} \times (X_1 \cup X_2).$$

Let $X_1 \xrightarrow{i_1} X_1 \amalg \xleftarrow{i_2} X_2$ be the maps defined by $i_1(x) := (1,x)$ and $i_2(x) := (2,x)$.

Question: Can the disjoint union of two sets be characterized up to isomorphism by a universal property, similar to the universal property (4.11) for the product of sets?

We observe that the images $i_1(X_1)$ and $i_2(X_2)$ are *disjoint* subsets of $X_1 \amalg X_2$ whose union is all of $X_1 \amalg X_2$. Hence any map f from $X_1 \amalg X_2$ to some set Y is uniquely determined by its restriction to the image of i_1 resp. i_2. Since the maps i_1, i_2 are injective, this means that f is uniquely determined by the compositions $f_1 := f \circ i_1$ and $f_2 := f \circ i_2$. As in the case of the Cartesian product, this property of the disjoint union of sets can neatly be expressed by the following commutative diagram.

$$\begin{tikzcd}
X_1 \arrow[d, "i_1"'] \arrow[dr, "f_1"] & \\
X_1 \amalg X_2 \arrow[r, dashed, "\exists! f"] & Y \\
X_2 \arrow[u, "i_2"] \arrow[ur, "f_2"'] &
\end{tikzcd}$$

Definition 4.17. Let X_1, X_2 be objects of a category $\mathcal{C}$. An object X in $\mathcal{C}$ is called the *coproduct* of X_1 and X_2 (often denoted $X_1 \amalg X_2$) if there are morphisms

$$X_1 \xrightarrow{i_1} X \xleftarrow{i_2} X_2$$

such that this pair of maps satisfies the universal property expressed by the following commutative diagram

$$\begin{tikzcd}
X_1 \arrow[d, "i_1"'] \arrow[dr, "f_1"] & \\
X \arrow[r, dashed, "\exists ! f"] & Y \\
X_2 \arrow[u, "i_2"] \arrow[ur, "f_2"'] &
\end{tikzcd} \tag{4.18}$$

Comparing the universal property of coproducts (4.18) with the universal property of products (4.11), we notice that these are basically the *same* diagrams; the only difference is that the direction of the maps in these diagram is the opposite. This motivates the terminology "coproduct"; the prefix "co" in category theory often referring to reversing all arrows in a diagram.

The universal of a coproduct X implies that the map

$$\mathcal{C}(X,Y) \longrightarrow \mathcal{C}(X_1,Y) \times \mathcal{C}(X_2,Y) \qquad f \mapsto (f \circ i_1, f \circ i_2) \tag{4.19}$$

is a bijection. In particular, it is easy to understand morphisms whose domain is a coproduct.

A coproduct might or might not exist. For example, in the category $\mathcal{C}$ of sets of cardinality $\leq n$, a coproduct of X_1, X_2 exists if and only if $|X_1| + |X_2| \leq n$, as can be shown by considerations analogous to those in Example 4.13. As in Lemma 4.14, any two coproducts are isomorphic.

Proposition 4.20. *Coproducts (of two arbitrary objects X_1, X_2) exist in the categories* Set, Top, Top_*, Vect *and* Grp. *The following table shows the usual notation and terminology for the coproduct in these categories.*

category	*coproduct of objects*
Set	$X_1 \amalg X_2$, *the disjoint union of sets* X_1, X_2
Top	$X_1 \amalg X_2$, *the disjoint union of topological spaces* X_1, X_2
Top_*	$X_1 \vee X_2$, *the wedge product of topological spaces* X_1, X_2
Grp	$X_1 * X_2$, *the free product of groups* X_1, X_2

The disjoint union of sets and its universal property was for us the example motivating the general definition of the coproduct, and hence the theorem holds in case of the category Set. To prove the result for the other categories, we will go through them one by one, first giving in each case an explicit construction of a candidate for the coproduct (including the morphisms i_1 and i_2), namely the disjoint union of topological spaces, the wedge product of pointed topological spaces, and the free product of groups. Then we show in each case that this object satisfies the universal property of the coproduct.

Definition 4.21. (The disjoint union of topological spaces). Let X_1, X_2 be topological spaces. The *disjoint union* $X_1 \amalg X_2$ is the topological space whose underlying set is the disjoint union of X_1 and X_2, considered as sets. The topology on the set $X_1 \amalg X_2$ is defined by declaring a subset $U \subset X_1 \amalg X_2$ to be open if and only if $i_1^{-1}(U)$ is an open subset of X_1 and $i_2^{-1}(U)$ is an open subset of X_2. With this topology on $X_1 \amalg X_2$ the maps

$$X_1 \xrightarrow{i_1} X_1 \amalg X_2 \xleftarrow{i_2} X_2 \tag{4.22}$$

are both continuous.

Lemma 4.23. *The diagram* (4.22) *satisfies the universal property* (4.18)*. In particular, the disjoint union of topological spaces is the coproduct in the category* Top*.*

Proof. It is clear that there is at most one continuous map $f\colon X_1 \amalg X_2 \to Y$ making the diagram (4.18) for given maps f_1, f_2, since the underlying set $X_1 \amalg X_2$ is the coproduct of the sets X_1, X_2, and hence there is exactly one map $f\colon X_1 \amalg X_2 \to Y$ making the diagram commutative (without insisting on its continuity).

So it remains to show that f is continuous if f_1 and f_2 are. So let V be an open subset of Y. Then $f^{-1}(V)$ is open, since $i_\ell^{-1}(f^{-1}(V)) = (f \circ i_\ell)^{-1}(V) = f_\ell^{-1}(V)$ is open, since f_ℓ is continuous for $\ell = 1, 2$. □

Let (X_1, x_1), (X_2, x_2) be pointed topological spaces. We need to come up with a candidate for the coproduct of these pointed spaces. Note that the disjoint union $X_1 \amalg X_2$ is not a good candidate, since we would like the maps (4.22)

$$X_1 \xrightarrow{i_1} X_1 \amalg X_2 \xleftarrow{i_2} X_2$$

to be basepoint preserving, but $i_1(x_1) \neq i_2(x_2)$. The way to fix this is to pass to a quotient space of $X_1 \amalg X_2$ where we identify these two points.

Definition 4.24. Let (X_1, x_1), (X_2, x_2) be pointed topological spaces. The quotient space

$$X_1 \vee X_2 := (X_1 \amalg X_2)/\{i_1(x_1), i_2(x_2)\}$$

equipped with the base point $*$ given by the equivalence class represented by these two points $i_1(x_1)$, $i_2(x_2)$ is the *wedge product* of the pointed spaces X_1, X_2.

Lemma 4.25. *For $\ell = 1, 2$ let $k_\ell \colon X_\ell \to X_1 \amalg X_2$ be the composition of $i_\ell \colon X_\ell \to X_1 \amalg X_2$ and the projection map $p \colon X_1 \amalg X_2 \to X_1 \vee X_2$. Then the diagram*

$$X_1 \xrightarrow{k_1} X_1 \vee X_2 \xleftarrow{k_2} X_2 \tag{4.26}$$

satisfies the universal property (4.18)*. In particular, the wedge product $X_1 \vee X_2$ of pointed topological spaces X_1, X_2 is the coproduct in the category Top_*.*

Proof. The universal property we need to check is expressed by the commutative diagram

$$\begin{tikzcd} X_1 \arrow[d, "k_1"'] \arrow[dr, "f_1"] & \\ X_1 \vee X_2 \arrow[r, dashed, "\exists! f"] & Y \\ X_2 \arrow[u, "k_2"] \arrow[ur, "f_2"'] & \end{tikzcd} \tag{4.27}$$

By the universal property of the disjoint union $X_1 \amalg X_2$, there is a unique continuous map $\widetilde{f} \colon X_1 \amalg X_2 \to Y$ such that $\widetilde{f} \circ i_1 = f_1$ and $\widetilde{f} \circ i_2 = f_2$. Since f_1, f_2 are basepoint preserving, $\widetilde{f}(i_1(x_1)) = \widetilde{f}(x_1) = y_0$ and $\widetilde{f}(i_2(x_2)) = \widetilde{f}(x_2) = y_0$, where $y_0 \in Y$ is the basepoint in Y. This implies that $\widetilde{f}$ factors through the quotient space $X_1 \vee X_2 = (X_1 \amalg X_1)/\{i_1(x_1), i_2(x_2)\}$, i.e., $\widetilde{f}$ can be written as composition

$$X_1 \amalg X_2 \xrightarrow{p} X_1 \vee X_2 \xrightarrow{f} Y$$

for a unique map f. This map is basepoint preserving. It also is continuous: by the continuity criterion for maps out of a quotient space 1.24, the map f is continuous if and only if the composition $f \circ p = \widetilde{f}$ is continuous. $\square$

Definition 4.28. Let X_1, X_2 be groups. Their *free product* $X_1 * X_2$ is the group whose elements are equivalence classes of words $s_1 \dots s_k$ whose letter s_i belong to X_1 or X_2 (we assume that X_1, X_2 are disjoint as sets). The equivalence relation $\sim$ on these words is generated by

(1) $s_1 \dots s_i \dots s_k \sim s_1 \dots \widehat{s_i} \dots s_k$ if s_i is the identity element of X_1 or X_2, and $s_1 \dots \widehat{s_i} \dots s_k$ is the word obtained by deleting the letter s_i.

(2) $s_1 \dots s_i s_{i+1} \dots s_k \sim s_1 \dots (s_i \cdot s_{i+1}) \dots s_k$ if s_i and s_{i+1} both belong to X_1 or to X_2 and $s_i \cdot s_{i+1} \in X_j$ denotes their product in that group.

The multiplication in $X_1 * X_2$ is induced by concatenation of words. The identity element is represented by the empty word, and the inverse of the element represented by the word $s_1 \dots s_k$ is given by $s_k^{-1} \dots s_1^{-1}$. Let

$$X_1 \xrightarrow{i_1} X_1 * X_2 \xleftarrow{i_2} X_2 \tag{4.29}$$

be the group homomorphisms given by sending an element s of X_1 or X_2 to the element $[s] \in X_1 * X_2$ represented by the one-letter-word s.

Lemma 4.30. *The diagram* (4.29) *satisfies the universal property* (4.18)*. In particular, the free product $X_1 * X_2$ of groups X_1, X_2 is the coproduct in the category* Grp*.*

The proof of this lemma is left as a homework problem.

4.2.3 Pushouts

Before defining what a pushout is in a category in Definition 4.33, we consider a motivating example of a pushout in the category of topological spaces.

Example 4.31. Let X_1, X_2 be open subsets of a topological space X. Then considering the inclusion maps relating X_1, X_2, X and $X_1 \cap X_2$ we have the following commutative square in the category Top of topological spaces:

$$\begin{tikzcd} X_1 \cap X_2 \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \arrow[d, "k_1"] \\ X_2 \arrow[r, "k_2"] & X \end{tikzcd} \tag{4.32}$$

Let $f_1 \colon X_1 \to Y$ and $f_2 \colon X_2 \to Y$ be continuous maps which agree on the subspace $X_1 \cap X_2$. Then there is a unique well-defined map $f \colon X \to Y$ whose restriction to X_1 is the map f_1 and whose restriction to X_2 is the map f_2. Moreover, by an earlier homework problem, the continuity of f_1 and f_2 imply the continuity of the map f (here we use the assumption that X_1, X_2 are open subsets of X).

We note that the last paragraph can be expressed as the following commutative diagram of topological spaces and continuous maps.

$$\begin{tikzcd} X_1 \cap X_2 \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \arrow[d, "k_1"] \arrow[ddr, bend left, "f_1"] & \\ X_2 \arrow[r, "k_2"] \arrow[drr, bend right, "f_2"'] & X \arrow[dr, dashed, "\exists! f"] & \\ & & Y \end{tikzcd}$$

The advantage of expressing a statement in form of a commutative diagram is that we can consider diagrams of the same shape in any category. Doing this for the diagram above leads to the following definition.

Definition 4.33. Let $\mathcal{C}$ be a category, and let

$$\begin{tikzcd} A \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \arrow[d, "k_1"] \\ X_2 \arrow[r, "k_2"] & X \end{tikzcd}$$

be a commutative diagram of objects and morphisms in $\mathcal{C}$. This diagram is a *pushout diagram* or *pushout square* if it satisfies the universal property expressed by the diagram

$$\begin{tikzcd} A \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \arrow[d, "k_1"] \arrow[ddr, bend left, "f_1"] & \\ X_2 \arrow[r, "k_2"] \arrow[drr, bend right, "f_2"'] & X \arrow[dr, dashed, "\exists! f"] & \\ & & Y \end{tikzcd} \tag{4.34}$$

The object X is called a *pushout* of the diagram

$$\begin{tikzcd} A \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \\ X_2 & \end{tikzcd} \tag{4.35}$$

We observe that if $\mathcal{C}$ is the category Set of sets, and A is the empty set, then the universal property expressed by the diagram (4.34) is that X is a coproduct of X_1 and X_2 (more generally, this is true if A is an object of the category $\mathcal{C}$ which is *initial*, which means that there is a unique morphism from A to any other object of $\mathcal{C}$). So a pushout is a generalization of a coproduct.

Being a pushout is a universal property for X; in particular, a pushout of a diagram of the shape (4.35) in a category $\mathcal{C}$ might or might not exist, and any two pushouts of the same diagram are isomorphic. Going back to Example 4.31, this fact that the diagram (4.32) is a pushout diagram implies that the space X is determined by the diagram of topological spaces

$$\begin{tikzcd} X_1 \cap X_2 \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \\ X_2 & \end{tikzcd}$$

since X is a pushout of this diagram.

Theorem 4.36. *The pushout of any diagram* (4.35) *exists in the categories* Top, Top_* *and* Grp. *The following table shows the usual notation and terminology for pushouts in these categories.*

category	*pushout*
Set	$X_1 \cup_A X_2$
Top	$X_1 \cup_A X_2$
Top_*	$X_1 \cup_A X_2$
Grp	$X_1 *_A X_2$, *the amalgamated free product*

There seems to be no standard terminology for $X_1 \cup_A X_2$. Both of the notations $X_1 \cup_A X_2$ and $X_1 *_A X_2$ suppress the dependence of this object on the morphisms j_1, j_2. To indicate the dependence, some people use the notation $X_1 \cup_{j_1,A,j_2} X_2$ and $X_1 *_{j_1,A,j_2} X_2$.

To prove this, we need to

(i) construct a *candidate object* X for the pushout of a diagram of the shape (4.35) in each of the categories $\mathcal{C}$ considered,

(ii) define morphisms $k_1 \colon X_1 \to X$, $k_2 \colon X_2 \to X$,

(iii) show that the square (4.32) commutes, and finally,

(iv) show that the square is in fact a pushout square.

Proof of Thm.4.36 in the category Set. Here is our first try: take as candidate X for the pushout the disjoint union $X_1 \amalg X_2$. Consider the diagram

$$\begin{tikzcd} A \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \arrow[d, "i_1"] \\ X_2 \arrow[r, "i_2"] & X_1 \amalg X_2 \end{tikzcd}$$

where i_1, i_2 are the standard inclusions into the disjoint union $X_1 \amalg X_2$. Alas, there is a problem at step (iii): this diagram is not commutative, since for $a \in A$, the element $i_1(j_1(a))$ belongs to the image of i_1, while $i_2(j_2(a))$ belongs to the image of i_2, and these two images are disjoint subsets of $X_1 \amalg X_2$.

We can fix this problem by replacing the disjoint union $X_1 \amalg X_2$ by its quotient, where we identify the points $i_1(j_1(a))$ and $i_2(j_2(a))$. Here is the formal definition for future reference.

Definition 4.37. Let $j_1\colon A \to X_1$ and $j_2\colon A \to X_2$ be maps of sets. Then we define

$$X_1 \cup_A X_2 := (X_1 \amalg X_2)/\sim,$$

where the equivalence relation $\sim$ is generated by $i_1(j_1(a)) \sim i_2(j_2(a))$ for $a \in A$. Here $i_\ell\colon X_\ell \to X_1 \amalg X_2$ are the inclusion maps featured in the definition of the disjoint union 4.37.

Let $p\colon X_1 \amalg X_2 \twoheadrightarrow X_1 \cup_A X_2$ be the projection map, and for $m = 1, 2$, let k_m be the composition

$$X_m \xrightarrow{i_m} X_1 \amalg X_2 \xrightarrow{p} X_1 \cup_A X_2.$$

Then by construction of $X_1 \cup_A X_2$, the diagram

$$\begin{tikzcd}
A \arrow[r, "j_1"] \arrow[d, "j_2"'] & X_1 \arrow[d, "k_1"] \\
X_2 \arrow[r, "k_2"] & X_1 \cup_A X_2
\end{tikzcd}$$

is commutative. So it remains to show that this diagram is a pushout square, the universal property expressed by the diagram (4.34). To do so, we replicate that diagram, but add to it the disjoint union $X_1 \amalg X_2$ and various maps with (co)domain $X_1 \amalg X_2$.

$$\begin{tikzcd}
 & & X_1 \arrow[dl, "i_1"] \arrow[dr, "k_1"] \arrow[drr, bend left, "f_1"] & & \\
A \arrow[urr, bend left, "j_1"] \arrow[drr, bend right, "j_2"'] & X_1 \amalg X_2 \arrow[rr, "p"] \arrow[rrr, dotted, bend right, "\widetilde{f}"'] & & X_1 \cup_A X_2 \arrow[r, dashed, "f"] & Y \\
 & & X_2 \arrow[ul, "i_1"] \arrow[ur, "k_1"'] \arrow[urr, bend right, "f_2"'] & &
\end{tikzcd}$$

The solid arrows are given and they make the diagram consisting of all solid arrows commutative. Our goal is to show that there is a unique map f making the diagram commutative. To construct f, we note that by the universal property of $X_1 \amalg X_2$ as coproduct, the maps f_1 and f_2 uniquely determine a map $\widetilde{f}$ the addition of which keeps the diagram commutative.

complete proof $\square$

Still to do here: Definition of free amalgamated product of groups, and its universal property as a pushout – using the same diagram as above, but in the category of groups

4.3 The Seifert van Kampen Theorem

Theorem 4.38. *Let U, V be open subsets of topological space X such that $U \cup V = X$ and $U \cap V$ is path connected. Let*

$$\begin{array}{ccc} U \cap V & \xrightarrow{j^U} & U \\ \downarrow{\scriptstyle j^V} & & \downarrow{\scriptstyle i^U} \\ V & \xrightarrow[i^V]{} & X \end{array}$$

*be the pushout diagram of topological spaces given by the inclusion maps (see Example **??** and Definition 4.33). Then for $x_0 \in U \cap V$ the induced commutative diagram of fundamental groups*

$$\begin{array}{ccc} \pi_1(U \cap V, x_0) & \xrightarrow{j^U} & \pi_1(U, x_0) \\ \downarrow{\scriptstyle j^V} & & \downarrow{\scriptstyle i^U} \\ \pi_1(V, x_0) & \xrightarrow[i^V]{} & \pi_1(X, x_0) \end{array}$$

is a pushout diagram in the category of groups.

The Seifert van Kampen Theorem says in particular that the fundamental group functor preserves pushouts (under suitable additional assumptions).

Corollary 4.39. *With the assumptions of the theorem, the fundamental group $\pi_1(X, x_0)$ is isomorphic to the free amalgamated product $\pi_1(U, x_0) *_{\pi_1(U \cap V, x_0)} \pi_1(V, x_0)$.*

Corollary 4.40. *Let (X_1, x_1), (X_2, x_2) be pointed topological spaces such that x_1, x_2 have contractible open neighborhoods U_1, U_2. Let $j_1 \colon X_1 \to X_1 \vee X_2$ and $j_2 \colon X_2 \to X_1 \vee X_2$ be the inclusion maps. Then the map*

$$\pi_1(X_1) * \pi_1(X_2) \longrightarrow \pi_1(X_1 \vee X_2)$$

given by

$$\pi_1(X_1) \ni c_1 \mapsto (j_1)_*(c) \in \pi_1(X_1 \vee X_2)$$
$$\pi_1(X_2) \ni c_2 \mapsto (j_2)_*(c_2) \in \pi_1(X_1 \vee X_2)$$

is an isomorphism of groups.

proof missing

Lemma 4.41. *For $n \geq 2$ the fundamental group $\pi_1(S^n)$ is trivial.*

proof missing

Groups are often described in terms of generators and relations. This description of a group G is called a *presentation* of G.

Definition 4.42. Let $S = \{s_1, s_2, \dots\}$ be a set, and let $r_1, \dots, r_\ell$ be words with letters s_i, s_i^{-1}. The words r_j in particular represent elements in the free group $F(S)$ generated by the set S (see Example 4.9(3)). Abusing notation, we write $r_j \in F(S)$. These data determine a group

$$\langle s_1, s_2, \cdots \mid r_1, \dots, r_\ell \rangle := F(S)/N,$$

where N is the normal subgroup of $F(S)$ generated by $r_1, \dots, r_\ell \in F(S)$. It is often written as $\langle S \mid R \rangle$, where $R := \{r_1, \dots, r_\ell\}$.

A *presentation* of a group G consists of a set S, a set $R = \{r_1, \dots, r_k\}$ of elements $r_j \in F(S)$, and a group isomorphism

$$\Phi\colon \langle S|R \rangle \longrightarrow G.$$

The group homomorphism

$$F(S) \xrightarrow{\;p\;}\!\!\!\!\rightarrow F(S)/N = \langle S \mid R \rangle \xrightarrow[\cong]{\Phi} G$$

maps the elements of S to elements of G which generate the group G. Hence the elements of S are called *generators* of G. The elements $r_j \in R$ map to the trivial element in G, and are hence called *relations*.

Proposition 4.43. *Let $L = \{a_1, \dots, a_k\}$ be a label set and let $w = a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}$, $\epsilon_i \in \{\pm 1\}$, be an n-letter words with letters from the alphabet $\{a_1, \dots, a_k\} \cup \{a_1^{-1}, \dots, a_k^{-1}\}$. Let $\Sigma(w) = P_n/\sim_w$ be the quotient space of the n-gon P_n via the edge identification determined by the word w (see Definition 2.18). Assume:*

(i) *Every label $a_i \in L$ occurs in w (i.e., a_i is a letter of w, or a_i^{-1} is a letter).*

(ii) *All vertices of P_n are equivalent for $\sim_w$.*

Then $\pi_1(\Sigma(w)) \cong \langle a_1, \dots, a_k \mid w \rangle$.

Example 4.44. The projective plane $\mathbb{RP}^2$ and the Klein bottle K can be described as polygons with edge identifications (see (2.17)). Hence by the previous result, it is easy to read off the fundamental group.

1. $\mathbb{RP}^2 \approx \Sigma(aa)$, and hence $\pi_1(\mathbb{RP}^2) \cong \langle a \mid aa \rangle = \mathbb{Z}/2$.

2. $K \approx \Sigma(aba^{-1}b)$, and hence $\pi_1(K) \cong \langle a, b \mid aba^{-1}b \rangle$.

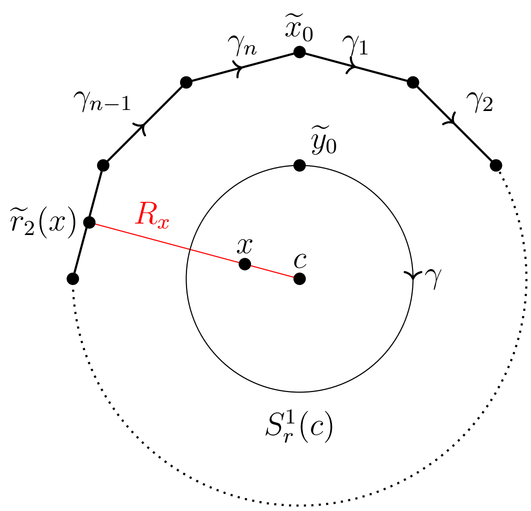
Similarly, we have described all compact connected 2-manifolds as polygons with edge identifications, and hence it is easy to read off a presentation of their fundamental group from this description.

Proof of Prop. 4.43. To use the Seifert van Kampen Theorem for the calculation of the fundamental group of $X = \Sigma(w) = P_n/\sim_w$, we need to write X as the union of open subsets $X_1, X_2 \subset X$. Let $p\colon P_n \twoheadrightarrow X$ be the projection map, let $c \in P_n$ be the center point of the regular n-gon P_n, let ∂P_n be the boundary of P_n, consisting of all edges, and let $\mathring{P}_n := P_n \setminus \partial P_n$ be the interior of P_n. Let

$$X_1 := p(\mathring{P}_n) \qquad X_2 := p(P_n \setminus \{c\}) \quad \text{and hence} \quad X_1 \cap X_2 = p(\mathring{P}_n \setminus \{c\}).$$

These are open subspaces of the quotient space X, since $p^{-1}(X_1) = \mathring{P}_n$ and $p^{-1}(X_2) = P_n \setminus \{c\}$ are open subsets of P_n.

Here is a picture of the polygon P_n with its centerpoint c. The other things shown in the picture will be explained as we go along.



Next we calculate the fundamental groups of X_1, X_2 and $X_1 \cap X_2$. This requires us to choose a basepoint in $X_1 \cap X_2$. Let $\widetilde{x}_0$ be the distinguished vertex of P_n, and let $x_0 := p(\widetilde{x}_0) \in X$. This is a good choice of a basepoint for X in the sense that each edge γ_i of P_n projects to a path in X which by assumption (ii) starts and ends at x_0, i.e., is a based loop in (X, x_0) and hence represents an element in $\pi_1(X, x_0)$.

However, applying the Seifert van Kampen Theorem *requires* the choice of a basepoint in the intersection $X_1 \cap X_2$, which is not the case for x_0. We choose a point $y_0 \in X_1 \cap X_2$ as follows. Let $S^1_r(c)$ be the circle of radius r around the center point $c \in P_n$, where $r > 0$ is chosen such that $S^1_r(c)$ is contained in the interior $\mathring{P}_n$. Let

$$\widetilde{y}_0 \in S^1_r(c) \subset \mathring{P}_n$$

be the intersection point of the straight line from $\widetilde{x}_0$ to c with the circle $S^1_r(c)$ (see picture), and let $y_0 = p(\widetilde{y}_0) \in X_1 \cap X_2$.

- The projection map p restricts to a homeomorphism from $\mathring{P}_n$ to X_1, since the equivalence relation $\sim_w$ only identified distinct points in ∂P_n with each other. Hence

$$\pi_1(X_1, y_0) \cong \pi_1(\mathring{P}_n, \widetilde{y}_0) = \{1\},$$

since P_n is convex.

- Similarly, p restricts to a homeomorphism of $\mathring{P}_n \setminus \{c\}$ to $X_1 \cap X_2$, and hence the induced map p_* yields an isomorphism

$$p_* \colon \pi_1(\mathring{P}_n \setminus \{c\}, \widetilde{y}_0) \xrightarrow{\cong} \pi_1(X_1 \cap X_2, y_0).$$

We claim that the the circle $S^1_r(c) \subset \mathring{P}_n \setminus \{c\}$ is a deformation retract of $\mathring{P}_n \setminus \{c\}$. A retraction

$$r \colon \mathring{P}_n \setminus \{c\} \longrightarrow S^1_r(c)$$

is given by mapping a point $x \in \mathring{P}_n \setminus \{c\}$ to the intersection point $r(x)$ of the straight line connecting x and c with the circle $S^1_r(c)$. A homotopy from the identity on $\mathring{P}_n \setminus \{c\}$ to the composition $i \circ r$ is given by

$$H \colon (\mathring{P}_n \setminus \{c\}) \times I \longrightarrow S^1_r(c) \qquad H(x,t) = (1-t)x + tr(x).$$

In particular, $\pi_1(\mathring{P}_n \setminus \{c\}, \widetilde{y}_0)$ is isomorphic to $\pi_1(S^1_r(c), \widetilde{y}_0) \cong \mathbb{Z}$, and the path γ that runs once counterclockwise around the circle $S^1_r(c)$ starting and ending at $\widetilde{y}_0$ represents a generator of $\pi_1(\mathring{P}_n \setminus \{c\}, \widetilde{y}_0)$. Hence

$$\pi_1(X_1 \cap X_2, y_0) \cong \mathbb{Z} \qquad \text{with generator} \qquad p_*([\gamma]) = [p \circ \gamma].$$

- We claim that $X_2 = (P_n \setminus \{c\})/\sim_w$ deformation retracts to the subspace $\partial P_n/\sim_w$. To prove this, we first show that ∂P_n is a deformation retract of $P_n \setminus \{c\}$. Let $\widetilde{r}_2 \colon P_n \setminus \{c\} \to \partial P_n$ be defined by mapping $x \in P_n \setminus \{c\}$ to the intersection point of the ray R_x starting at c through x with the boundary ∂P_n (see picture). Let $\widetilde{i}_2 \colon \partial P_n \to P_n \setminus \{c\}$ be the inclusion map. Then

$$\widetilde{H}_2 \colon (P_n \setminus \{c\}) \times I \longrightarrow \partial P_n \qquad \text{defined by} \qquad \widetilde{H}_2(x,t) := (1-t)x + t\widetilde{r}_2(x)$$

is a homotopy from the identity to $\widetilde{i}_2 \circ \widetilde{r}_2$. We observe that the retraction map $\widetilde{r}_2$ and the homotopy $\widetilde{H}_2$ is compatible with the equivalence relation $\sim_w$. Hence they induce well-defined maps r_2 and H_2 on the quotient spaces that make the diagrams

$$\begin{tikzcd} P_n \setminus \{c\} \arrow[r, "\widetilde{r}_2"] \arrow[d, two heads] & \partial P_n \arrow[d, two heads] \\ (P_n \setminus \{c\})/\sim_w \arrow[r, "r_2"] & \partial P_n/\sim_w \end{tikzcd} \qquad \begin{tikzcd} (P_n \setminus \{c\}) \times I \arrow[r, "\widetilde{H}_2"] \arrow[d, two heads] & \partial P_n \arrow[d, two heads] \\ (P_n \setminus \{c\})/\sim_w \times I \arrow[r, "H_2"] & \partial P_n/\sim_w \end{tikzcd}$$

commutative. A by now familiar argument shows that the continuity of the maps $\widetilde{r}_2$, $\widetilde{H}_2$ implies the continuity of the maps r_2 and H_2.

This proves our claim that X_2 deformation retracts to $\partial P_n/\sim_w$. We note that $\widetilde{r}_2(\widetilde{y}_0) = \widetilde{x}_0$ and hence $r_2(y_0) = x_0$. It follows that the retraction map r_2 induces an isomorphism

$$(r_2)*\colon \pi_1(X_2, y_0) \xrightarrow{\cong} \pi_1(\partial P_n/\sim_w, x_0).$$

don't delete until proof is done which is used to label and orient the edges of P_n according to the word w, as explained just before Definition 2.18. We recall that the label and orientation of the j^{th} edge (counting clockwise starting from the distinguished vertex) is determined by $a_{i_j}^{\epsilon_j}$, the j^{th} letter of the word w: the label of the j^{th} edge is a_{i_j} and the arrow along this edge is pointing clockwise if $\epsilon_j = +1$, and counterclockwise if $\epsilon_j = -1$.

□

5 Covering spaces

5.1 Homotopy lifting property for covering spaces

The goal of this section is to prove the path lifting and homotopy lifting properties of covering maps expressed by Lemma 3.18 and Proposition 3.21 respectively. We begin by recalling the definition of a covering map (Definition 3.16), and then state their homotopy lifting property.

Definition 5.1. A continuous map $p\colon \widetilde{X} \to X$ is a *covering map* if p is surjective, and if for each $x \in X$ there is an open neighborhood U with the property that

- $p^{-1}(U)$ is the disjoint union of open subsets $U_a \subset \widetilde{X}$, $a \in A$, and
- for every $i \in I$ the restriction $p_{|U_a}\colon U_a \to U$ is a homeomorphism.

Any open subset $U \subset X$ with this property is called *evenly covered*. If $p\colon \widetilde{X} \to X$ is a covering map, the space $\widetilde{X}$ is called a *covering space* of X.

Proposition 5.2. (Homotopy lifting for covering maps) *Let $p\colon \widetilde{X} \to X$ be a covering map, let $f\colon Y\times I \to X$ be a homotopy, and let $\widetilde{f}_0\colon Y\times\{0\} \to \widetilde{X}$ be a lift of $f_0\colon Y\times\{0\} \to X$, the restriction of f to $Y \times \{0\} \subset Y \times I$. Then there is a unique lift $\widetilde{f}\colon Y \times I \to \widetilde{X}$ which restricts on $Y \times \{0\}$ to $\widetilde{f}_0$.*

For a one-point-space Y, the map $f\colon I \to X$ is a path in X, and $\widetilde{f}\colon I \to \widetilde{X}$ is a lift of this path with given starting point $\widetilde{f}_0 \in \widetilde{X}$. So this special case expresses the path-lifting property of covering maps (Lemma 3.18).

Proof. We first prove the homotopy lifting property in the special case $Y = \mathrm{pt}$, i.e., the path lifting property. So let $f\colon I \to X$ be a path, and let $\widetilde{f}_0 \in p^{-1}(f(0)) \subset \widetilde{X}$. Our goal is show that there is a unique path $\widetilde{f}\colon I \to \widetilde{X}$ which is a lift of f (i.e., $p \circ \widetilde{f} = f$) with prescribed starting point $\widetilde{f}(0) = \widetilde{f}_0$.

It is easy to find such a lift if $f([0,1])$ is contained in an evenly covered subset $U \subset X$: the preimage $p^{-1}(U)$ is the union of disjoint open subsets $U_a \subset \widetilde{W}$, $a \in A$, and the required starting point $\widetilde{f}_0$ belongs to one of these sets U_a. Since $p_{|U_a}\colon U_a \to U$ is a homeomorphism, a lift $\widetilde{f}\colon I \to \widetilde{X}$ is simply given by the composition

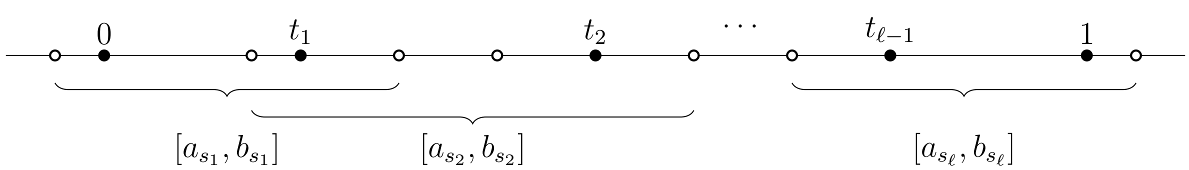
$$I \xrightarrow{\ f\ } U \xrightarrow[\cong]{p_{|U_a}^{-1}} U_a \subset \widetilde{X}.$$

So the idea is to chop the path f into short segments so that each segment is contained in an evenly covered subset of X. Then the construction above is used on each segment, constructing a lifted path $\widetilde{f}$ step by step, one segment at a time.

Claim. There is a partition $0 = t_0 < t_2 < \cdots < t_k = 1$ such that for each $i = 1, \dots, k$ the image $f([t_{i-1}, t_i])$ is contained in an evenly covered open set $U \subset X$.

To prove the claim, we note that for each $s \in I$, the point $f(s) \in X$ is contained in some evenly covered open neighborhood $U_s \subset X$. Then $f^{-1}(U_s)$ is an open neighborhood of s, and hence there is some interval (a_s, b_s) such that $s \in (a_s, b_s) \cap I \subset f^{-1}(U)$ (since these form a basis for the subspace topology of $I \subset \mathbb{R}$). The open subsets $(a_s, b_s) \cap I$, $s \in I$, form an open cover of I, and hence by compactness of I, there is a finite set $S = \{s_1, \dots, s_\ell\}$ such that the intervals $(a_s, b_s) \cap I$ for $s \in S$ still cover I.

Among the open intervals (a_s, b_s), $s \in S$, there is one interval that contains 0, without loss of generality the interval (a_{s_1}, b_{s_1}). If $b_{s_1} \le 1$, then there is another such interval that contains the point b_{s_1}, without loss of generality the interval (a_{s_2}, b_{s_2}), and so on. Hence we can assume that the intervals (a_{s_i}, b_{s_i}) for $i = 1, \dots, k$ for $k \le \ell$ cover I and $b_{s_i} \in (a_{s_{i+1}}, b_{s_{i+1}})$ as shown in the picture below. In particular, $(a_{s_i}, b_{s_i}) \cap (a_{s_{i+1}}, b_{s_{i+1}}) = (a_{s_{i+1}}, b_{s_i})$ intersect nontrivially, and we can choose a point t_i in that intersection.



We note that the interval $[t_{i-1}, t_i]$ is contained in the interval (a_{s_i}, b_{s_i}) which is contained in the preimage $f^{-1}(U_{s_i})$. Hence $f([t_{i-1}, t_i])$ is contained in the evenly covered subset $U_{s_i} \subset X$ which proves the claim.

Construction of a lift $\widetilde{f}\colon I \to \widetilde{X}$ with $\widetilde{f}(0) = \widetilde{f}_0$. The construction is inductive. Assume that we have already constructed a lift $\widetilde{f}\colon [0, t_{i-1}] \to \widetilde{X}$ of the path $f\colon [0, t_{i-1}] \to X$ with starting point $\widetilde{f}(0) = \widetilde{f}_0 \in \widetilde{X}$. Then the path $f\colon [t_{i-1}, t_i] \to X$ is contained in an evenly covered subset of X and hence has a lift $\widetilde{f}\colon [t_{i-1}, t_i] \to \widetilde{X}$ with any prescribed starting point in $\widetilde{f}(t_{i-}) \in p^{-1}(f(t_{i-1}))$ as described at the beginning of the proof. In particular, we can choose a lift with starting point whose starting point is the endpoint of the previously constructed lift $\widetilde{f}\colon [0, t_{i-1}] \to \widetilde{X}$. This guarantees that these two lifts fit together to form a lift (lifts are required to be continuous!) $\widetilde{f}\colon [0, t_i] \to \widetilde{X}$ with $\widetilde{f}(0) = \widetilde{f}_0$. This proves the existence of a lift.

Uniqueness of the lift $\widetilde{f}\colon I \to \widetilde{X}$ with $\widetilde{f}(0) = \widetilde{f}_0$. Assume that $\widetilde{f}$, $\widetilde{f}'$ are two lifts of $f\colon [0,1] \to X$ with the same starting point $\widetilde{f}_0 \in \widetilde{X}$. Assume $\widetilde{f}(t) \neq \widetilde{f}'(t)$ for some $t \in I$. Using the partition of I given by the Claim above, let i be the smallest number such that $\widetilde{f}(t_{i-1}) = \widetilde{f}'(t_{i-1})$, but $\widetilde{f}(t) \neq \widetilde{f}'(t)$ for some $t \in (t_{i-1}, t_i]$. By construction, $f([t_{i-1}, t_i])$ is contained in an evenly covered subset $U \subset X$, and hence $\widehat{f}\colon [t_{i-1}, t_i] \to \widetilde{X}$ and $\widehat{f}'\colon [t_{i-1}, t_i] \to \widetilde{X}$ are paths in

$$p^{-1}(U) = \bigcup_{a \in A} U_a.$$

Assume that the common starting point $\widetilde{f}(t_{i-1}) = \widetilde{f}'(t_{i-1})$ belongs to U_{a_0}. Since $\widetilde{f}(t) \neq \widetilde{f}'(t)$ for some $t \in (t_{i-1}, t_i]$, one of these two points can not be a point of the sheet U_{a_0}, say $\widehat{f}(t) \neq U_{a_0}$. This is not possible, since otherwise $\{s \in [t_{i-1}, t_i] \mid \widetilde{f}(s) \in U_{a_0}\}$ and $\{s \in [t_{i-1}, t_i] \mid \widetilde{f}(s) \notin U_{a_0}\}$ would be two disjoint, non-emtpy, open subspaces of $[t_{i-1}, t_i]$ whose union is this interval. This contradict the fact that intervals are connected.

Uniqueness of the lift $\widetilde{f}\colon Y \times I \to \widetilde{X}$ extending $\widetilde{f}_0\colon Y \times \{0\} \to \widetilde{X}$. For each $y \in Y$ the map $I \ni s \mapsto \widetilde{f}(y, s) \in \widetilde{X}$ is a lift with starting point $\widetilde{f}_0(y)$ of the path $s \mapsto f(y, s) \in X$. Hence the uniqueness of $\widetilde{f}$ is an immediate consequence of the uniqueness of pathlifting with given starting point.

Construction of a lift $\widetilde{f}\colon N \times I \to \widetilde{X}$ for a neighborhood N of a point $y_0 \in Y$.

We fix a point $y_0 \in Y$. Our strategy for constructing the lift $\widetilde{f}\colon N \times I \to \widetilde{X}$ is a modification of the method we used for the construction of the lifted path in the case $Y = \mathrm{pt}$: we find a partition $0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1$ of the interval I and an open neighborhood N of $y \in Y$ such that $f(N \times [t_{i-1}, t_i])$ is contained in some evenly covered subset of X for each $i = 1, \dots, k$. Then we use induction over i to construct a lift on $N \times [0, t_i]$ for $i = 1, \dots, k$.

To construct the partition $0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1$ we choose for every point $s \in I$ an evenly covered open neighborhood $U_s \subset X$ of the point $f(y_0, s) \in X$. Then $f^{-1}(U_s) \subset Y \times I$ is an open neighborhood of $(y_0, s) \in Y \times I$. Hence there are open neighborhoods

$$y_0 \in N_s \subset Y \qquad \text{and} \qquad s \in (a_s, b_s) \cap I \subset I$$

such that the product $N_s \times (a_s, b_s) \cap I$ is contained in $f^{-1}(U_s)$. The collection $\{(a_s, b_s) \cap I \mid s \in I\}$ is an open cover of I, and hence by compactness of I there is a finite subset

$S = \{s_1, \dots, s_\ell\} \subset I$ such that the intervals $(a_s, b_s) \cap I$ for $s \in S$ cover I. Then as before, there is a partition $0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1$ of I such that $[t_{i-1}, t_i] \subset (a_{s_i}, b_{s_i}) \cap I$. It follows that for each $i = 1, \dots, k$

$$f(N \times [t_{i-1}, t_i]) \subset f(N \times (a_{s_i}, b_{s_i}) \cap I) \subset U_{s_i} \qquad \text{for } N := \bigcap_{s \in S} N_s. \tag{5.3}$$

Now we can construct a lift $\widetilde{f}\colon N \times [0, t_i]$ of $f\colon N \times [0, t_i]$ by induction over i. Suppose we have already constructed a lift $\widetilde{f}\colon N \times [0, t_{i-1}]$ that restricts to $\widetilde{f}_0$ on $N \times \{0\}$. By (5.3) the image $f(N \times [t_{i-1}, t_i])$ is contained in the evenly covered subset $U := U_{s_i} \subset X$. Hence the image of the already constructed lift $\widetilde{f}_{|N \times t_{i-1}}\colon N \times t_{i-1} \to \widetilde{X}$ is contained in

$$p^{-1}(U) = \bigcup_{a \in A} U_a,$$

the union of the disjoint open subsets $U_a \subset \widetilde{X}$. Replacing N by the connected component of N that contains y_0, we may assume that N is connected. Hence the image of $\widetilde{f}_{|N \times t_{i-1}}$ is contained is contained in U_a for some $a \in A$.

Then it is easy to define the lift $\widetilde{f}$ on $N \times [t_{i-1}, t_i]$ as before by

$$\widetilde{f}(y, s) := p_{|U_a}^{-1}(f(y, s)) \qquad \text{for } y \in N,\ s \in [t_{i-1}, t_i].$$

This map agrees with the previously constructed lift $\widetilde{f}\colon N \times [0, t_{i-1}] \to \widetilde{X}$ on $N \times \{t_{i-1}\}$, and hence they fit together to define a lift $\widetilde{f}\colon N \times [0, t_i]$ as claimed.

Construction of the lift $\widetilde{f}\colon Y \times I \to \widetilde{X}$. In our construction above, we fixed the point y_0 and then found a lift $\widetilde{f}$ on $N \times I$ where N is some open neighborhood of y_0. This can be done for *any point* y_0, but we should write N_{y_0} instead of N and $\widetilde{f}_{y_0}$ instead of $\widetilde{f}$ to indicate that these data depend on the point y_0. Now take two points $y_0, y_1 \in Y$, assume $N_{y_0} \cap N_{y_1} \neq \emptyset$, and let us compare

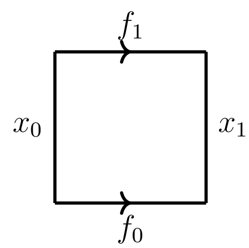
$$\widetilde{f}_{y_0}(y, s) \quad \text{and} \quad \widetilde{f}_{y_1}(y, s) \qquad \text{for } y \in N_{y_0} \cap N_{y_1} \text{ and } s \in I.$$

We observe that for fixed $y \in N_{y_0} \cap N_{y_1}$ these are paths in $\widetilde{X}$ which are lifts of the path $s \mapsto f(y, s)$ in X, both starting at the point $\widetilde{f}_0(y) \in \widetilde{X}$. Hence these paths agree by the uniqueness of pathlifting. This shows that the lifts $\widetilde{f}_{y_0}$, $\widetilde{f}_{y_1}$ agree on their common domain $(N_{y_0} \cap N_{y_1}) \times I$. Hence they fit together to give a lift on $(N_{y_0} \cup N_{y_1}) \times I$. Since this can be done for any point $y_0 \in Y$, we can construct a lift $\widetilde{f}\colon Y \times I \to \widetilde{X}$ which restricts to the given map $\widetilde{f}_0\colon Y \times \{0\} \to \widetilde{X}$. $\square$

5.2 Applications of the homotopy lifting property

First we will apply the homotopy lifting property for covering maps in the case of homotopies between paths in X. In other words, the homotopy is a map $f\colon I \times I \to X$, which gives a homotopy from the path f_0 to the path f_1. Here $f_t\colon I \to X$ for any $t \in I$ is the path defined by $f_t(s) := H(s,t)$.

In this case we are in particular interested in *path homotopies* or *homotopies relative* ∂I, which means that the paths $f_t\colon I \to X$ for $t \in I$ all have the same starting point x_0 and the same endpoint x_1. In other words, the map $f\colon I \times I \to X$ maps the left vertical edge of the square $I \times I$ to x_0 and the right vertical edge to x_1. The restriction of f to the lower (resp. upper) edge gives the path f_0 (resp. f_1). We express this in a picture by labeling the edges of $I \times I$ appropriately as follows:



$f\colon I \times I \to X$

Lemma 5.4. *Let $p\colon \widetilde{X} \to X$ be a covering map and let $f\colon I \times I \to X$ be a path homotopy. Let $\widetilde{x}_0 \in p^{-1}(x_0)$, where $x_0 \in X$ is the common starting point of the paths f_t. Let $\widetilde{f}_0\colon I \to \widetilde{X}$ be the unique lift of f_0 with $\widetilde{f}_0(0) = \widetilde{x}_0$, and let $\widetilde{f}\colon I \times I \to \widetilde{X}$ be the unique lift of f such that $\widetilde{f}(s,0) = \widetilde{f}_0(s)$. Then $\widetilde{f}$ is a path homotopy.*

Proof. We need to check that $\widetilde{\ }\, f\colon I \times I \to \widetilde{X}$ restricts on the vertical edges to *constant* maps. The homotopy $f\colon I \times I \to X$ is by assumption a path homotopy, i.e., it maps the left (resp. right) vertical edge points $x_0 \in X$ (resp. $x_1 \in X$). Hence $t \mapsto \widetilde{f}(0,t)$ is a lift of the constant path $c_{x_0}\colon I \to X$. But any lift $\widetilde{\gamma}$ of a constant path must be constant, (the constant path $c_{\widetilde{x}_0}$ starting at any point $\widehat{x}_0 \in p^{-1}(x_0)$ clearly is lift, and hence by the uniqueness of path lifting, $c_{\widetilde{x}_0}$ is the only lift with that starting point). This shows that $\widetilde{f}$ is in fact a path homotopy. $\square$

Corollary 5.5. *Let $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ be a covering map. Then the induced homomorphism*

$$p_*\colon \pi_1(\widetilde{X}, \widetilde{x}_0) \longrightarrow \pi_1(X, x_0)$$

is injective.

Proof. Let $\widetilde{\gamma}_0, \widetilde{\gamma}_1$ be based loops in $(\widetilde{X}, \widetilde{x}_0)$ and assume that $p_*([\widetilde{\gamma}]) = p_*([\widetilde{\gamma}_1]) \in \pi_1(X, x_0)$. In other words, there is a path homotopy $f\colon I \times I \to X$ with $f_0 = p \circ \widetilde{\gamma}_0$ and $f_1 = p \circ \widetilde{\gamma}_1$.

By the homotopy lifting property there is a unique lift $\widetilde{f}\colon I\times I\to\widetilde{X}$ with $\widetilde{f}_0=\widetilde{\gamma}_0$. By the lemma, $\widetilde{f}$ is again a path homotopy from $\widetilde{f}_0=\widetilde{\gamma}_0$ to $\widetilde{f}_1$.

Both paths, $\widetilde{f}_1$ and $\widetilde{\gamma}_1$ are lifts of the path $f_1=\partial\circ\widetilde{\gamma}_1$ with the same starting point, since $\widetilde{f}_1(0)=\widetilde{f}_0(0)$ (since $\widetilde{f}$ is a path homotopy), and $\widetilde{f}_0(0)=\widetilde{\gamma}_0(0)=\widetilde{x}_0=\widetilde{\gamma}_1(0)$. Hence by the uniqueness of path lifting, this implies $\widetilde{f}_1=\widetilde{\gamma}_1$, and so $\widetilde{f}$ is a indeed a path homotopy from $\widetilde{\gamma}_0$ to $\widetilde{\gamma}_1$, proving that $[\widetilde{\gamma}_0]=[\widetilde{\gamma}_1]\in\pi_1(\widetilde{X},\widetilde{x}_0)$. $\square$

Corollary 5.6. *Let γ_0, γ_1 be two paths in X with starting point $x_0\in X$ and endpoint $x_1\in X$. Let $\widetilde{x}_0\in p^{-1}(x_0)$ and let $\widetilde{\gamma}_0,\widetilde{\gamma}_1\colon I\to\widetilde{X}$ be lifts of γ_0 resp. γ_1 with starting point $\widetilde{x}_0\in\widetilde{X}$. If γ_0 and γ_1 are path homotopic, then $\widetilde{\gamma}_0(1)=\widetilde{\gamma}_1(1)$, i.e., the endpoints of the lifts $\widetilde{\gamma}_0$ and $\widetilde{\gamma}_1$ agree.*

In the special case of the covering map $p\colon\mathbb{R}\to S^1$, we proved this statement to show that the winding number of a based loop $\gamma\colon(I,\partial I)\to S^1$ (which is defined as the endpoint $\widetilde{\gamma}(1)\in\mathbb{Z}$ of a lift $\widetilde{\gamma}$ to γ with starting point $\widetilde{\gamma}(0)=0$) depends only the homotopy class $[\gamma]\in\pi_1(S^1,1)$.

Proof. Let $f\colon I\times I\to X$ be path homotopy from γ_0 to γ_1, and let $\widetilde{f}\colon I\times I\to\widetilde{X}$ be the unique lift of f with $\widetilde{f}_0=\widetilde{\gamma}_0$. By the lemma $\widetilde{f}$ is then a path homotopy. In particular, $\widetilde{f}_1(0)=\widetilde{f}_0(0)=\widetilde{\gamma}_0(0)=\widetilde{x}_0$, and hence $\widetilde{f}_1$ is a lift of γ_1 with starting point $\widetilde{x}_0$. Since $\widetilde{\gamma}_1$ is also a lift of γ_1 with $\widetilde{\gamma}_1(0)=\widetilde{x}_0$, uniqueness of path lifting implies $\widetilde{f}_1=\widetilde{\gamma}_1$. It follows that $\widetilde{f}$ is path homotopy from $\widetilde{\gamma}_0$ to $\widetilde{\gamma}_1$ and hence in particular, $\widetilde{\gamma}_0(1)=\widetilde{\gamma}_1(1)$. $\square$

Corollary 5.7. *Let γ be a based loop in (X,x_0), and let $\widetilde{\gamma}\colon I\to\widetilde{X}$ be the lift of γ with starting point $\widetilde{\gamma}(0)=\widetilde{x}_0$. Then $\widetilde{\gamma}$ is a based loop in $(\widetilde{X},\widetilde{x}_0)$ if and only if $[\gamma]$ belongs to the subgroup $p_*\pi_1(\widetilde{X},\widetilde{x}_0)$ of $\pi_1(X_0,x_0)$.*

Proof. If $\widetilde{\gamma}$ is a based loop in $(\widetilde{X},\widetilde{x}_0)$, then $[\gamma]=[p\circ\widetilde{\gamma}]=p_*[\widetilde{\gamma}]$ belongs to $p_*\pi_1(\widetilde{X},\widetilde{x}_0)$. Conversely, if $[\gamma]$ is an element of $p_*\pi_1(\widetilde{X},\widetilde{x}_0)$, then γ is path homotopic to $p\circ\widetilde{\gamma}'$ for some based loop $\widetilde{\gamma}'$ in $(\widetilde{X},\widetilde{x}_0)$. The paths $\widetilde{\gamma}$ and $\widetilde{\gamma}'$ have the same starting point, and hence by the previous corollary, the path homotopy between $\gamma=p\circ\widetilde{\gamma}$ and $p\circ\widetilde{\gamma}'$ implies that the endpoints of $\widetilde{\gamma}$ and $\widetilde{\gamma}'$ agree, and hence $\widetilde{\gamma}(1)=\widetilde{\gamma}'(1)=\widetilde{x}_0$. $\square$

We recall that the winding number of a based loop γ in $(S^1,1)$ is defined by $W(\gamma):=\widetilde{\gamma}(1)\in\mathbb{Z}$ where $\widetilde{\gamma}\colon I\to\mathbb{R}$ is a path starting at the base point $0\in\mathbb{R}$ which is the unique lift of γ for the covering map $p\colon\mathbb{R}\to S^1$ defined by $p(s)=e^{2\pi s}$. We note that $p^{-1}(1)$, the fiber over the basepoint $1\in S^1$, is equal to $\mathbb{Z}$.

This construction generalizes to covering maps $p\colon(\widetilde{X},\widetilde{x}_0)\to(X,x_0)$: if γ is a based loop in (X,x_0), we define

$$W(\gamma):=\widetilde{\gamma}(1)\in p^{-1}(x_0),$$

for the unique lift $\widetilde{\gamma}\colon I \to \widetilde{X}$ with $\widetilde{\gamma}(0) = \widetilde{x}_0$. A crucial property of the winding number $W(\gamma)$ is that it depends only on the homotopy class $[\gamma] \in \pi_1(S^1, 1)$. It turns out that in the case of a general covering map, a stronger statement holds: $W(\gamma)$ depends only on the right coset $H[\gamma] \in H\backslash G := \{Hg \mid g \in G\}$, where $G = \pi_1(X, x_0)$ and $H \subset G$ is the subgroup $H := p_*\pi_1(\widetilde{X}, \widetilde{x}_0)$.

Proposition 5.8. *Let $\widetilde{X}$ be path-connected and let $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ be a covering map. Then the map*

$$\Psi\colon H\backslash G \longrightarrow p^{-1}(x_0) \qquad \text{given by } \Psi(H[\gamma]) := W(\gamma) \text{ as defined above}$$

is a well-defined bijection. In particular, the number of sheets of $\widetilde{X} \to X$ (the cardinality of the fiber $p^{-1}(x_0)$) is equal to the index $[G : H]$ of the subgroup $H \subset G$ (the cardinality of $H\backslash G$).

Proof. To show that Ψ is well-defined, let $g \in G = \pi_1(X, x_0)$ be represented by a based loop γ in (X, x_0) and let $h \in H = p_*\pi_1(\widetilde{X}, \widetilde{x}_0)$ be represented by a based loop δ in (X, x_0). We let $\widetilde{\gamma}, \widetilde{\delta}\colon I \to \widetilde{X}$ be the unique lifts of γ resp. δ starting at $\widetilde{x}_0$. Since $[\delta]$ is in $p_*\pi_1(\widetilde{X}, \widetilde{x}_0)$ its lift $\widetilde{\delta}$ is a based loop in $(\widetilde{X}, \widetilde{x}_0)$ by Corollary 5.7. It follows the endpoint of $\widetilde{\delta}$ is the starting point of $\widetilde{\gamma}$ and hence the concatenation $\widetilde{\delta} * \widetilde{\gamma}$ is defined. Moreover,

$$p \circ (\widetilde{\delta} * \widetilde{\gamma}) = (p \circ \widetilde{\delta}) * (p \circ \widetilde{\gamma}) = \delta * \gamma,$$

and hence $\widetilde{\delta} * \widetilde{\gamma}$ is the unique lift of $\delta * \gamma$ with starting point $\widetilde{x}_0$. This implies that

$$\Psi(hg) = \Psi([\delta][\gamma]) = \Psi([\delta * \gamma]) = W(\delta * \gamma) = (\widetilde{\delta} * \widetilde{\gamma})(1) = \widetilde{\gamma}(1) = W(\gamma) = \Psi([\gamma]) = \Psi(g)$$

and hence W is well-defined.

It is clear that W is surjective, since due to the assumption that $\widetilde{X}$ is path-connected for any $\widetilde{x} \in p^{-1}(x_0)$ there is a path $\widetilde{\gamma}\colon I \to \widetilde{X}$ from $\widetilde{x}_0$ to x. Then $\gamma := p \circ \widetilde{\gamma}$ is a loop in (X, x_0), and hence $\Psi([\gamma]) = W(\gamma) = \widetilde{\gamma}(1) = \widetilde{x}$.

To see that Ψ is injective, let $g_1, g_2 \in \pi_1(X, x_0)$ be elements with $\Psi(g_1) = \Psi(g_2)$. Let γ_i be a based loop in (X, x_0) representing g_i and let $\widetilde{\gamma}_i\colon I \to \widetilde{X}$ be its unique lift with $\widetilde{\gamma}_i(0) = \widetilde{x}_0$. Then $\Psi(g_i) = \widetilde{\gamma}_i(1)$, and hence our assumption $\Psi(g_1) = \Psi(g_2)$ implies that the paths $\widetilde{\gamma}_1$, $\widetilde{\gamma}_1$ have the same endpoints. In particular, the concatenation $\widetilde{\gamma}_1 * \overline{\widetilde{\gamma}}_2$ is defined and is a based loop in $(\widetilde{X}, \widetilde{x}_0)$. It follows that

$$p_*[\widetilde{\gamma}_1 * \overline{\widetilde{\gamma}}_2] = [p \circ (\widetilde{\gamma}_1 * \overline{\widetilde{\gamma}}_2)] = [\gamma_1 * \bar{\gamma}_2] = [\gamma_1][\gamma_2]^{-1} = g_1 g_2^{-1},$$

showing that $g_1 g_2^{-1} \in H$ and hence the left cosets Hg_1 and Hg_2 agree. $\square$

Let K be a finite group that acts freely on a space Y. Then the projection $p\colon Y \to Y/K$ onto the orbit space Y/K is a covering space (this is a homework problem). To use the proposition above, let us set $\widetilde{X} := Y$ and $X := Y/K$,and let us assume that $\widetilde{X}$ is *simply connected*, i.e., $\widetilde{X}$ is path-connected and $\pi_1(\widetilde{X}, \widetilde{x}_0)$ vanishes (since $\widetilde{X}$ is path connected, up to isomorphism the fundamental group of $\widetilde{X}$ is independent of the basepoint we use). Then the proposition yields a bijection

$$\Psi\colon G \xrightarrow{\cong} p^{-1}(x_0)$$

We note that the fiber $p^{-1}(x_0)$ is the orbit $K\widetilde{x}_0 = \{k\widetilde{x}_0 \mid k \in K\}$ of the basepoint $\widetilde{x}_0 \in \widetilde{X}$. The assumption that K acts freely implies that the map $K \to p^{-1}(x_0)$, $k \mapsto k\widetilde{x}_0$ is a bijection.

Lemma 5.9. *The composition of the bijections*

$$\Psi\colon G \to p^{-1}(x_0) \qquad \text{and} \qquad p^{-1}(x_0) \to K, \quad k\widetilde{x}_0 \mapsto k$$

is a group isomorphism.

Proof. Let γ_1 and γ_2 be based loops in (X, x_0) and let $\widetilde{\gamma}_i\colon I \to \widetilde{X}$ be the lift of γ_i starting at $\widetilde{x}_0$. Let $k_i \in K$ be the image of $[\gamma_i] \in G$ under the bijection above, i.e., $\widetilde{\gamma}_i(1) = k_i\widetilde{x}_0 \in p^{-1}(x_0)$. To determine $G([\gamma_1][\gamma_2]) = G([\gamma_1 * \gamma_2])$, we need to find the unique lift $\widetilde{\gamma_1 * \gamma_2}$ of $\gamma_1 * \gamma_2$.

A first guess might be $\widetilde{\gamma_1 * \gamma_2} = \widetilde{\gamma}_1 * \widetilde{\gamma}_2$, but that concatenation does not make sense in general, since $\widetilde{\gamma}_1(1) = k_1\widetilde{x}_0$, while $\widetilde{\gamma}_2(0) = \widetilde{x}_0$. However, this suggests to use the action of $k_1 \in K$ to obtain the path $k_1\gamma_2\colon I \to \widetilde{X}$, defined by $(k_1\gamma_2)(s) = k_1\gamma_2(s)$. It has starting point $k_1\gamma_2(0) = k_1\widetilde{x}_0$, and hence can be concatenated with $\widetilde{\gamma}_1$. Hence

$$\widetilde{\gamma_1 * \gamma_2} = \widetilde{\gamma}_1 * k_1\widetilde{\gamma}_2,$$

with endpoint $\widetilde{\gamma_1 * \gamma_2}(1) = k_1\widetilde{\gamma}_2(1) = k_1k_2\boldsymbol{\widetilde{x}}_0$. This shows that the image of $[\gamma_1 * \gamma_2]$ under the composite bijection is k_1k_2, and hence this composition is a group isomorphism. $\square$

Example: calculation of the fundamental group of $\mathbb{RP}^n$ and the lens space $\text{Ł}^{2n-1}(\mathbb{Z}/k)$

Give a covering space $p\colon \widetilde{X} \to X$, it will be important for us to lift not just homotopies, but more general maps $f\colon Y \to X$. In other words, we are looking for a base point preserving map $\widetilde{f}$ making the diagram

$$\begin{tikzcd} & (\widetilde{X}, \widetilde{x}_0) \arrow[d, "p"] \\ (Y, y_0) \arrow[ur, dashed, "\widetilde{f}"] \arrow[r, "f"] & (X, x_0) \end{tikzcd} \tag{5.10}$$

commutative.

There is an obvious necessary condition for the existence of a lift $\widetilde{f}$: such a lift induces a commutative diagram of fundamental groups

$$\begin{tikzcd} & \pi_1(\widetilde{X}, \widetilde{x}_0) \arrow[d, "p_*"] \\ \pi_1(Y, y_0) \arrow[ur, dashed, "\widetilde{f}_*"] \arrow[r, "f_*"] & \pi_1(X, x_0) \end{tikzcd}, \tag{5.11}$$

and hence $f_*\pi_1(Y, y_0)$ is contained in $p_*\pi_1(\widetilde{X}, \widetilde{x}_0)$. We will show that this is also a sufficient condition for the existence of a lift $\widetilde{f}$, provided the topological space Y isn't too crazy.

Definition 5.12. A topological space Y is *locally path-connected* if for any point $y \in Y$ and any neighborhood U of y there is an open neighborhood $V \subset U$ which is path-connected. More generally, if P is any property of a topological space (e.g., compact, connected,...), then Y is *locally* P if for any point $y \in Y$ and any neighborhood U of y there is an open neighborhood $V \subset U$ such that V has property P.

Example 5.13. (Path-connected versus locally path-connected). There are many examples of spaces which are locally path-connected, but non path-connected, for example the disjoint union of path-connected spaces is locally path-connected. An example of a space which is not locally path-connected is provided by the *topologist's sine curve* (**??**), consisting of the union of the graph of the function $(0, \infty) \to \mathbb{R}$, $x \mapsto \sin(1/x)$ and the vertical line segment $\{0\} \times [-1, +1]$. As discussed then, the topologist's sine curve is connected, but not path-connected. The same argument shows in fact that any open neighborhood V of a point y on the vertical line segment is not path-connected (since it always contains points on the graph of $\sin(1/x)$; those cannot be reached by paths starting at y).

Even more interesting is that there are spaces which is path-connected, but not locally path-connected, for example, the *Warsaw circle*. This is a variant of the topologist's sine curve obtained by restricting the graph of $\sin(1/x)$ to some finite interval $(0, a)$ and connecting the point $(a, \sin(1/a))$ on the graph with the point $(0, 0)$ via an arc in $\mathbb{R}^2$ which intersects the topologist's sine curve only in those two points as shown in the figure below.

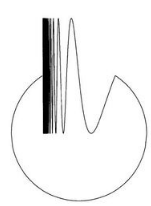


Figure 1: The Warsaw circle

The Warsaw circle is path-connected, since any point on the vertical line can be connected via a path running along the added arc to any point of graph. Adding that arc does not change the fact that any open neighborhood of a point on the vertical line segment is not path-connected, provided it is small enough that it doesn't contain the arc. In particular, the Warsaw circle is not locally path-connected.

Proposition 5.14. (General Lifting Criterion). *Let* $p\colon (\widetilde{X},\widetilde{x}_0) \to (X,x_0)$ *be a covering map and let* $f\colon (Y,y_0) \to (X,x_0)$ *be a basepoint preserving map whose domain* Y *is path-connected and locally path-connected. Then a lift* $\widetilde{f}$ *in the diagram* (5.11) *exists if and only if* $f_*\pi_1(Y,y_0) \subset p_*\pi_1(\widetilde{X},\widetilde{x}_0)$. *There is at most one such lift.*

Remark 5.15. The hypothesis that Y is locally path-connected cannot be dropped. This follows by showing that the statement above does not hold if Y is the Warsaw circle W, which is proved by showing

(a) The fundamental group of W vanishes, but

(b) the map $f\colon W \to S^1$ which wraps the Warsaw circle once around the circle S^1 does not have a lift $\widetilde{f}\colon W \to \mathbb{R}$.

Proof. We have argued above that $f_*\pi_1(Y,y_0) \subset \pi_1(\widetilde{X},\widetilde{x}_0)$ is a necessary condition for the existence of a lift $\widetilde{f}$. It is also easy to see that there is at most one such lift: if $\widetilde{f}$, $\widetilde{f}'$ are two lifts of f and $y \in Y$, let $\gamma\colon I \to Y$ be a path in Y from y_0 to y. Then $\widetilde{f}\circ\gamma$ and $\widetilde{f}'\circ\gamma$ are two paths in $\widetilde{X}$ which are both lifts of the path $f\circ\gamma$ in X with starting point $\widetilde{x}_0$. The uniqueness of lifted paths then implies $\widetilde{f}(y) = \widetilde{f}\circ\gamma(1) = \widetilde{f}'\circ\gamma(1) = \widetilde{f}'(y)$.

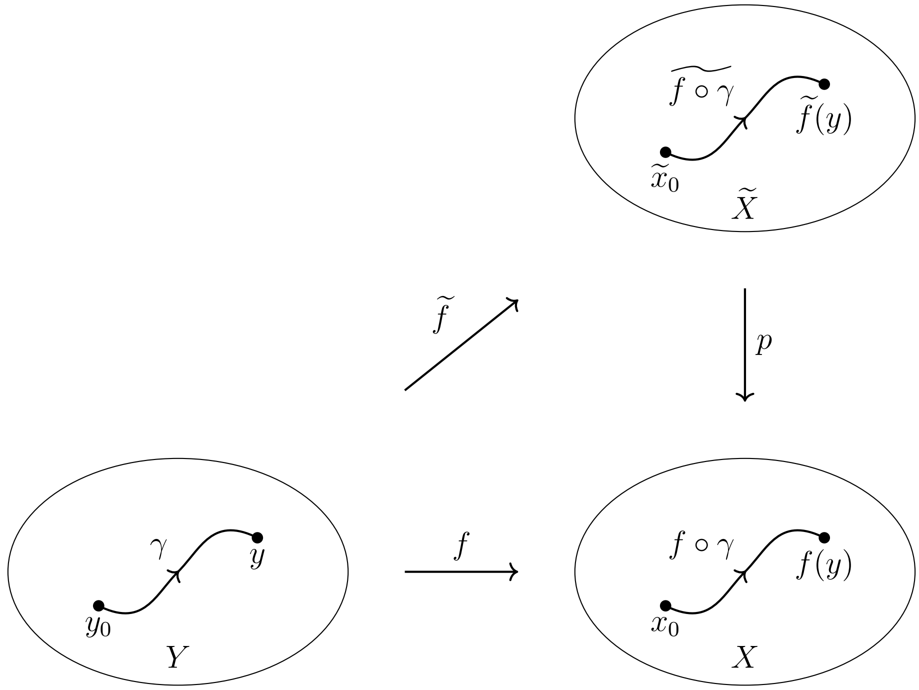
The idea for *constructing* the lift $\widetilde{f}\colon Y \to \widetilde{X}$ is to use the *existence* of lifts of paths, similar to the way we used *uniqueness of path-lifting* to prove the *uniqueness of* $\widetilde{f}$: we define the map

$$\widetilde{f}\colon Y \longrightarrow \widetilde{X} \qquad \text{by} \qquad \widetilde{f}(y) := (\widetilde{f\circ\gamma})(1)$$

where $\gamma\colon I \to Y$ is path from y_0 to y, and

$$\widetilde{f\circ\gamma}\colon I \to \widetilde{X} \text{ is a lift of } f\circ\gamma\colon I \to X \text{ with starting point } \widetilde{x}_0.$$

The following figure illustrates the various paths involved and their endpoints.



To show that $\widetilde{f}$ is well-defined, we need to verify that $\widetilde{f}(y)$ is independent of the choice of the path γ from the basepoint y_0 to y. So suppose that $\gamma' \colon I \to Y$ is another path from y_0 to y. The concatenation $\gamma * \bar{\gamma}'$ is then a based loop in (Y, y_0) and hence the loop $f \circ (\gamma * \bar{\gamma}') = (f \circ \gamma) * (f \circ \bar{\gamma}')$ is a based loop in (X, x_0) which represents an element in the image of $f_* \colon \pi_1(Y, y_0) \to \pi_1(X, x_0)$. By assumption, this element is then in the image of $p_* \colon \pi_1(\widetilde{X}, \widetilde{x}_0) \to \pi_1(X, x_0)$, which by Corollary 5.5 implies that the loop $(f \circ \gamma) * (f \circ \bar{\gamma}')$ lifts to a based loop $\widetilde{\delta} \colon I \to \widetilde{X}$ in $(\widetilde{X}, \widetilde{x}_0)$. By uniqueness of lifted paths, the first half of $\widetilde{\delta}$ is $\widetilde{f \circ \gamma}$ and the second half is $\widetilde{f \circ \gamma'}$ traversed backwards, with the common midpoint $\widetilde{f \circ \gamma}(1) = \widetilde{f \circ \gamma'}(1)$. This shows that $\widetilde{f}$ is well-defined.

To prove that $\widetilde{f}$ is continuous, it suffices to show that the restriction of $\widetilde{f}$ to a suitable open neighborhood of $y_1 \in Y$ is continuous. A convenient choice in this context is to choose $f^{-1}(U)$ where U is an evenly covered neighborhood of $f(y_1) \in X$. Using the assumption that X is locally path-connected we can pass to a smaller open neighborhood $V \subset f^{-1}(U)$ of y_1, which is path-connected.

Our goal is to show that $\widetilde{f}_{|V}$ is continuous. To calculate $\widetilde{f}(y)$ for $y \in V$, we choose a path γ_1 from y_0 to y_1, and we choose a path α_y from y_1 to y which is contained in the pathconnected neighborhood V of y_1. Then $\gamma_1 * \alpha_y$ is a path from y_0 to y, which we can use to calculate $\widetilde{f}(y)$ as the endpoint of a lift of the path

$$f \circ (\gamma_1 * \alpha_y) = (f \circ \gamma_1) * (f \circ \alpha_y). \tag{5.16}$$

So let $\widetilde{f\circ\gamma_1}\colon I\to\widetilde{X}$ be the unique lift with starting point $\widetilde{x}_0$ of the path $f\circ\gamma_1\colon I\to X$. Then by construction $\widetilde{f}(y_1)=\widetilde{f\circ\gamma_1}(1)$. Let $\widetilde{f\circ\alpha_y}\colon I\to\widetilde{X}$ be the unique lift of $f\circ\alpha_y$ with starting point $\widetilde{f}(y_1)$. Then the concatenation $\widetilde{f\circ\gamma_1}*\widetilde{f\circ\alpha_y}$ is the unique lift with starting point $\widetilde{x}_0$ of the path (5.16). Again by construction $\widetilde{f}(y)$ is the endpoint of this concatenated path and hence

$$\widetilde{f}(y)=\widetilde{f\circ\alpha_y}(1).$$

At first glance, it seems difficult to argue that $\widetilde{f}(y)$ depends continuously on y, since the y-dependence of the right hand side comes from the y-dependence of the path α_y, which seems hard to control since it involves the choice of a path.

To the rescue comes the fact that the path $f\circ\alpha_y$ is contained in the evenly covered subset $U\subset X$. Hence the lift $\widetilde{f\circ\alpha_y}\colon I\to\widetilde{X}$ is contained in $p^{-1}(U)$ which is the union of disjoint open subsets U_a, $a\in A$ since U is evenly covered. Since I is connected, its image under the map $\widetilde{f\circ\alpha_y}$ must be contained in U_a for some $a\in A$. Since the restriction $p_{U_a}\colon U_a\to U$ is a homeomorphism, it follows that

$$\widetilde{f\circ\alpha_y}=p^{-1}_{|U_a}\circ f\circ\alpha_y.$$

Hence,

$$\widetilde{f}(y)=\widetilde{f\circ\alpha_y}(1)=p^{-1}_{|U_a}\circ f\circ\alpha_y(1)=p^{-1}_{|U_a}\circ f(y)\qquad\text{for } y\in V.$$

This shows that $\widetilde{f}_{|V}=p^{-1}_{|U_a}\circ f$ is the composition of continuous maps and hence continuous. Since f is continuous in some open neighborhood of every point of Y, this implies that $\widetilde{f}$ is continuous. □

5.3 Classification of covering spaces

The goal of this section is to classify covering spaces. More precisely, we consider the covering spaces $p\colon\widetilde{X}\to X$ of a fixed topological space X as objects of a category $\mathrm{Cov}(X)$, and aim to understand this category by

- determining the isomorphism classes of objects of $\mathrm{Cov}(X)$, and
- determining the set of morphisms between two covering spaces of X.

To illustrate what we have in mind, we consider the category $\mathsf{Vect}_K^{\mathrm{fin}}$ of finite dimensional vector spaces over K and linear maps. Then

(i) Mapping a finite dimensional vector space V to its dimension $\dim_K V\in\mathbb{N}_0$ (the natural numbers including 0) gives a bijection

$$\mathrm{ob}(\mathsf{Vect}_K^{\mathrm{fin}})/\mathrm{isom}\ \xrightarrow{\cong}\ \mathbb{N}_0$$

between the isomorphism classes of objects of $\mathsf{Vect}_K^{\mathrm{fin}}$ and $\mathbb{N}_0$.

(ii) For vector spaces V, W of dimension m resp. n, there is a bijection

$$\operatorname{mor}(V, W) \xrightarrow{\cong} M_{n\times m}(K)$$

between the morphisms from V to W and the set of $n \times m$-matrices with coefficients in K.

A covering space over X restricts to a covering space over each path-connected component of X, and so it suffices to analyze covering spaces over path-connected spaces. Similarly, if $p\colon \widetilde{X} \to X$ is a covering space over a path-connected space X, then $p_{|\widetilde{X}_i}\colon \widetilde{X}_i \to X$ is a covering space for each path-connected component of $\widehat{X}_i$ of $\widetilde{X}$. Hence it suffices to classifying covering spaces $p\colon \widetilde{X} \to X$ assuming that X and $\widetilde{X}$ are path-connected.

Definition 5.17. Let X be a path-connected topological space. Let $\operatorname{Cov}(X)$ be the *category of path-connected covering spaces of the space* X defined as follows:

- The objects of $\operatorname{Cov}(X)$ are covering spaces $p\colon \widetilde{X} \to X$ of X with path-connected total space $\widetilde{X}$.
- The morphisms of $\operatorname{Cov}(X)$ from a covering space $p\colon \widetilde{X} \to X$ to a covering space $p'\colon \widetilde{X}' \to X$ are maps $f\colon \widetilde{X} \to \widetilde{X}'$ which make the diagram

$$\begin{tikzcd} \widetilde{X} \arrow[rr, "\phi"] \arrow[dr, "p"'] & & \widetilde{X}' \arrow[dl, "p'"] \\ & X & \end{tikzcd}$$

commutative.

- the composition of morphisms in $\operatorname{Cov}(X)$ is given by composing the maps $f\colon \widetilde{X} \to \widetilde{X}'$ and $f'\colon \widetilde{X}' \to \widetilde{X}''$; the identity morphism of $\widetilde{X} \to \widetilde{X}$ is the identity map of $\widetilde{X}$.

There is a variant of this category, namely the category $\operatorname{Cov}_*(X, x_0)$ of *pointed path-connected covering spaces* of the pointed space (X, x_0), where the objects are based covering spaces $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ with $\widetilde{X}$ path-connected, and the morphisms are maps $f\colon (\widetilde{X}, \widetilde{x}_0) \to (\widetilde{X}', \widetilde{x}'_0)$ that are compatible with the projection maps to X.

While the definition of the category of coverings of a topological space X does not require assumptions on X, we need assumptions in order to use the tools at our disposal to analyze that category:

- the General Lifting Criterion 5.14 assumes that the domain of the lift is path-connected and locally path-connected;

- the existence of the universal covering space $\widetilde{X}^u$ of a space X requires X to be semi-locally simply-connected, i.e., every point $x \in X$ has an open neighborhood U such that the induced homomorphism $\pi_1(U, x) \to \pi_1(X, x)$ is the trivial map.

So typically we will make the following requirement.

Definition 5.18. We will call a space X *nice* if X is path-connected, locally path-connected and semi-locally simply connected.

Theorem 5.19. (Classification of path-connected pointed covering spaces). *Let X be a nice space with basepoint x_0.*

(i) *Mapping a pointed covering space $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ to the subgroup $p_*\pi_1(\widetilde{X}, \widetilde{x}_0) < G := \pi_1(X, x_0)$ gives a bijection*

$$\mathrm{ob}(\mathrm{Cov}_*(X)) \xrightarrow{\cong} \{\textit{subgroups } H < G\}. \tag{5.20}$$

(ii) *Let $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ and $p\colon (\widetilde{X}', \widetilde{x}'_0) \to (X, x_0)$ be pointed covering spaces of X, and let $H := p_*\pi_1(\widetilde{X}, \widetilde{x}_0)$, $H' := p_*\pi_1(\widetilde{X}', \widetilde{x}'_0)$ be the corresponding subgroups of G. Then*

$$\mathrm{mor}((\widetilde{X}, \widetilde{x}_0), (\widetilde{X}', \widetilde{x}'_0)) = \begin{cases} * & \textit{if } H < H' \\ \emptyset & \textit{otherwise} \end{cases}$$

Here $*$ denotes the set with one element.

Proof. Let ϕ be a morphism

$$\begin{tikzcd} (\widetilde{X}, \widetilde{x}_0) \arrow[rr, "\phi"] \arrow[dr, "p"'] & & (\widetilde{X}', \widetilde{x}'_0) \arrow[dl, "p'"] \\ & X & \end{tikzcd}$$

in the category $\mathrm{Cov}_*(X)$. Then ϕ can be viewed as a *lift* of the map p. According to our General Lifting Criterion 5.14, there is at most one lift ϕ, and a lift ϕ exists if and only if $p_*\pi_1(\widetilde{X}, \widetilde{x}_0)$ is contained in $p_*\pi_1(\widetilde{X}', \widetilde{x}'_0)$. This proves part (ii).

To show that the map in part (i) is injective, assume that $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ and $p'\colon (\widetilde{X}', \widetilde{x}'_0) \to (X, x_0)$ are covering spaces such that $p_*\pi_1(\widetilde{X}, \widetilde{x}_0) = p_*\pi_1(\widetilde{X}', \widetilde{x}'_0)$. Then by part (ii), there exist morphisms

$$\phi\colon (\widetilde{X}, \widetilde{x}_0) \to (\widetilde{X}', \widetilde{x}'_0) \qquad \text{and} \qquad \psi\colon (\widetilde{X}', \widetilde{x}'_0) \to (\widetilde{X}, \widetilde{x}_0).$$

Then the compositions $\psi \circ \phi$ and $\phi \circ \psi$ must be the identity morphisms of $(\widetilde{X}, \widetilde{x}_0)$ resp. $(\widetilde{X}', \widetilde{x}'_0)$, since there is at most one morphism between any pair of objects.

The tedious part of the proof is to verify surjectivity of the map (5.20), i.e., the construction of a pointed covering space $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ such that $p_*(\widetilde{X}, \widetilde{x}_0)$ is a prescribed subgroup H of G. This is first done for $H = \{1\}$, i.e., $\widetilde{X}$ is a simply connected covering space of X, also called *the universal covering space of* X, which we will denote by $\widetilde{X}^u$. □

Missing: outline of the construction of the universal covering and of coverings corresponding to subgroups H of G; classification of (unpointed) coverings.

Remark 5.21. The theorem above can be used to prove a statement in group theory, namely that any subgroup H of a free group is again a free group. The proof of this needs in addition to Theorem 5.19(i) the following two facts which are quite intuitive:

(a) We can consider any graph X as a topological space (edges from a vertex v to the same vertex v are allowed). Then for any connected graph X, its fundamental group is a *free group* (since every connected graph is homotopy equivalent to a wedge of copies of S^1).

(b) Any covering space $p\colon \widetilde{X} \to X$ of a graph X is again a graph (the set of vertices of $\widetilde{X}$ consists of the preimages of vertices of X, the interiors of edges of $\widetilde{X}$ are the preimage of open edges in X).

A free group F is the fundamental group of some graph X (e.g., the graph with one vertex and an edge for any generator of F). Then by Theorem 5.19(i), for any subgroup $H < F = \pi_1(X)$, there is a covering space $p\colon \widetilde{X} \to X$ with $p_*\pi_1(\widetilde{X}) = H$. Since $p_*\colon \pi_1(\widetilde{X}) \to \pi_1(X)$ is injective, H is then isomorphic to $\pi_1(\widetilde{X})$. By fact (b), $\widetilde{X}$ is a graph, and hence by fact (a) its fundamental group is a free group.

5.4 The Seifert van Kampen Theorem via K-coverings

The goal of this section is twofold:

- To define K-coverings over a space X (see Definition 5.22) and to classify them up to isomorphism for spaces X which are *nice* in the sense of Definition 5.18, i.e., X is path-connected, locally path-connected and semi-locally simply connected.
- To use this classification result to prove the Seifert van Kampen Theorem for spaces X which are locally path-connected, and semi-locally simply connected.

Definition 5.22. Let K be a group, and let X be a topological space. A *K-covering over X* or *principal K-bundle over X* is a covering space $p\colon \widetilde{X} \longrightarrow$ equipped with a continuous K-action $K \times \widetilde{X} \to \widetilde{X}$ on the total space X. We require that

(i) The K-action on $\widetilde{X}$ preserves the fibers of p, i.e., $p(k \cdot \widetilde{x}) = p(\widetilde{x})$ for $k \in K$ and $\widetilde{x} \in \widetilde{X}$. In particular, the K-action on $\widetilde{X}$ restricts to a K-action on any fiber $p^{-1}(x) \subset \widetilde{X}$.

(ii) The K-action on any fiber $p^{-1}(x)$ is free and transitive. Equivalently, for any $\widetilde{x} \in p^{-1}(x)$ the map

$$K \to p^{-1}(x) \qquad \text{given by} \qquad k \mapsto k \cdot \widetilde{x} \tag{5.23}$$

is a bijection (the transitivity of the action is equivalent to the surjectivity of this map, while the freeness of the action is equivalent to the injectivity of this map).

Example 5.24. (Examples of K-covering spaces)

(a) Let $p\colon \mathbb{R} \to S^1$ be our old friend of a covering space, given by $p(t) = e^{2\pi it}$. Let the group $\mathbb{Z}$ act on $\mathbb{R}$ by translations, i.e., the action map $\mathbb{Z} \times \mathbb{R} \to \mathbb{R}$ is given by $(n,t) \mapsto n + t$. It is easy to check that this $\mathbb{Z}$-action is free and transitive on the fibers of p.

(b) Let $p\colon \widetilde{X}^u \to X$ be the universal covering of a nice space X. Let $K = \pi_1(X, x_0)$ be the fundamental group of X which acts on $\widetilde{X}^u = \{[\gamma] \mid \gamma \text{ is a path with starting point } x_0\}$ via the map

$$\pi_1(X, x_0) \times \widetilde{X}^u \longrightarrow \widetilde{X}^u \qquad ([\alpha], [\gamma]) \mapsto [\alpha * \gamma].$$

It is clear that the K-action preserves the fibers of p, and Proposition 5.8 implies that the action of K is free and transitive on the fibers. Specializing to $X = S^1$, we obtain back our first example.

(c) An example of a K-covering of X which is not the universal cover of X is the following. Let K be the cyclic group of three elements and let $p\colon S^1 \to S^1$ be the covering given by $z \mapsto z^3$. To describe the action, it will be convenient to think of K not as $\mathbb{Z}/3$, but as the group of third roots of unity, i.e., as $K = \{\zeta \in S^1 \subset \mathbb{C} \mid \zeta^3 = 1\}$. Let K act on S^1 via the map

$$G \times S^1 \longrightarrow S^1 \qquad (\zeta, z) \mapsto \zeta z.$$

This is an action map since for $\zeta, \zeta' \in G$ and $z \in S^1$ we have $\zeta(\zeta' z) = (\zeta\zeta')z$. It has the required properties:

(i) $p(\zeta z) = (\zeta z)^3 = \zeta^3 z^3 = z^3 = p(z)$.

(ii) G acts transitively on the fibers $p^{-1}(x)$ for any $x \in S^1$, since if $z, z' \in p^{-1}(x)$, then $z^3 = p(z) = p(z') = (z')^3$ and hence $1 = z^{-3}(z')^3 = (z^{-1}z')^3$. In other words, $\zeta := z^{-1}z'$ is a third root of unity and hence $\zeta z = z'$ shows that z' is in the G-orbit of z.

(iii) If $\zeta \in G$ fixes some point $z \in S^1$, then $\zeta z = z$, and hence $\zeta = zz^{-1} = 1$. In other words, G acts freely.

(d) A slight variation of this example is given by the covering $p\colon \mathbb{C}^\times \to \mathbb{C}^\times$, $z \mapsto z^3$ (here $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$). As above, let G be the group of third roots of unity and let G act on the total space $\mathbb{C}^\times$ by $(\zeta, z) \mapsto \zeta z$. The same arguments as above show that this is a G-covering. In fact, restricting this G-covering of $\mathbb{C}^\times$ to the subspace $S^1 \subset \mathbb{C}^\times$, we obtain the first G-covering.

To understand the structure of a K-covering $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ we use the idea we have already utilized a couple of times: we take a based loop $\gamma\colon (I, \partial I) \to (X, x_0)$ and consider its unique lift $\widetilde{\gamma}\colon I \to \widetilde{X}$ with starting point $\widetilde{\gamma}(0) = \widetilde{x}_0$. Then the endpoint $\widetilde{\gamma}(1)$ belongs to the fiber $p^{-1}(x_0)$. The bijection (5.23) shows that there is a *unique* element $k \in K$ such that $\widetilde{\gamma}(1) = k \cdot \widetilde{x}_0$, called the holonomy of $\widetilde{X}$ along the loop γ. Here is the formal definition:

Definition 5.25. Let $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ be a K-covering, and let γ be a based loop in (X, x_0). The *holonomy of* $\widetilde{X}$ *along the loop* γ is the element $\mathrm{Hol}^{\widetilde{X}}(\gamma) \in K$ uniquely determined by the equation

$$\widetilde{\gamma}(1) = \mathrm{Hol}^{\widetilde{X}}(\gamma) \cdot \widetilde{x}_0, \tag{5.26}$$

where $\widetilde{\gamma}$ is the unique lift of γ with $\widetilde{\gamma}(0) = \widetilde{x}_0$.

In the special case 5.24(a) of the standard covering $\mathbb{R} \to S^1$ where $K = \mathbb{Z}$ acts on $\mathbb{R}$ by translation, the holonomy $\mathrm{Hol}^{\mathbb{R}}(\gamma) \in \mathbb{Z}$ of a loop γ in $(S^1, 1)$ is exactly the winding number $W(\gamma) \in \mathbb{Z}$ as defined in Definition 3.19.

We recall that the endpoint of the lift $\widetilde{\gamma}(1)$ depends only on the homotopy class $[\gamma] \in \pi_1(X, x_0)$. Hence the holonomy $\mathrm{Hol}^{\widetilde{X}}(\gamma) \in K$ depends only on $[\gamma]$, and we obtain a well-defined *holonomy map*

$$\mathrm{Hol}^{\widetilde{X}}\colon \pi_1(X, x_0) \longrightarrow K \qquad \text{given by} \qquad [\gamma] \mapsto \mathrm{Hol}^{\widetilde{X}}(\gamma). \tag{5.27}$$

Lemma 5.28. *The holonomy map* (5.27) *is a group homomorphism.*

Proof. Let γ, β be two based loops in (X, x_0) and let $\widetilde{\gamma}, \widetilde{\beta}\colon I \to \widetilde{X}$ be lifts of γ resp. β with starting poing $\widetilde{x}_0$. Then by the definition of holonomy, we have

$$\widetilde{\gamma}(1) = \mathrm{Hol}(\gamma) \cdot \widetilde{x}_0 \qquad \text{and} \qquad \widetilde{\beta}(1) = \mathrm{Hol}(\beta) \cdot \widetilde{x}_0.$$

To evaluate the holonomy of the concatenated loop $\gamma * \beta$, we need to construct a lift $\widetilde{\gamma * \beta}$ of $\gamma * \beta$. We note that in general we cannot form the concatenation $\widetilde{\gamma} * \widetilde{\beta}$, since the endpoint of $\widetilde{\gamma}$ is $\mathrm{Hol}(\gamma) \cdot \widetilde{x}_0$, while the starting point of $\widetilde{\beta}$ is $\widetilde{x}_0$. However, for any $k \in K$, we can create a lift of β with starting point $k \cdot \widetilde{x}_0$ by letting $k \in K$ to the path $\widetilde{\beta}$ to create a new path

$$k \cdot \widetilde{\beta}\colon I \to \widetilde{X} \qquad \text{defined by} \qquad (k \cdot \widetilde{\beta})(s) := k \cdot \widetilde{\beta}(s).$$

In particular for $k = \mathrm{Hol}(\gamma)$, the path $\mathrm{Hol}(\gamma) \cdot \widetilde{\beta}$ has starting point $\mathrm{Hol}(\gamma)\widetilde{x}_0$, and hence the concatenation

$$\widetilde{\gamma} * (\mathrm{Hol}(\gamma) \cdot \widetilde{\beta}) \qquad \text{is a lift of } \gamma * \beta$$

with starting point $\widetilde{x}_0$. By definition of holonomy, its endpoint is $\mathrm{Hol}(\gamma * \beta) \cdot \widetilde{x}_0$. Hence

$$\begin{aligned}\mathrm{Hol}(\gamma * \beta) \cdot \widetilde{x}_0 =& (\widetilde{\gamma} * (\mathrm{Hol}(\gamma) \cdot \widetilde{\beta}))(1) = (\mathrm{Hol}(\gamma) \cdot \widetilde{\beta})(1) \\ =& \mathrm{Hol}(\gamma) \cdot (\mathrm{Hol}(\beta) \cdot \widetilde{x}_0) = (\mathrm{Hol}(\gamma) \cdot \mathrm{Hol}(\beta)) \cdot \widetilde{x}_0\end{aligned}$$

It follows that $\mathrm{Hol}(\gamma * \beta) = \mathrm{Hol}(\gamma) \cdot \mathrm{Hol}(\beta)$, which proves that the holonomy map is a group homomorphism. □

The K-coverings of a fixed topological space X are the objects of a category $K\text{-Cov}(X)$, whose morphisms are maps between the total spaces of K-coverings compatible with the K-action and the projection maps to X. For the proof of the Seifert van Kampen Theorem we are interested in the pointed version of this category defined as follows.

Definition 5.29. Let (X, x_0) be a pointed topological space and K a group. The category $K\text{-Cov}_*(X, x_0)$ of *pointed K-coverings of* (X, x_0) is defined as follows. The objects are K-coverings $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$. A morphism ϕ from $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ to $p'\colon (\widetilde{X}', \widetilde{x}'_0) \to (X, x_0)$ is a map $\phi\colon (\widetilde{X}, \widetilde{x}_0) \to (\widetilde{X}', \widetilde{x}'_0)$ with the following two properties:

(i) the diagram

$$\begin{tikzcd} (\widetilde{X}, \widetilde{x}_0) \arrow[rr, "\phi"] \arrow[dr, "p"'] & & (\widetilde{X}', \widetilde{x}'_0) \arrow[dl, "p'"] \\ & X & \end{tikzcd}$$

is commutative, and

(ii) the map $\phi\colon \widetilde{X} \to \widetilde{X}'$ is K-equivariant, i.e., $\phi(k\widetilde{x}) = k\phi(\widetilde{x})$ for all $k \in K$, $\widetilde{x} \in \widetilde{X}$.

The composition in $K\text{-Cov}_*(X, x_0)$ is given by composing these K-equivariant covering maps.

Theorem 5.30. *Let X be a nice space with basepoint x_0.*

(i) *There is a bijection*

$$\mathrm{ob}(K\text{-Cov}_*(X, x_0))/\mathrm{isom} \xrightarrow{\cong} \mathrm{Hom}(\pi_1(X, x_0), K)$$

given by sending a pointed K-covering space $p\colon (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ to the holonomy map $\mathrm{Hol}^{\widetilde{X}}\colon \pi_1(X, x_0) \to K$.

(ii) *There is at most one morphism ϕ from a pointed K-covering $(\widetilde{X}, \widetilde{x}_0)$ to another pointed K-covering $(\widetilde{X}', \widetilde{x}_0')$. Such a morphism exists if and only if*

$$\mathrm{Hol}^{\widetilde{X}} = \mathrm{Hol}^{\widetilde{X}'} \in \mathrm{Hom}(\pi_1(X, x_0), K).$$

Proof missing

With these tools in our toolbox, we are ready to prove the Seifert van Kampen Theorem (Theorem 4.38). We recall the statement.

Theorem 5.31. *Let U, V be open subsets of topological space X such that $U \cup V = X$ and $U \cap V$ is path connected. Then the commutative diagram of fundamental groups (with respect to a basepoint $x_0 \in U \cap V$)*

$$\begin{tikzcd} \pi_1(U\cap V, x_0) \arrow[r, "j^U"] \arrow[d, "j^V"] & \pi_1(U, x_0) \arrow[d, "i^U"] \\ \pi_1(V, x_0) \arrow[r, "i^V"'] & \pi_1(X, x_0) \end{tikzcd}$$

is a pushout diagram in the category of groups.

Proof. We need to show that the commutative diagram of groups

$$\begin{tikzcd} \pi_1(U\cap V, x_0) \arrow[r, "j^U_*"] \arrow[d, "j^V_*"'] & \pi_1(U, x_0) \arrow[d, "i^U_*"] \\ \pi_1(V, x_0) \arrow[r, "i^V_*"'] & \pi_1(X, x_0) \end{tikzcd} \tag{5.32}$$

is a pushout diagram. Let $U' \subset U$ be the *path component* of U containing the basepoint x_0, which consists of all points $x \in U$ for which there is a path starting in x_0 and ending in x. This subspace of U is a path-connected space and the inclusion map $i\colon U' \to U$ induces an isomorphism $i_*\colon \pi_1(U', x_0) \xrightarrow{\cong} \pi_1(U, x_0)$, since every based loop in (U, x_0) and homotopy between based loops is necessarily contained in U'.

To verify that diagram (5.32) is a pushout square, we need to check whether it has the universal property expressed by the following diagram of groups and group homomorphisms

$$\begin{tikzcd} \pi_1(U\cap V, x_0) \arrow[r, "j^U_*"] \arrow[d, "j^V_*"'] & \pi_1(U, x_0) \arrow[d, "i^U_*"] \arrow[ddr, bend left, "f_1"] & \\ \pi_1(V, x_0) \arrow[r, "i^V_*"] \arrow[drr, bend right, "f_2"'] & \pi_1(X, x_0) \arrow[dr, dashed, "\exists! f"] & \\ & & G \end{tikzcd} \tag{5.33}$$

This diagram contains many group homomorphisms from fundamental groups of U, V, $U \cap V$ and X to the group G. Via Theorem 5.30 these homomorphisms can be interpreted geometrically as G-coverings over these topological spaces (up to isomorphism). In particular, the homomorphism $f_2 \in \operatorname{Hom}(\pi_1(U, x_0), G)$ corresponds to some G-covering

$$p_1 \colon (\widetilde{U}, \widetilde{x}_1) \longrightarrow (U, x_0).$$

This G-covering restricts to a G-covering $(\widetilde{U}_{|U \cap V}, \widetilde{x}_1) \longrightarrow (U \cap V, x_0)$, which by the Addendum corresponds to the composition

$$\pi_1(U \cap V, x_0) \xrightarrow{j_*^U} \pi_1(U, x_0) \xrightarrow{f_1} G.$$

Similarly, $f_2 \colon \pi_1(V, x_0) \to G$ corresponds to a G-covering $p_2 \colon (\widetilde{V}, \widetilde{x}_2) \to (V, x_0)$. Its restriction to $U \cap V$ is a G-covering $(\widetilde{V}_{|U \cap V}, \widetilde{x}_2) \longrightarrow (U \cap V, x_0)$ which corresponds to the composition $f_2 \circ j_*^V \colon \pi_1(U \cap V, x_0) \to G$. By the commutativity of the outer square of the diagram (**??**), $f_1 \circ j_*^U = f_2 \circ j_*^V$. By the Classification Theorem for G-coverings, this implies that there is an isomorphism between the corresponding based covering spaces over $U \cap V$, i.e., there is a G-equivariant map basepoint preserving map ϕ making the following diagram commutative: □

Proof incomplete

6 Smooth manifolds

The goal of this second part of the semester is to do **Calculus on manifolds**. This can be motivated by physics. According to Newton's law, $F = ma$, where F is the force acting on a point particle of mass m and acceleration a. More explicitly, writing $x(t) \in \mathbb{R}^3$ for the position of the point particle at time t, this equation is the second order differential equation

$$m\ddot{x}(t) = F(x(t)).$$

In particular, the trajectory $x(t)$ of the particle can be determined by solving this differential equation if the position $x(t_0)$ and the velocity $\dot{x}(t_0)$ are know at some initial time t_0.

If the particle is somehow restricted to moving on the sphere $S^2 \subset \mathbb{R}^3$, then the same general discussion applies, but leads to a second order differential equation on the 2-sphere. The same is typically true when describing the position of extended objects, say rod, which can be done by giving the position v of the center of mass, and a unit vector u pointing in the direction of the rod. Then the set of such pairs (v, u) forms a manifold of dimension 5, and determining the trajectory $(v(t), u(t))$ amounts to solving a differential equation on that manifold.

6.1 Smooth structures

Let U be an open subset of $\mathbb{R}^n$, and let $f\colon U \to \mathbb{R}$ be a real valued function. The two main constructions of Calculus are

- **Differentiation**, which associates to f its (total) derivative

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n \tag{6.1}$$

- **Integration**, which associates to f the integral

$$\int_U f(x) dx_1 \dots dx_n. \tag{6.2}$$

Needless to say, some conditions on the function f are needed to ensure that df exists (f should be differentiable) or that the integral over U exists (f should be integrable). For our geometric purposes here, we will assume that the functions we consider are *smooth*, i.e., can be differentiated as often as desired, which will ensure that all derivatives or integrals we will consider below exist. This is a condition much stronger than needed, but it will be pretty clear how the theory of smooth manifolds can be modified to get away with less amount of differentiability.

Definition 6.3. Let U be an open subset of $\mathbb{R}^n$. A function $f\colon U \to \mathbb{R}$ is *smooth* if for all n-tupels $(k_1, \dots, k_n)$, $k_i \in \mathbb{N}$, the corresponding partial derivative

$$\frac{\partial^k f}{\partial x^{k_1} \dots \partial x_n^{k_n}}(x)$$

exist for all points $x \in U$; here $k = \sum_{i=1}^n k_i$. A map $f = (f_1, \dots, f_m)\colon U \to V \subset \mathbb{R}^m$ is smooth if all its component functions f_i are smooth. This map f is a *diffeomorphism* if f is a bijection and its inverse map $f^{-1}\colon V \to U$ is smooth as well (it turns out that this can only happen if $m = n$).

Goals of this section.

(i) Define what a "smooth function" $f\colon M \to R$ on a manifold M is.

(ii) Define the total derivative df of a smooth function. What kind of object is a df?

(iii) Define suitable objects on an n-manifold M that can be integrated over M (if M is an open subset of $\mathbb{R}^n$, these are just of the form $f(x)dx_1 \dots dx_n$).

Before addressing (i) in the section, we remark that df is a *smooth section of the cotangent bundle of* M, while the object sought in (iii) will turn out to be a *smooth section of a vectorbundle* built from the cotangent bundle. This explains our interest in vector bundles and their sections which will be discussed in the following sections.

Definition 6.4. Let M be a topological manifold of dimension n. A *chart* for M is an open subset $U \subset M$ and a homeomorphism

$$M \underset{\text{open}}{\supset} U \xrightarrow{\phi} \phi(U) \underset{\text{open}}{\subset} \mathbb{R}^n.$$

A collection of charts $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ is an *atlas* for M if $\{U_\alpha\}_{\alpha \in A}$ is an open cover of M, i.e., if $M = \bigcup_{\alpha \in A} U_\alpha$.

Example 6.5. (Examples of charts and atlases).

1. Let $M = S^n$ be the sphere of dimension n. For $i = 0, \dots, n$ let $U_i^+ := \{(x_0, \dots, x_n) \in S^n \mid x_i > 0\}$ and $U_i^- := \{(x_0, \dots, x_n) \in S^n \mid x_i < 0\}$. Let $\phi_i^\pm \colon U_i^\pm \to B_1^n$ be the homeomorphism given by
$$\phi_i^\pm(x_0, \dots, x_n) = (x_1, \dots, x_{i-1}, \widehat{x}_i, x_{i+1}, \dots, x_n).$$
Then (U_i^+, ϕ_i^+) and (U_i^-, ϕ_i^-) are $2(n+1)$ charts for the manifold S^n. Since every point of S^n belongs to some hemisphere $U_i^\pm$, this collection of charts form an atlas for S^n.

2. A smaller atlas of S^n consisting of just two charts is obtained by using the homeomorphisms
$$\phi^\pm \colon U^\pm := S^n \setminus \{(0, \dots, \pm 1)\} \xrightarrow{\approx} \mathbb{R}^n$$
given by stereographic projections.

3. Let M be the projective space $\mathbb{RP}^n$, considered as the quotient space $\mathbb{R}^n \setminus \{0\}/x \sim \lambda x$ for $x \in \mathbb{R}^n \setminus \{0\}$ and $\lambda \in \mathbb{R} \setminus \{0\}$. Let $U_i \subset \mathbb{RP}^n$ be the open subset given by $U_i := \{([x_0, \dots, x_n] \in \mathbb{RP}^n \mid x_i \neq 0\}$. Then the map
$$U_i \xrightarrow{\phi_i} \mathbb{R}^n \qquad [x_0, \dots, x_n] \mapsto (\frac{x_0}{x_i}, \dots, \frac{x_{i-1}}{x_i}, \widehat{\frac{x_i}{x_i}}, \frac{x_{i+1}}{x_i}, \dots, \frac{x_n}{x_i})$$
is a homeomorphism with inverse given by $\phi_i^{-1}(v_1, \dots, v_n) = [v_1, \dots, v_{i-1}, 1, v_i, \dots, v_n]$. In particular, the collection of charts $\{(U_i, \phi_i)\}_{i=0,\dots,n}$ is an atlas for $\mathbb{RP}^n$.

Smoothness of a function $\mathbb{R}^n \supset U \to \mathbb{R}$ is a *local* property in the sense that if U is the union of open subsets $U_\alpha \subset U$ with $\bigcup_{\alpha \in A} U_\alpha = U$, then f is smooth if and only if the restriction f_{U_α} is smooth for all $\alpha \in A$. Let M be an n-manifold with an atlas $(U_\alpha, \phi_\alpha)_{\alpha \in A})$, and let $f\colon M \to \mathbb{R}$ be a map. The observation above suggests to define that $f_{|U_\alpha}\colon U_\alpha \to \mathbb{R}$ is *smooth* if the composition

$$\mathbb{R}^n \supset \phi_\alpha(U_\alpha) \xrightarrow{\phi_i^{-1}} U_\alpha \xrightarrow{f_{|U_\alpha}} \mathbb{R}$$

is smooth (unlike $U_\alpha \subset M$, the image $\phi_\alpha(U_\alpha)$ is an open subset of $\mathbb{R}^n$, and hence we already know by Definition 6.3 what a smooth function with domain $\phi_\alpha(U_\alpha)$ is). Then we define f to be *smooth* if its restriction $f_{|U_\alpha}\colon U_\alpha \to \mathbb{R}$ to each $U_\alpha \subset M$ is smooth.

The problem with the proposed definition. Suppose there are two charts (U,ϕ), (V,ψ) belonging to the atlas $\{(U_\alpha,\phi_\alpha)\}_{\alpha\in A}$ with $U\cap V \neq \emptyset$. Then according to the above definition, there would be *two* ways to determine whether the restriction $f_{|U\cap V}$ is smooth: we could use the chart (U,ϕ) and check whether the composition

$$\mathbb{R}^n \supset \phi(U\cap V) \xrightarrow[\approx]{\phi^{-1}} U\cap V \xrightarrow{f} \mathbb{R}$$

is smooth. Alternatively, we could use the chart (V,ψ) and check for smoothness of the composition

$$\mathbb{R}^n \supset \psi(U\cap V) \xrightarrow[\approx]{\psi^{-1}} U\cap V \xrightarrow{f} \mathbb{R}$$

The problem is that these two ways to test for smoothness of the function f on $U\cap V$ might **not** yield the same answer. We note that the second map $f\circ\psi^{-1}$ can be expressed as the composition

$$\begin{tikzcd} \psi(U\cap V) \arrow[r, "\phi\circ\psi^{-1}", "\approx"'] \arrow[rr, bend right, "f\circ\psi^{-1}"'] & \phi(U\cap V) \arrow[r, "f\circ\phi^{-1}"] & \mathbb{R} \end{tikzcd}$$

We also note that $\phi\circ\psi^{-1}$ is a map between open subsets of $\mathbb{R}^n$, and hence we can check by Definition 6.3 whether it is smooth. This map is a homeomorphism, but in general, there is no reason that this map should be smooth. It follows that if $f\circ\phi^{-1}$ is smooth, the map $f\circ\psi^{-1}$ in general won't be smooth. In other words, the smoothness test for f restricted to $U\cap V$ using the chart (U,ϕ) in general won't give the same answer as the smoothness test using the chart (V,ψ).

Let M be a topological manifold. How can we define whether a map $f\colon M\to\mathbb{R}$ is smooth?

First try. We call a f smooth at a point x if

Still missing: transition map, smooth compatible, smooth atlas, smooth structure, smooth manifold, smooth function on a manifold, smooth maps between manifolds

6.2 Tangent space

We begin by reviewing the differential of a smooth map between open subsets of Euclidean spaces. Our goal is to extend this definition to smooth maps between manifolds.

Definition 6.6. Let $\mathbb{R}^m \supset U \xrightarrow{F} V \subset \mathbb{R}^n$ be a smooth map. For a point $p \in U$, the *differential of F at p* is the linear map

$$dF(p) \colon \mathbb{R}^m \longrightarrow \mathbb{R}^n \tag{6.7}$$

which corresponds to the Jacobi matrix

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1}(p) & \dots & \frac{\partial F_1}{\partial x_m}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1}(p) & \dots & \frac{\partial F_n}{\partial x_m}(p) \end{pmatrix}$$

via the usual mechanism, i.e., the i-th column vector of the Jacobi matrix is equal $dF(p)$ applied to $e_i \in \mathbb{R}^m$, where $\{e_i\}_{i=1,\dots,m}$ is the standard basis of $\mathbb{R}^m$.

Theorem 6.8. (Chain Rule). *Let $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^p$ be open subsets, and let*

$$U \xrightarrow{\quad F \quad} V \xrightarrow{\quad G \quad} V$$

be smooth maps. Then for $p \in U$ the differential $d(G \circ F)(p)$ is the composition

$$\mathbb{R}^m \xrightarrow{\quad dF(p) \quad} \mathbb{R}^n \xrightarrow{\quad dG(F(p)) \quad} \mathbb{R}^p.$$

Our goal is to generalize the construction of the differential to smooth maps $F \colon M \to N$ between manifolds; i.e., for $p \in M$ we want to construct the *differential* $dF(p)$ which should be a linear map. One might suspect that the domain of $dF(p)$ is $\mathbb{R}^m$, where $m = \dim M$, and its codomain is $\mathbb{R}^n$, $n = \dim N$. It turns out to be more involved, namely $dF(p)$ is a linear map

$$dF(p) \colon T_pM \longrightarrow T_{F(p)}N, \tag{6.9}$$

where T_pM is an m-dimensional vector space associated to M and $p \in M$, called the *tangent space of M at the point p*.

So our next goal is to define the tangent space T_pM; in fact, we will provide *two* definitions, the "geometric" definition, denoted $T_p^{\mathrm{geo}}M$ and the "algebraic" definition, denoted $T_p^{\mathrm{alg}}M$. The reason for dealing with both, rather than settling on one of these is that both have their pros and cons, and hence it's good to know both of them.

6.2.1 The geometric tangent space

The strategy for for how to construct either $T_p^{\mathrm{geo}}M$ or $T_p^{\mathrm{alg}}M$ is basically the same: if M is an open subset of $\mathbb{R}^n$, the to be constructed tangent space $T_p^{\mathrm{geo}}M$ resp. $T_p^{\mathrm{alg}}M$ should be isomorphic to $\mathbb{R}^n$. So in both cases, we first observe that there is a very peculiar, roundabout way to think about a vector $v \in \mathbb{R}^n$. The only redeeming quality of this is that this very peculiar, roundabout way still makes sense when we are no longer looking at a point p of some open subset of $\mathbb{R}^n$, but for p a point in a smooth manifold M.

Observation. Let $M \subset \mathbb{R}^n$ be an open subset, and let p be a point of M. We define an equivalence relation $\sim$ on smooth paths $\gamma\colon (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ as follows

$$((-\epsilon_1, \epsilon_1) \xrightarrow{\gamma_1} M) \ \sim \ ((-\epsilon_2, \epsilon_2) \xrightarrow{\gamma_2} M) \qquad \text{if and only if} \qquad \gamma_1'(0) = \gamma_2'(0)$$

It is evident that the map

$$\{\gamma\colon (-\epsilon, \epsilon) \to M \mid \gamma \text{ is smooth}\}/\sim \quad \to \quad \mathbb{R}^n \qquad \text{given by} \qquad [\gamma] \mapsto \gamma'(0)$$

is a bijection. In other words, these equivalence classes of smooth paths is just a very complicated way to think about vectors in $\mathbb{R}^n$. The only redeeming quality of doing this is that this construction works in the more general case where M is a smooth manifold rather than an open subset of $\mathbb{R}^n$ and motivates the following definition.

Definition 6.10. (The geometric tangent space). Let M be a smooth n manifold and $p \in M$. We define the *geometric tangent space of M at p* to be

$$T_p^{\mathrm{geo}}M := \{\gamma\colon (-\epsilon, \epsilon) \to M \mid \gamma \text{ is smooth and } \gamma(0) = p\}/\sim,$$

where two such paths $\gamma_1\colon (-\epsilon_1, \epsilon_1) \to M$ and $\gamma_2\colon (-\epsilon_2, \epsilon_2) \to M$ are declared equivalent if for some smooth chart (U, ϕ) with $p \in U$ the tangent vectors of the paths $\phi \circ \gamma_1$ and $\phi \circ \gamma_2$ in $\mathbb{R}^n$ have the same tangent vector at $t = 0$ (we might restrict the domain to γ_i to a smaller interval around 0 such that the compositions $\phi \circ \gamma_i$ are defined. We observe that the seemingly stronger requirement that $(\phi \circ \gamma_1)'(0) = (\phi \circ \gamma_2)'(0)$ for *all* smooth charts (U, ϕ) with $p \in U$ is actually equivalent to the condition above.

We note that if the manifold M is an open subset of $\mathbb{R}^n$, we can use the smooth chart (M, i) provided by the inclusion map $i\colon M \to \mathbb{R}^n$. So in this case, the geometric tangent space $T_p^{\mathrm{geo}}M$ is equal to the quotient space discussed above and the map

$$T_p^{\mathrm{geo}}M \xrightarrow{\cong} \mathbb{R}^n \qquad [\gamma] \mapsto \gamma'(0) \tag{6.11}$$

is a bijection. So in this case, the geometric tangent space can be identified with $\mathbb{R}^n$ via this bijection, which we will often do without further comment.

Example 6.12. The idea of extracting the more concrete object $\gamma'(0)$ from the "abstract beast" $[\gamma] \in T_p^{\text{geo}} M$ works in more general situations, for example for $M = S^n \subset \mathbb{R}^{n+1}$ (or more generally for submanifolds $M \subset \mathbb{R}^{n+k}$, a concept we will define a little later). It is easy to show that if γ is a smooth path in S^n, then its composition $i \circ \gamma$ with the inclusion map $i \colon S^n \to \mathbb{R}^{n+1}$ is a smooth path in $\mathbb{R}^{n+1}$, and hence it has a tangent vector $(i \circ \gamma)'(0) \in \mathbb{R}^{n+1}$. It is not hard to show that

$$T_p^{\text{geo}} S^n \longrightarrow \mathbb{R}^{n+1} \qquad \text{given by} \qquad [\gamma] \mapsto (i \circ \gamma)'(0)$$

is a well-defined injective map. From the geometric picture it is clear that the tangent vector $(i \circ \gamma)'(0)$ of the path $i \circ \gamma$ should be perpendicular to the vector $p \in \mathbb{R}^{n+1}$. This can also be verified by the following calculation. Let $f \colon \mathbb{R}^{n+1} \to \mathbb{R}$ be the function $f(x_0, \dots, x_n) = x_0^2 + \cdots + x_n^2$, which can be used to describe S^n as $S^n = f^{-1}(1) \subset \mathbb{R}^{n+1}$. Hence if γ is a smooth path in S^n, then $f \circ \gamma$ is constant and so its derivative vanishes. On the other hand, we can calculate the derivative $(f \circ \gamma)'(0)$ via the chain rule and obtain the following equation:

$$0 = (f \circ \gamma)'(0) = \langle (\text{grad} f)(\gamma(0)), \gamma'(0) \rangle.$$

For $\gamma(0) = p = (x_0, \dots, x_n)$ we have $\text{grad} f = (\frac{\partial f}{\partial x_0}, \dots, \frac{\partial f}{\partial x_n}) = (2x_0, \dots, 2x_n) = 2p$. It follows that $0 = \langle p, \gamma'(0) \rangle$, i.e., the tangent vector $\gamma'(0)$ is perpendicular to p. Again, it is not difficult to show that every vector $v \in \mathbb{R}^{n+1}$ perpendicular to p is the tangent vector of some path γ in S^n. Summarizing, we obtain a bijection

$$T_p^{\text{geo}} S^n \xrightarrow{\cong} \{v \in \mathbb{R}^{n+1} \mid v \text{ is perpendicular to } p\} \qquad \text{given by} \qquad [\gamma] \mapsto \gamma'(0)$$

Definition 6.13. (The geometric differential). Let M, N be smooth manifolds, and let $F \colon M \to N$ be a smooth map. Then for $p \in M$ the *induced map of geometric tangent spaces* or the *differential of F at p* is the map

$$T_p^{\text{geo}} M \xrightarrow{F_*^{\text{geo}}} T_{F(p)}^{\text{geo}} N \qquad \text{given by} \qquad [\gamma] \mapsto [F \circ \gamma]$$

We might also write F_* if it is clear that we are using the geometric version of the tangent space.

Of course, we should justify the notation $dF(p)$ by showing that the map F_*^{geo} agrees with the differential as defined via the Jacobian matrix if M, N are open subsets of Euclidean space. This is the content of the next Lemma.

Lemma 6.14. *Let $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ be open subsets, let $F \colon U \to V$ be a smooth map, and let $p \in U$. Then the diagram*

$$\begin{tikzcd} T_p^{\text{geo}} U \arrow[r, "F_*^{\text{geo}}"] \arrow[d, "\cong"] & T_{F(p)}^{\text{geo}} V \arrow[d, "\cong"] \\ \mathbb{R}^m \arrow[r, "dF(p)"] & \mathbb{R}^n \end{tikzcd}$$

Here the vertical maps are the bijections (6.11)*, and the bottom horizontal map is the usual differential of* F *at* p *(see* (6.7)*) corresponding to the Jacobian matrix.*

Proof. Let $\gamma\colon(-\epsilon,\epsilon)\to U$ be a smooth path. Then the left vertical map sends $[\gamma]\in T_p^{\mathrm{geo}}U$ to $\gamma'(0)\in\mathbb{R}^m$, which via $dF(p)$ maps to $(dF(p))(\gamma'(0))\in\mathbb{R}^n$.

Going the other way, $F_*\mathrm{geo}([\gamma])=[F\circ\gamma]$, which via the right vertical map is send to $(F\circ\gamma)'(0)$. Using the chain rule,

$$(F\circ\gamma)'(0)=(dF(\gamma(0))(\gamma'(0))=(dF(p))(\gamma'(0)),$$

which proves that the diagram is commutative. □

skip this here, and do that for the algebraic tangent space, followed by the functor point of view.

Lemma 6.15. (Chain rule for the induced map of geometric tangent spaces). *Let* M*,* N*,* P *be smooth manifolds and let* $F\colon M\to N$*,* $G\colon N\to P$ *be smooth maps. Then for* $p\in M$ *the following diagram is commutative:*

$$\begin{tikzcd} T_p^{\mathrm{geo}}M \arrow[r, "F_*^{\mathrm{geo}}"'] \arrow[rr, bend left, "(G\circ F)_*^{\mathrm{geo}}"] & T_{F(p)}^{\mathrm{geo}}N \arrow[r, "G_*^{\mathrm{geo}}"'] & T_{G(F(p))}^{\mathrm{geo}}P \end{tikzcd}$$

Proof.

$$G_*^{\mathrm{geo}}(F_*^{\mathrm{geo}}([\gamma]))=G_*^{\mathrm{geo}}([F\circ\gamma])=[G\circ(F\circ\gamma)]=[(G\circ F)\circ\gamma]=(G\circ F)_*^{\mathrm{geo}}([\gamma]).$$

□

Corollary 6.16. *If* F *is a diffeomorphisms with inverse* G*, then* F_*^{geo} *is a bijection with inverse* G_*^{geo}*.*

In particular, if M is a smooth manifold with a smooth chart $M\supset U\xrightarrow{\phi}V:=\phi(U)\subset\mathbb{R}^n$, then ϕ is a diffeomorphism, and hence for $p\in U$ we have bijections

$$T_p^{\mathrm{geo}}M=T_p^{\mathrm{geo}}U\xrightarrow[\cong]{\phi_*^{\mathrm{geo}}}T_{\phi(p)}^{\mathrm{geo}}V\xrightarrow{\cong}\mathbb{R}^n \tag{6.17}$$

Here the last map is the bijection (6.11), and the equality $T_p^{\mathrm{geo}}M=T_p^{\mathrm{geo}}U$ follows from immediately from the definition of the geometric tangent space.

This motivates the algebraic tangent space; so it should be discussed in that section.

The above construction of the geometric tangent space $T_p^{\mathrm{geo}}M$ and the differential

$$F_*^{\mathrm{geo}}\colon T_p^{\mathrm{geo}}M\to T_q^{\mathrm{geo}}N$$

is pretty straightforward, geometrically intuitive and it agrees with what we want if the manifolds involved are open subsets of Euclidean space. So, what's not to like about the geometric tangent space?

The biggest drawback is that the geometric tangent space $T_p^{\text{geo}}M$ is not a vector space in an obvious way. We can use the bijection (6.17) to *define* a vector space structure on $T_p^{\text{geo}}M$ (which turns out to be independent of the choice of the chart (U, ϕ)), but this vector space structure does not have a direct geometric description.

6.2.2 The algebraic tangent space

The definition of the algebraic tangent space $T_p^{\text{alg}}M$ at a point p of a smooth manifold M is more involved and less intuitive than that of the geometric tangent space $T_p^{\text{geo}}M$. Its big advantage is that unlike the geometric tangent space, the algebraic tangent space has an evident vector space structure.

To motivate the definition of the algebraic tangent space, we consider the directional derivative of a smooth function $f \in C^\infty(M)$ defined as follows.

Definition 6.18. (Directional derivative). Let M be a smooth manifold, $f \in C^\infty(M)$ a smooth function, and $\gamma\colon (-\epsilon, \epsilon) \to M$ a smooth curve with $\gamma(0) = p$. Then the *directional derivative of f in the direction of γ* is

$$DD_\gamma f := (f \circ \gamma)'(0) = \lim_{t \to 0} \frac{f(\gamma(t)) - f(p)}{t} \in \mathbb{R}.$$

The directional derivative can also be expressed in terms of the differential f_* of $f\colon M \to \mathbb{R}$ as following result shows.

Lemma 6.19. *Let $f\colon M \to \mathbb{R}$ be a smooth map, and $\gamma\colon (-\epsilon, \epsilon) \to M$ a smooth curve with $\gamma(0) = p$. Then the directional derivative $DD_\gamma f$ is the image of the tangent vector $[\gamma] \in T_p^{\text{geo}}M$ under the composition*

$$T_p^{\text{geo}}M \xrightarrow{f_*} T_{f(p)}^{\text{geo}}\mathbb{R} \xrightarrow{\cong} \mathbb{R},$$

where the first map is the differential of f, and the second map is the isomorphism (6.11)*. In particular, $DD_\gamma f$ depends only on $[\gamma] \in T_p^{\text{geo}}M$.*

Proof. This is an immediate consequence of the definitions of the maps involved. The differential f_* maps $[\gamma] \in T_p^{\text{geo}}M$ to $[f \circ \gamma] \in T_{f(p)}^{\text{geo}}\mathbb{R}$ and the isomorphism $T_{f(p)}^{\text{geo}}\mathbb{R} \xrightarrow{\cong} \mathbb{R}$ maps $[f \circ \gamma] \in T_{f(p)}^{\text{geo}}\mathbb{R}$ to $(f \circ \gamma)'(0) \in \mathbb{R}$. □

Lemma 6.20. *Let M be a smooth manifold, let $[\gamma] \in T_p^{\text{geo}}M$ be a tangent vector at $p \in M$, and let $DD_\gamma\colon C^\infty(M) \to \mathbb{R}$ be the associated directional derivative. Then DD_γ has the following properties:*

(i) DD_γ *is a linear map.*

(ii) DD_γ *has the following product rule property:*

$$DD_\gamma(f \cdot g) = DD_\gamma(f) \cdot g(p) + f(p) \cdot DD_\gamma(g) \qquad \text{for } f, g \in C^\infty(M). \tag{6.21}$$

Proof. To prove part (i), let $f, g \in C^\infty(M)$. Then

$$DD_\gamma(f+g) = ((f+g)\circ\gamma)'(0) = (f\circ\gamma+g\circ\gamma)'(0) = (f\circ\gamma)'(0)+(g\circ\gamma)'(0) = DD_\gamma(f)+DD_\gamma(g).$$

The proof that $DD_\gamma(cf) = cDD_\gamma(f)$ for $c \in \mathbb{R}$ is a similarly straightforward calculation.

To prove part (ii), let $f, g \in C^\infty(M)$. Then

$$\begin{aligned} DD_\gamma(f \cdot g) =&((f \cdot g) \circ \gamma)'(0) = ((f \circ \gamma) \cdot (g \circ \gamma))'(0) \\ =&(f \circ \gamma)'(0) \cdot (g \circ \gamma)(0) + (f \circ \gamma)(0) \cdot (g \circ \gamma)'(0) \\ =&DD_\gamma f \cdot g(p) + f(p) \cdot DD_\gamma(g). \end{aligned}$$

□

Another obvious, but rather useful property of $DD_\gamma(f) = (f \circ \gamma)'(0)$ is that it *does only depend on the restriction of f to an arbitrary small open neighborhood $U \subset M$ of p.* More succinctly, this property can be expressed using the terminology of germs.

Definition 6.22. (Germs of smooth functions). Let M be a smooth manifold and $p \in M$. We consider smooth functions $f\colon U \to \mathbb{R}$ defined in an open neighborhood U of p, and define an equivalence relation on such functions by declaring $f\colon U \to M$ equivalent to $f'\colon U' \to \mathbb{R}$ if there is some open neighborhood $U'' \subset U \cap U'$ such that $f_{|U''} = f'_{|U''}$. Such an equivalence class is called a *germ of a smooth function at p*. We will write $C^\infty_p(M)$ for the set of germs of smooth functions at p, and $[f]_p \in C^\infty_p(M)$ for the germ represented by smooth function f defined in some open neighborhood of p.

Using this terminology, we can say that $[\gamma] \in T^{\text{geo}}_p M$ determines a map

$$DD_\gamma\colon C^\infty_p(M) \longrightarrow \mathbb{R}. \tag{6.23}$$

We note that the usual addition/multiplication of smooth functions is compatible with the germ equivalence relation defined above, and hence with can add and multiply elements of $C^\infty_p(M)$. This shows that $C^\infty_p(M)$ is an algebra. Lemma 6.20 implies that the map (6.23) is linear and satisfies the product rule (6.21) (with $f, g \in C^\infty_p(M)$ instead of $f, g \in C^\infty(M)$).

Definition 6.24. (Derivation). A map $D\colon C_p^\infty(M) \to \mathbb{R}$ which is linear and has the product rule property

$$D(fg) = D(f)g(p) + f(p)D(g) \qquad \text{for } f, g \in C_p^\infty(M)$$

is a *derivation*. We write $\mathrm{Der}(C_p^\infty(M), \mathbb{R})$ for the set of all derivations $C_p^\infty(M) \to \mathbb{R}$. We note that the sum $D + D'$ of derivations D, D' is again a derivation, and so is cD for $c \in \mathbb{R}$. In particular, $\mathrm{Der}(C_p^\infty(M), \mathbb{R})$ is a vector space.

Hence there is a well-defined map

$$DD^M\colon T_p^{\mathrm{geo}}M \longrightarrow \mathrm{Der}(C_p^\infty(M), \mathbb{R}) \qquad \text{given by} \qquad [\gamma] \mapsto DD_\gamma. \tag{6.25}$$

We will show that the map DD is a bijection, first in the case that M is an open subset of Euclidean space (Lemma 5.9), then for a general manifold M (Lemma 5.9). This motivates the following definition.

Definition 6.26. (The algebraic tangent space). Let M be a smooth manifold. Then the *algebraic tangent space at a point* $p \in M$ is the vector space $T_p^{\mathrm{alg}}M := \mathrm{Der}(C_p^\infty(M), \mathbb{R})$.

Remark 6.27. The map $C^\infty(M) \to C_p^\infty(M)$ which sends a smooth function f to its germ $[f]_p$ is an algebra homomorphism, and hence induces a homomorphism

$$\mathrm{Der}(C_p^\infty(M), \mathbb{R}) \longrightarrow \mathrm{Der}(C^\infty(M), \mathbb{R}).$$

This map is an isomorphism (this is a homework problem), and hence some authors, for example Lee in his book on smooth manifolds, prefer to work with smooth functions on M rather their germs at p. However, working with germs has technical advantages as will be pointed out later in this section.

The next goal is to define for a smooth map $F\colon M \to N$ with $F(p) = q$ the differential of F as a map from $T_p^{\mathrm{alg}}M$ to $T_q^{\mathrm{alg}}N$. We note that the smooth map $F\colon M \to N$ induces an algebra homomorphism $F^*\colon C^\infty(N) \to C^\infty(M)$ given by $F^*(g) := g \circ F$. It also induces an algebra homomorphism

$$F_p^*\colon C_q^\infty(M) \longrightarrow C_p^\infty(N)$$

by mapping a germ $[N \supset V \xrightarrow{g} \mathbb{R}] \in C_q^\infty(N)$ to $[M \supset F^{-1}(V) \xrightarrow{g\circ F} \mathbb{R}] \in C_p^\infty(M)$. These maps are compatible in the sense that the diagram

$$\begin{tikzcd}
C^\infty(N) \arrow[r, "F^*"] \arrow[d] & C^\infty(M) \arrow[d] \\
C_q^\infty(N) \arrow[r, "F_p^*"] & C_p^\infty(M)
\end{tikzcd}$$

is commutative. Here the vertical maps send a smooth function on N (resp. M) to its germ at q (resp. p).

Definition 6.28. (The algebraic differential) Let $F\colon M \to N$ be a smooth map with $F(p) = q$. Then the *(algebraic) differential*

$$F_*^{\mathrm{alg}}\colon T_p^{\mathrm{alg}}M = \mathrm{Der}(C_p^\infty(M), \mathbb{R}) \longrightarrow T_q^{\mathrm{alg}}N = \mathrm{Der}(C_q^\infty(N), \mathbb{R})$$

is defined by $F_*^{\mathrm{alg}}(D) := D \circ F_p^*$.

Of course, for this definition to make sense, we need to check that the composition

$$C_q^\infty(N) \xrightarrow{F_p^*} C_p^\infty(M) \xrightarrow{D} \mathbb{R}$$

is in fact a derivation. This is left as a homework problem.

Lemma 6.29. (Chain rule for the differential). *Let M, N, P be smooth manifolds and let $F\colon M \to N$, $G\colon N \to P$ be smooth maps. Then for $p \in M$ the following diagram is commutative:*

$$\begin{tikzcd}
T_p^{\mathrm{alg}}M \arrow[r, "F_*^{\mathrm{alg}}"'] \arrow[rr, bend left, "(G\circ F)_*^{\mathrm{alg}}"] & T_{F(p)}^{\mathrm{alg}}N \arrow[r, "G_*^{\mathrm{alg}}"'] & T_{G(F(p))}^{\mathrm{alg}}P
\end{tikzcd}$$

The proof of this statement is also left as a homework problem. We note that if $F\colon M \to N$ is a diffeomorphism and $G\colon N \to M$ is the inverse of F, then the chain rule implies that $G_*^{\mathrm{alg}} \circ F_*^{\mathrm{alg}}$ and $F_*^{\mathrm{alg}} \circ G_*^{\mathrm{alg}}$ are the identity maps on $T_p^{\mathrm{alg}}M$ resp. $T_{F(p)}^{\mathrm{alg}}N$. In particular, this implies the following result.

Corollary 6.30. *If $F\colon M \to N$ is a diffeomorphism, then the differential $F_*^{\mathrm{alg}}\colon T_p^{\mathrm{alg}}M \to T_{F(p)}^{\mathrm{alg}}N$ is a vector space isomorphism.*

Remark 6.31. There is a categorical way to think about the chain rule. Let Man_* be the the following category. The objects of Man_* are smooth manifolds M equipped with a basepoint $p \in M$. A morphism from (M, p) to (N, q) is a smooth map $F\colon M \to N$ with $F(p) = q$. Then our constructions in this section can be interpreted as a functor from the category Man_* to the category $\mathsf{Vect}_{\mathbb{R}}$ of real vector spaces, which sends an object $(M, p) \in \mathsf{Man}_*$ to the tangent space $T_p^{\mathrm{alg}}M \in \mathsf{Vect}_{\mathbb{R}}$ and a morphism $F\colon (M, p) \to (N, q)$ in Man_* to the differential $F_*^{\mathrm{alg}}\colon T_p^{\mathrm{alg}}M \to T_q^{\mathrm{alg}}N$. The chain rule is then the statement that this in fact gives a *functor*, namely that this construction is compatible with compositions (it is clear that this construction maps the identity map of (M, p) to the identity map of $T_p^{\mathrm{alg}}M$).

From this point of view the corollary above follows from the observation that a functor maps isomorphisms in Man_* (i.e., diffeomorphisms) to isomorphism in $\mathsf{Vect}_{\mathbb{R}}$.

Remark 6.32. Now we want to comment on the technical advantage to work with germs. Let M be a smooth manifold and $U \subset M$ be an open neighborhood of p, then by definition of germs of functions $C_p^\infty(U) = C_p^\infty(M)$. Hence

$$T_p^{\text{alg}}U = \text{Der}(C_p^\infty(U), \mathbb{R}) = \text{Der}(C_p^\infty(M), \mathbb{R}) = T_p^{\text{alg}}M.$$

In particular, we do not need a smooth map $F\colon M \to N$ to obtain a differential $F_*^{\text{alg}}\colon T_p^{\text{alg}}M \to T_{F(p)}^{\text{alg}}N$; it suffices to have a smooth map F $M \subset U \xrightarrow{F} N$:

$$T_p^{\text{alg}}M = T_p^{\text{alg}}U \xrightarrow{F_*^{\text{alg}}} T_{F(q)}^{\text{alg}}N.$$

For example, if $M \supset U \xrightarrow{\phi} V \subset \mathbb{R}^m$ is a smooth chart, then ϕ is a diffeomorphism and hence we obtain a vector space isomorphism

$$T_p^{\text{alg}}M = T_p^{\text{alg}}U \xrightarrow{\quad\phi_*^{\text{alg}}\quad} T_{\phi(p)}^{\text{alg}}V \cong \mathbb{R}^m.$$

The fact that a smooth chart (U, ϕ) with $p \in U$ determines an isomorphism $T_p^{\text{alg}}M \cong \mathbb{R}^m$ is extremely useful for calculations, since it provides us with an explicit basis for the tangent space $T_p^{\text{alg}}M$ given by the images of the standard basis elements $e_i \in \mathbb{R}^m$ under this isomorphism.

After introducing the two flavors of tangent spaces (the geometric and the algebraic), it is now time to show that these two approaches are equivalent.

Lemma 6.33. *Let U be an open subset of $\mathbb{R}^m$ and $p \in U$. Then the directional derivative map*

$$DD\colon T_p^{\text{geo}}U \longrightarrow T_p^{\text{alg}}U = \text{Der}(C_p^\infty(U), \mathbb{R})$$

is a bijection.

Proof. We recall that there there is a bijection $\mathbb{R}^m \longrightarrow T_p^{\text{geo}}U$ given by sending a vector $v \in \mathbb{R}^m$ to the element $[\gamma_v] \in T_p^{\text{geo}}U$ represented by the straight line path $\gamma_v(t) = p + tv$. Hence the statement of the lemma follows once we prove the following stronger statement that the composition

$$\Phi\colon \mathbb{R}^n \xrightarrow{\quad\cong\quad} T_p^{\text{geo}}U \xrightarrow{\quad D\quad} \text{Der}(C_p^\infty(U), \mathbb{R})$$

is an isomorphism of vector spaces.

We begin by providing an explicit formula for $(\Phi(v))(f) \in \mathbb{R}$, where $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, $f \in C^\infty(M)$, and $(\Phi(v))(f)$ is the derivation $\Phi(v) \in \operatorname{Der}(C^\infty(M), \mathbb{R})$ applied to the function f.

$$\begin{aligned}(\Phi(v))(f) =& D_{\gamma_v} f = (f \circ \gamma_v)'(0) = (\operatorname{grad} f)_p \cdot \gamma_v'(0) = (\operatorname{grad} f)_p \cdot v \\ =& \begin{pmatrix} \frac{\partial f}{\partial x^1}(p) \\ \vdots \\ \frac{\partial f}{\partial x^n}(p) \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(p) v_i\end{aligned}$$

This shows that Φ is a linear map.

To show that Φ is injective, it suffices to show the that kernel Φ is trivial. So assume that $v = (v_1, \dots, v_n) \in \ker \Phi$, i.e., for *every* smooth function f the derivation $\Phi(v)$ applied to v gives zero. Let $x^i \colon \mathbb{R}^n \to \mathbb{R}$ be the i^{th} coordinate function defined by $x^i(x_1, \dots, x_n) = x_i$ (so x_i is the i^{th} component of $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, while x^i is a smooth function). Then $(\text{grad } x^i)_p = e_i \in \mathbb{R}^n$, where e_i is the i^{th} vector of the standard basis of $\mathbb{R}^n$, which has 1 in the i^{th} component and 0 in all other components. It follows that

$$(\Phi(v)) = (\text{grad } x^i)_p \cdot v = e_i \cdot v = v_i.$$

Hence the kernel of Φ is trivial.

To prove that Φ is surjective, we need to show that for any derivation $D \in \operatorname{Der}(C^\infty(U), \mathbb{R})$ there is a vector $v = (v_1, \dots, v_n)$ such that for any $f \in C^\infty(M)$ we have

$$D(f) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(p) v_i \tag{6.34}$$

To prove this, we use the Taylor expansion of f around the point p. For simplicity, and without loss of generality, we assume $p = 0$. Then

$$f = f(0) + \sum_{i=1}^n \frac{\partial f}{\partial x^i}(0) x^i + \sum_{i,j} R_{i,j} x^i x^j,$$

where $R_{i,j}$ are smooth functions. Applying the derivation D to this equation, using the linearity of D, we obtain

$$D(f) = D(f(0)) + \sum_{i=1}^n \frac{\partial f}{\partial x^i}(0) D(x^i) + \sum_{i,j} D(R_{i,j} x^i x^j).$$

The second term on the right hand side has the desired form (6.34), with $v_i = D(x^i)$. So it suffices to show that first and third term are zero.

To show that D applied to any constant function $c \in C^\infty(\mathbb{R}^n)$ vanishes (e.g., the constant function $f(0)$), we first show $D(1) = 0$ for the constant function $1 \in C^\infty(\mathbb{R}^n)$:

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

and hence $D(c) = D(c \cdot 1) = cD(1) = 0$ (by linearity of D).

To show that $D(Rx^i x^j) = 0$ for any $R \in C^\infty(\mathbb{R}^n)$, we calculate:

$$D(Rx^i x^j) = D((Rx^i)x^j) = D(Rx^i)x^j(0) + (Rx^i)(0)D(x^j) = 0,$$

since both functions, Rx^i and x^j vanish at $0 \in \mathbb{R}^n$. $\square$

Lemma 6.35. (i) *Let $F\colon M \to N$ be a smooth map with $F(p) = q$. Then the diagram*

$$\begin{tikzcd} T_p^{\mathrm{geo}} M \arrow[r, "F_*^{\mathrm{geo}}"] \arrow[d, "DD"'] & T_q^{\mathrm{geo}} N \arrow[d, "DD"] \\ T_p^{\mathrm{alg}} M \arrow[r, "F_*^{\mathrm{alg}}"] & T_q^{\mathrm{alg}} N \end{tikzcd}$$

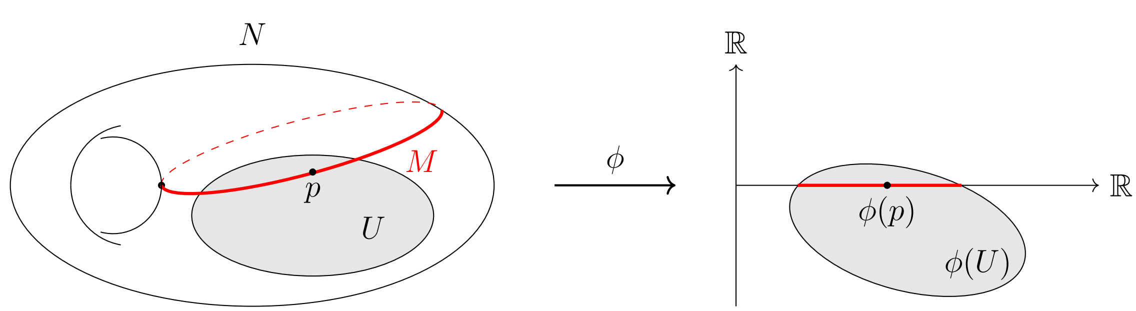
is a commutative diagram.

(ii) *For any smooth manifold M, the map $DD\colon T_p^{\mathrm{geo}} M \to T_p^{\mathrm{alg}} M$ is a bijection.*

This proof of this lemma is a homework problem. Hint for part (ii): Use a smooth chart (U, ϕ) with $p \in U$ to show that Lemma 6.33 implies the desired statement.

6.3 Smooth submanifolds

In this section we define what it means for a subset $M \subset N$ of a smooth manifold N to be a *submanifold*. If M is of the form $M = F^{-1}(q)$ for a smooth map $F\colon N \to Q$ and $q \in Q$, we give a sufficient condition on F and q for M to be a submanifold. Before giving the formal definition of a submanifold, here is a picture of a 2-manifold N and a submanifold M of dimension 1. The map $N \supset U \xrightarrow{\phi} \mathbb{R}^2$ shown in a picture is an example of a submanifold chart for $M \subset N$ (as defined below).



Definition 6.36. Let N be a smooth manifold of dimension $n = m + k$. A subset $M \subset N$ is a *smooth submanifold* of N of dimension m if the inclusion $M \hookrightarrow N$ locally is isomorphic to the inclusion $\mathbb{R}^m \hookrightarrow \mathbb{R}^{m+k}$, i.e., if for every $p \in M$ there is a smooth chart

$$N \underset{\text{open}}{\supset} U \xrightarrow{\phi} \mathbb{R}^{m+k} = \mathbb{R}^m \times \mathbb{R}^k$$

with $p \in U$ such that

$$\phi(M \cap U) = \phi(U) \cap (\mathbb{R}^m \times \{0\}).$$

A chart (U, ϕ) with this property will be called a *submanifold chart* for $M \subset N$. The integer k is referred to as the *codimension* of the submanifold M of N.

Remark 6.37. The restriction of ϕ to $M \cap U$ gives a chart $\phi_| \colon M \cap U \to \mathbb{R}^m$ for M. In particular, M is a topological manifold of dimension m. We claim that if (U, ϕ), (V, ψ) are two smooth charts of the special form described in the definition above, then the corresponding charts $\phi_|$, $\psi_|$ for M are smoothly compatible. To check this, we need to show that the transition map

$$\mathbb{R}^m \supset \phi(M \cap U \cap V) \xrightarrow{\phi_|^{-1}} M \cap U \cap V \xrightarrow{\psi_|} \psi(M \cap U \cap V) \subset \mathbb{R}^m$$

is smooth. Note that this map is given by restricting the transition map

$$\mathbb{R}^{m+k} \supset \phi(U \cap V) \xrightarrow{\phi^{-1}} U \cap V \xrightarrow{\psi} \psi(U \cap V) \subset \mathbb{R}^{m+k}$$

to $\mathbb{R}^m \subset \mathbb{R}^{m+k}$. Since the charts (U, ϕ), (V, ψ) are smoothly compatible, the transition function $\psi \circ \phi^{-1}$ is smooth, and hence also its restriction to $\mathbb{R}^m$. This shows that $(M \cap U, \phi_|)$ and $(M \cap V, \psi_|)$ are smoothly compatible. Hence the atlas for M obtained from these submanifold charts for N is in fact smooth, and hence determines a smooth structure on the topological manifold M. Summarizing: a submanifold M of a smooth manifold N is in fact a *smooth manifold*.

Example 6.38. We will show that the sphere $S^n \subset \mathbb{R}^{n+1}$ is a submanifold of codimension 1. It is easy to see that any hyperplane in $\mathbb{R}^{n+1}$ is a submanifold of codimension 1. The intersection of S^n and a (linear) hyperplane in $\mathbb{R}^{n+1}$ is a sphere of dimension $n - 1$, and hence it has codimension 2 in $\mathbb{R}^{n+1}$.

Definition 6.39. Let $F \colon N^n \to Q^k$ be a smooth map between manifolds of dimension n resp. k. For $p \in N$ and $q := F(p) \in Q$, let

$$F_* \colon T_pN \to T_qQ$$

be the linear map given by the differential of F. A point $p \in N$ is a *regular point of* F if the differential $F_* \colon T_pN \to T_qQ$ is surjective; p is called a *critical point of* F otherwise. A point $q \in Q$ is called a *regular value of* F if every point $p \in F^{-1}(q)$ is a regular point; q is called a *critical value* of F otherwise. We note that $F^{-1}(q) = \emptyset$ implies that q is a regular value.

Example 6.40. Let $\mathbb{R}^n \supset U \xrightarrow{F} \mathbb{R}$ be a smooth function, $p \in U$, and $q = F(p) \in \mathbb{R}$. Then there is a commutative diagram

$$\begin{tikzcd} T_pU \arrow[r, "F_*"] \arrow[d, "\cong"'] & T_q\mathbb{R} \arrow[d, "\cong"] \\ \mathbb{R}^n \arrow[r, "dF(p)"] & \mathbb{R} \end{tikzcd},$$

where $dF(p)\colon \mathbb{R}^n \to \mathbb{R}$ is the usual Jacobian of F at the point p. More explicitly, this sends a vector $v \in \mathbb{R}^n$ to the matrix product of the Jacobian matrix at the point p and the vector v. In the case at hand, the Jacobian matrix is just the gradient vector $(\mathrm{grad} F)(p)$, and $(df(p))(v) = (\mathrm{grad} F)(p) \cdot v \in \mathbb{R}$, the dot product of the gradient vector and v. In particular, p is a regular point of F if and only if $(\mathrm{grad} F)(p) \neq 0$ (argument: if $(\mathrm{grad} F)(p) \neq 0$, then some component of $(\mathrm{grad} F)(p)$, say the i^{th} component of $(\mathrm{grad} F)(p)$ is non-zero. Then the dot product $(\mathrm{grad} F)(p) \cdot e_i$ is non-zero, which implies that $dF(p)$ is surjective).

In particular, the critical points of F are the points $p \in U \subset \mathbb{R}^n$ for which the gradient $(\mathrm{grad} F)(p)$ vanishes, a statement that might be familiar from calculus. Those points are the only points in U where F might have a relative minimum or maximum.

Theorem 6.41. *Let $F\colon N^n \to Q^k$ be a smooth map of smooth manifolds N, Q of dimension n resp. k. If q is a regular value of F, then $M = F^{-1}(q)$ is a manifold of codimension k. In particular,* $\dim M = \dim N - k$.

Example 6.42. (Submanifolds of the form $M = F^{-1}(q)$ for a regular value q).

(a) $S^n = \{x \in \mathbb{R}^{n+1} \mid ||x||^2 = 1\}$; in other words, $S^n = F^{-1}(1)$ where $F\colon \mathbb{R}^{n+1} \to \mathbb{R}$ is the smooth map given by $F(x) = ||x||^2$. We claim that $1 \in \mathbb{R}$ is a regular value of F and hence $S^n \subset \mathbb{R}^{n+1}$ is a submanifold of codimension 1.

To prove the claim, we need to show that every $x \in F^{-1}(1) = S^n \subset \mathbb{R}^{n+1}$ is a regular point of F, i.e., the gradient $\mathrm{grad} F(x)$ is non-zero for $x \in S^n$. For the calculation of $\mathrm{grad} F(x)$, we write $x = (x_0, \dots, x_n) \in \mathbb{R}^{n+1}$ and

$$F(x) = ||x||^2 = x_0^2 + x_1^2 + \cdots + x_n^2.$$

Hence $\mathrm{grad} F(x) = (2x_0, \dots, 2x_n) = 2x \neq 0$ for $x \in S^n$.

(b) Let $\mathrm{SL}_2(\mathbb{R}) = \{A \in M_{n\times n}(\mathbb{R}) \mid \det(A) = 1\}$ the special linear group. This description shows that $\mathrm{SL}_2(\mathbb{R}) = F^{-1}(1)$, where $F\colon M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ is defined by $F(A) := \det(A)$. Hence to show that $\mathrm{SL}_n(\mathbb{R})$ is a codimension 1 submanifold of $M_{n\times n}(\mathbb{R})$, it suffices to show that $1 \in \mathbb{R}$ is a regular value of F, i.e., that every $A \in \mathrm{SL}_n(\mathbb{R}) = F^{-1}(1)$ is a regular point of F.

This requires a calculation of the differential $F_*\colon T_A M_{n\times n}(\mathbb{R}) \to T_{F(A)}\mathbb{R}$. This could be done as in the previous example by identifying the vector space $M_{n\times n}(\mathbb{R})$ with $\mathbb{R}^{n^2}$ and calculating the gradient vector of the function F. We prefer to avoid this identification, using instead the geometric tangent space and the bijection

$$T_A^{\mathrm{geo}} M_{n\times n}(\mathbb{R}) \xrightarrow{\cong} M_{n\times n}(\mathbb{R}) \qquad \text{given by} \qquad [\gamma] \mapsto \gamma'(0).$$

We have used this bijection before for $\mathbb{R}^n$, but this works equally well for any finite dimensional vector space, e.g., the vector space $M_{n\times n}(\mathbb{R})$. The inverse of this bijection is simply given by mapping $V \in M_{n\times n}(\mathbb{R})$ to the element $[\gamma_V] \in T_A^{\mathrm{geo}} M_{n\times n}(\mathbb{R})$ represented by the straight line path $\gamma_V\colon (-\epsilon, \epsilon) \to M_{n\times n}(\mathbb{R})$ defined by $\gamma_V(t) := A + tV$. Slightly abusing language, we write $dF(A)\colon M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ for the linear map making the following diagram commutative:

$$\begin{tikzcd}
T_A^{\mathrm{geo}} M_{n\times n}(\mathbb{R}) \arrow[r, "F_*^{\mathrm{geo}}"] \arrow[d, "\cong"'] & T_{F(A)}^{\mathrm{geo}} \mathbb{R} \arrow[d, "\cong"] \\
M_{n\times n}(\mathbb{R}) \arrow[r, "dF(A)"] & \mathbb{R}
\end{tikzcd}$$

We calculate $(dF(A))(V)$ by chasing $V \in M_{n\times n}(\mathbb{R})$ around the diagram:

$$\begin{tikzcd}
{[A+tV]} \arrow[r, mapsto] & {[F(A+tV)]} \arrow[d, mapsto] \\
V \arrow[u, mapsto] & \frac{d}{dt}_{|t=0} F(A+tV)
\end{tikzcd}$$

and hence $(dF(A))(V) = \frac{d}{dt}_{|t=0} F(A+tV)$. To calculate this derivative explicitly for our function $F(A) = \det(A)$, it will be useful to think of $\det(A)$ as a multilinear function $\det(a_1, a_2, \dots, a_n)$ of the column vectors $a_1, \dots, a_n$ of A. Similarly, writing v_i for the i^{th} column of the matrix V, we calculate:

$$F(A + tV) = \det(a_1 + tv_1, \dots, a_n + tv_n).$$

The product rule then implies

$$\frac{d}{dt} F(A+tV) = \sum_{i=1}^{n} \det(a_1 + tv_1, \dots, a_{i-1} + tv_{i-1}, v_i, a_{i+1} + tv_{i+1}, \dots, a_n + tv_n).$$

Hence

$$(dF(A))(V) = \frac{d}{dt}_{|t=0} F(A+tV) = \sum_{i=1}^{n} \det(a_1, \dots, v_i, \dots, a_n).$$

To show that the linear map $dF(A)\colon M_{n\times n} \to \mathbb{R}$ is surjective, it suffices to find some $V \in M_{n\times n}(\mathbb{R})$ such that $(dF(A))(V) \neq 0$. We note that for $V = A$ we obtain

$$(dF(A))(V) = (dF(A))(A) = \sum_{i=1}^{n} \det(a_1, \dots, a_i, \dots, a_n) = n \det(A) = n$$

for $A \in \mathrm{SL}_n(\mathbb{R})$. This shows that every $A \in \mathrm{SL}_n(\mathbb{R}) = F^{-1}(1)$ is a regular point of F, and hence $1 \in \mathbb{R}$ is a regular value of F. It follows from Theorem 6.41 that $\mathrm{SL}_n(\mathbb{R})$ is a submanifold of $M_{n\times n}(\mathbb{R})$ of codimension 1. In particular, $\mathrm{SL}_n(\mathbb{R})$ is a smooth manifold of dimension $\dim M_{n\times n}(\mathbb{R}) - 1 = n^2 - 1$.

(c) We recall that the Stiefel manifold $V_k(\mathbb{R}^n)$ of orthonormal k-frames in $\mathbb{R}^n$ is defined by

$$V_k(\mathbb{R}^n) = \{(a_1, \dots, a_k) \mid a_i \in R^n,\ ||a_i||^2 = 1,\ a_i \perp a_j \text{ for } i \neq j\}.$$

Thinking of the vectors a_i as the column vectors of a matrix $A \in M_{n\times k}(\mathbb{R})$, we recognize $V_k(\mathbb{R}^n)$ as a subset of $M_{n\times k}(\mathbb{R})$. We would like to express the conditions on the vectors a_i in terms of the matrix A. We consider the matrix product $A^tA \in M_{k\times k}(\mathbb{R})$, where $A^t \in M_{k\times n}(\mathbb{R})$ is the transpose matrix, whose *row vectors* are the vectors a_i. We observe that $(A^tA)_{ij}$ (the ij-component of the $k \times k$ matrix A^tA) is given by

$$(A^tA)_{ij} = a_i \cdot a_j,$$

the dot product of the vectors a_i and a_j. In particular, the requirement $a_i \perp a_j$ is equivalent to the vanishing of the off-diagonal entries of A^tA, and the requirement $||a_i||^2 = 1$ for all i is equivalent to the diagonal entries of A^tA being 1. If follows that

$$V_k(\mathbb{R}^n) = \{A \in M_{n\times k}(\mathbb{R}) \mid A^tA = I_k \in M_{k\times k}(\mathbb{R})\},$$

where I_k is the identity $k \times k$ matrix.

The proof of Theorem 6.41 is based on the Inverse Function Theorem. Before stating it, we would like to put it into context. Let $U, V \subset \mathbb{R}^n$ be open subsets and let $F\colon U \to V$ be a diffeomorphism. Then for any $p \in U$ the differential $F_*\colon T_pU \to T_{F(p}V$ is an isomorphism. It is natural to ask whether the converse of this statement holds as well. The answer is “no” as the following example shows.

Example 6.43. Let $F\colon \mathbb{C} \to \mathbb{C}^\times$ be the smooth map given by $F(z) := e^{2\pi iz}$. This map is not a diffeomorphism, since $F(z+1) = F(z)$ and hence F is not injective. However, it is easy to see that the differential $F_*\colon T_z\mathbb{C} \to T_{F(z)}\mathbb{C}^\times$ is an isomorphism for every $z \in \mathbb{C}$. There are a number of ways to do this. For people familiar with holomorphic functions and their derivatives, this is obvious by calculating $\frac{\partial F}{\partial z}$. Alternatively, one regard F as a smooth map $F\colon \mathbb{R}^2 \to \mathbb{R}^2 \setminus \{0\}$ and show that it Jacobian matrix is invertible at every point $(x, y) \in \mathbb{R}^2$.

We would like to illustrate that this calculation is easy to do using the geometric tangent spaces.

Domain and codomain of F are open subsets of $\mathbb{C}$ (viewed as real vector space), and hence we have a commutative diagram

$$\begin{tikzcd} T_z^{\mathrm{geo}}\mathbb{C} \arrow[r, "F_*"] \arrow[d, "\cong"'] & T_{F(z)}^{\mathrm{geo}}\mathbb{C}^\times \arrow[d, "\cong"] \\ \mathbb{C} \arrow[r, "dF(z)"] & \mathbb{C} \end{tikzcd}.$$

Here the vertical bijections are given by mapping a tangent vector $[\gamma]$ represented by a smooth path γ to $\gamma'(0)$, and its inverse is given by mapping a vector $w \in \mathbb{C}$ to the tangent vector $[z+tw] \in T_z^{\mathrm{geo}}\mathbb{C}$ represented by the straight line path $t \mapsto z+tw$. Hence we can calculate the Jacobian $dF(z)$ by chasing the vector $w \in \mathbb{C}$ around the diagram:

$$\begin{tikzcd} {[z+tw]} \arrow[r, mapsto] & {[F(z+tw)]} \arrow[d, mapsto] \\ w \arrow[u, mapsto] & \frac{d}{dt}_{|t=0}F(z+tw) \end{tikzcd}$$

Hence $(dF(z))(w) = \frac{d}{dt}_{|t=0}F(z+tw) = \frac{d}{dt}_{|t=0}e^{2\pi i(z+tw)} = (e^{2\pi i(z+tw)}2\pi i w)_{|t=0} = 2\pi i e^{2\pi i}w$. This shows that the Jacobian $dF(z)\colon \mathbb{C} \to \mathbb{C}$ is given by multiplication by the complex number $2\pi i e^{2\pi i}$. This number is non-zero for every $z \in \mathbb{C}$, and hence $dF(z)$ is invertible for every $z \in \mathbb{C}$.

We observe that the map $F\colon \mathbb{C} \to \mathbb{C}^\times$ is a covering map whose restriction to $\mathbb{R} \subset \mathbb{C}$ is our first example of a covering map $\mathbb{R} \to S^1$, $t \mapsto e^{2\pi i t}$. This shows that F is *locally* a diffeomorphism (namely by restricting it to a path component of the preimage $F^{-1}(U)$ of an evenly covered subspace $U \subset \mathbb{C}^\times$.

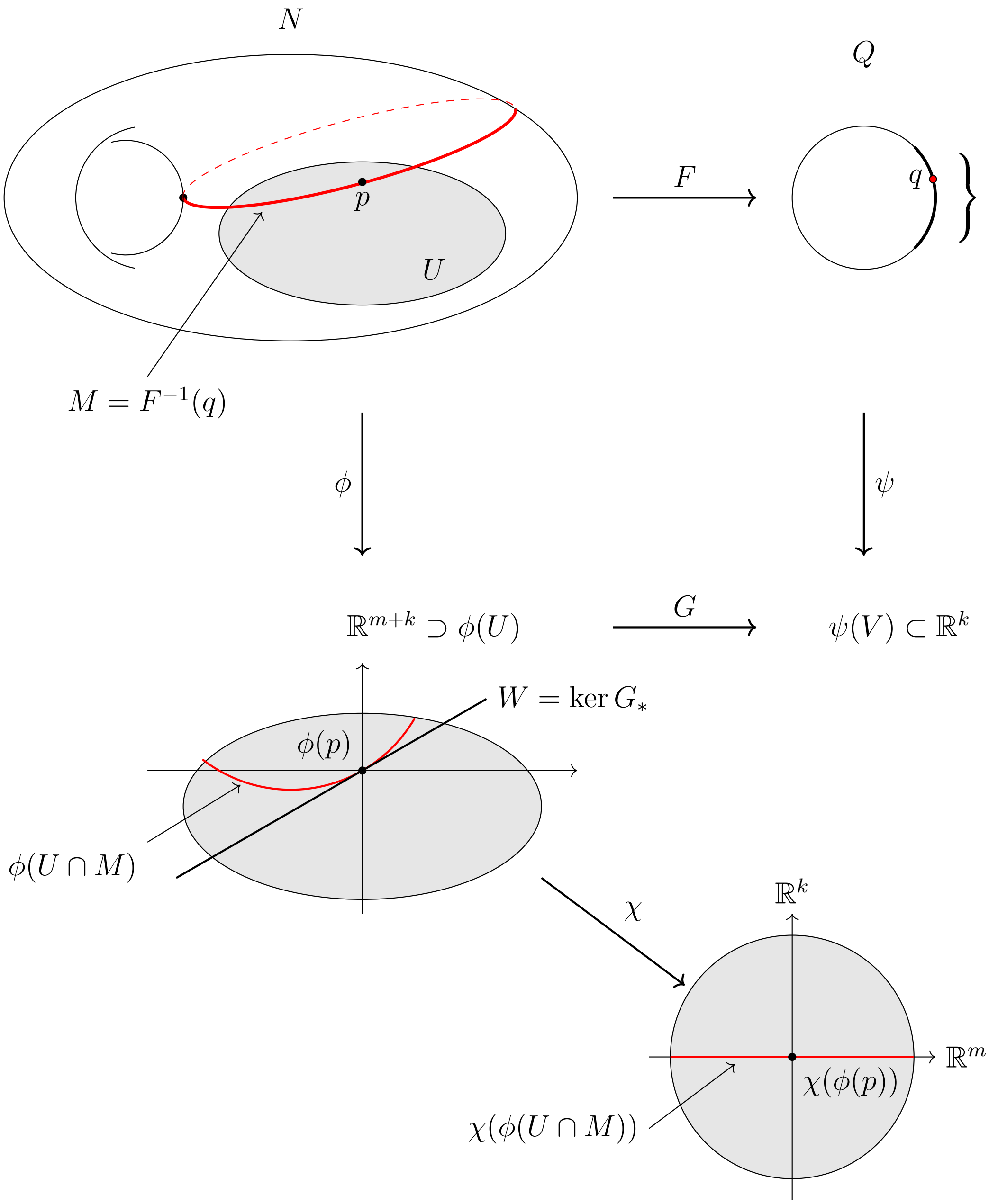
According to the Inverse Function Theorem, this is true in great generality.

Theorem 6.44. (The Inverse Function Theorem). *Let $U, V \subset \mathbb{R}^n$ be open subsets, let $F\colon U \to V$ be a smooth map, and assume that the differential $T_pM \xrightarrow{F_*} T_{F(p)}N$ is an isomorphism. Then F is a local diffeomorphism at p, i.e., there is an open neighborhood $U_x \subset U$ of x such that the restriction of F to U_x is a diffeomorphism onto is image.*

Proof of Theorem 6.41. To prove the theorem, we need to construct a submanifold chart at every point $p \in M$. The idea is to modify a smooth chart $N \supset U \xrightarrow{\phi} \mathbb{R}^n$ by post-composing it with a suitable diffeomorphism χ of open subsets of $\mathbb{R}^n$ to obtain as submanifold chart $\chi \circ \phi$.

Let $N \supset U \xrightarrow{\phi} \mathbb{R}^n$ be a smooth chart with $p \in U$, and $Q \subset V \xrightarrow{\psi} \mathbb{R}^k$ be a smooth chart with $q \in V$. Without loss of generality we can assume $\phi(p) = 0$, $\psi(q) = 0$ (by post-composing with a translation) and $F(U) \subset V$ (by restricting the diffeomorphism ϕ to a smaller open neighborhood of p). Let $G\colon \phi(U) \to \psi(V)$ be the smooth map defined by $G := \psi \circ F \circ \phi^{-1}$.

The situation is depicted in the following figure.



In this picture the level set $M = F^{-1}(q)$ and $\phi(U \cap M)$ are drawn in red. To construct a submanifold chart of $M \subset N$ at $p \in M$ it suffices to find a submanifold chart for $\phi(U \cap M) \subset \phi(U)$ at $0 \in \phi(U) \subset \mathbb{R}^{m+k}$. In other words, we need to find a smooth map

$$\chi\colon \phi(U) \longrightarrow \mathbb{R}^m \times \mathbb{R}^k$$

satisfying the following properties:

(a) $\chi(\phi(U \cap M)) \subset \mathbb{R}^m = \mathbb{R}^m \times \{0\} \subset \mathbb{R}^m \times \mathbb{R}^k$, and

(b) the restriction of χ to a neighborhood of 0 is a diffeomorphism onto its image.

Let $\chi^1\colon \phi(U) \to \mathbb{R}^m$, $\chi^2\colon \phi(U) \to \mathbb{R}^k$ be the component maps of χ, i.e.,

$$\chi(x) = (\chi^1(x), \chi^2(x)) \in \mathbb{R}^m \times \mathbb{R}^k \qquad \text{for all } x \in \phi(U).$$

Then the condition (a) can be satisfied by defining χ^2 to be the smooth function G, which by construction has the property $G(x) = 0$ if and only if $x \in \phi(U \cap M)$. For the construction of χ_1 we note that by the Inverse Function Theorem 6.44, condition (b) holds provided the differential of χ at 0 is an isomorphism.

To construct χ^1, let W be the kernel of the differential $G_*\colon T_0\phi(U) \to T_0\psi(V)$. To determine the dimension of W we use the commutative diagram of differentials

$$\begin{array}{ccc} T_pU & \xrightarrow{F_*} & T_qV \\ {\scriptstyle \phi_*}\big\downarrow{\scriptstyle \cong} & & {\scriptstyle \cong}\big\downarrow{\scriptstyle \psi} \\ T_0\phi(U) & \xrightarrow{G_*} & T_0\psi(V). \end{array}$$

Here the vertical maps isomorphisms since they are the differentials of the diffeomorphisms ϕ resp. ψ. The differential F_* is surjective by the assumption that q is a regular value for F, and hence $p \in F^{-1}(q)$ is a regular point of F. Consequently, the differential G_* is also surjective, and G_* induces an isomorphism $T_0\phi(U)/\ker G_* \cong T_0\psi(V)$. Hence

$$k = \dim T_0\psi(V) = \dim T_0\phi(U) - \dim\ker G_* = n - \dim W = m + k - \dim W,$$

which implies $\dim W = m$. Let χ_1 be the composition

$$\mathbb{R}^{m+k} \supset \phi(U) \xrightarrow{\pi} W \xrightarrow[\cong]{h} \mathbb{R}^m,$$

where π is the orthogonal projection from $\mathbb{R}^{m+k}$ onto its subspace W, and $h\colon W \to \mathbb{R}^m$ is any linear isomorphism.

It only remains to show that the differential $\chi_*\colon T_0\phi(U) \to T_{(0,0)}(\mathbb{R}^m \times \mathbb{R}^n)$ is an isomorphism, since then the Inverse Function Theorem 6.44 implies that then the restriction of χ

to an open neighborhood of $0 \in \phi(U)$ is a diffeomorphism onto its image. To prove this, let $\pi^1 \colon \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m$ and $\pi^2 \colon \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^k$ be the projection maps and consider the composition

$$T_0\phi(U) \xrightarrow{\chi_*} T_{(0,0)}(\mathbb{R}^m \times \mathbb{R}^k) \underset{\cong}{\xrightarrow{\pi^1_* \times \pi^2_*}} T_0\mathbb{R}^m \times T_0\mathbb{R}^k$$

By a homework problem the second map is an isomorphism, and hence

$$\begin{aligned} \ker \chi_* &= \ker(\pi^1_* \circ \chi_*) \cap \ker(\pi^2_* \circ \chi_*) = \ker \chi^1_* \cap \ker \chi^2_* \\ &= \ker \chi^1 \cap \ker G_* = \ker(\chi^1_{|W}) = \ker(h) = 0. \end{aligned}$$

We comment that the differential χ^1_* is equal to χ^1 since χ^1 is a linear map. This shows that χ_* is a monomorphism, and hence an isomorphism since domain and codomain of χ_* are vector spaces of the same dimension. □

Let N be smooth manifold of dimension n, and let $M \subset N$ be a submanifold of dimension m. We note that the inclusion map $i \colon M \to N$ is *smooth*; to check smoothness at a point $p \in M$, we can use a smooth chart $N \supset U \xrightarrow{\phi} \mathbb{R}^n$, which is a submanifold chart in the sense of Definition 6.36. This in particular implies that the restriction

$$\phi_{|U\cap M} \colon U \cap M \to \mathbb{R}^m \subset \mathbb{R}^n$$

i s a smooth chart for the submanifold M. Then the composition

$$\mathbb{R}^m \cap \phi(U) \xrightarrow{\phi^{-1}_{|U\cap M}} U \cap M \xrightarrow{i} U \xrightarrow{\phi} \phi(U)$$

is the inclusion map $\mathbb{R}^m \cap \phi(U) \hookrightarrow \phi(U) \subset \mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^k$ and hence smooth.

Lemma 6.45. *Let N be a smooth manifold, $M \subset N$ submanifold and $i \colon M \to N$ the inclusion map. Then for $p \in M$ the differential*

$$T_pM \xrightarrow{i_*} T_pN$$

is a monomorphism.

This shows that we can use i_* to identify T_pM with a subspace of T_pN, given by the image of i_*.

Proof. Let $[\gamma], [\gamma'] \in T^{\mathrm{geo}}_pM$ be two tangent vectors represented by smooth paths $\gamma \colon (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ resp. $\gamma' \colon (-\epsilon', \epsilon') \to M$ with $\gamma'(0) = p$. We note that $i_*[\gamma], i_*[\gamma'] \in T^{\mathrm{geo}}_pN$ are represented by the same paths $i \circ \gamma$ resp. $i \circ \gamma'$. Using the submanifold chart (U, ϕ), we note that

$$\phi \circ i \circ \gamma = \phi_{|U\cap M} \circ \gamma \qquad \text{and} \qquad \phi \circ i \circ \gamma' = \phi_{|U\cap M} \circ \gamma'.$$

In particular,

$$\begin{aligned} i_*[\gamma] = i_*[\gamma'] &\iff (\phi \circ i \circ \gamma)'(0) = (\phi \circ i \circ \gamma')'(0) \\ &\iff (\phi_{|U\cap M} \circ \gamma)'(0) = (\phi_{|U\cap M} \circ \gamma')(0) \iff [\gamma] = [\gamma'] \end{aligned}$$

□

Proposition 6.46. *Let $F\colon N \to Q$ be a smooth map, and let $M := F^{-1}(q) \subset N$ be the submanifold given by a regular value $q \in Q$ of F. Then*

$$T_pM = \ker\left(T_pN \xrightarrow{F_*} T_qQ\right).$$

Proof. Let us first check that the subspaces T_pM and $\ker F_*$ of T_pN have the same dimension. Let $m = \dim M$, $n = \dim N$, and $k = \dim Q$. Since M is a submanifold of codimension k in N, we have $m + k = n$. The assumption that $q \in Q$ is a regular value of F in particular implies that $p \in F^{-1}(q)$ is a regular point of F, i.e., that the differential $T_pN \xrightarrow{F_*} T_qQ$ is surjective. Hence T_qQ, the image of F_* is isomorphic to its domain modulo its kernel, i.e., $T_pN/\ker F_*$, and so

$$k = \dim T_qQ = \dim(T_pN/\ker F_*) = \dim T_pN - \dim\ker F_* = n - \dim\ker F_*.$$

It follows that $\dim\ker F_* = n - k = m$, which matches with $\dim T_pM = \dim M = m$.

So the two subspaces T_pM and $\ker F_*$ of T_pN have the same dimension, and hence it suffices to show that one of these subspaces is contained in the other.

To show that T_pM is contained in $\ker F_*$, it suffices to show that the composition

$$T_pM \xrightarrow{i_*} T_pN \xrightarrow{F_*} T_qQ$$

is the zero map. Using the geometric description of the tangent spaces, let $\gamma\colon (-\epsilon, \epsilon) \to M$ be a smooth path with $\gamma(0) = p$ representing an element $[\gamma] \in T_pM$. Then $F_*(i_*[\gamma]) = [F \circ i \circ \gamma] \in T_qQ$ is the trivial element, since the composition $F \circ i \circ \gamma$ is the constant path with value q in Q. □

The proposition above enables us to calculate the tangent spaces of submanifolds given as preimages of regular values, e.g., in all the examples 6.42.

Example 6.47. Let $\mathbb{R}^{n+1} \supset U \xrightarrow{f} \mathbb{R}$ be a smooth map with regular value $q \in \mathbb{R}$. Then $M := f^{-1}(q)$ is a codimension 1 submanifold of $\mathbb{R}^{n+1}$. The tangent space T_pM at a point $p \in M$ is by Proposition 6.46 equal to the kernel of $f_*\colon T_0U \to T_0\mathbb{R}$. The differential f_* can be identified with the Jacobian of f at the point p, or equivalently with the map

$$\mathbb{R}^{n+1} \longrightarrow \mathbb{R} \qquad \text{given by} v \mapsto (\mathrm{grad} f)_p \cdot v.$$

In particular, $T_pM = \ker f_* = \{v \in \mathbb{R}^{n+1} \mid (\mathrm{grad} f)_p \cdot v = 0\} = \{v \in \mathbb{R}^{n+1} \mid v \perp (\mathrm{grad} f)_p\}$, the subspace of $\mathbb{R}^{n+1}$ consisting of all vectors v perpendicular to the gradient vector of f at the point p.

7 Smooth vector bundles

The goal of this section is to define the notion of smooth section of a vector bundle E over a smooth manifold M. We will begin with examples motivating the need for this notion, follow it up with a preliminary definition which captures some but not all the required features and end up with the technical Definition **??** of smooth vector bundles and their sections.

Example 7.1. Let M be a smooth manifold and let $\gamma\colon \mathbb{R}\times M \to M$ be a smooth action of the group $\mathbb{R}$. For $p \in M$ let $\gamma_p\colon \mathbb{R}\to M$ be the smooth path given by $\gamma_p(t) := \gamma(t,p)$. The path γ_p represents an element $[\gamma_p] \in T_p^{\mathrm{geo}}M$ in the tangent space of M at p, using the geometric description of the tangent space. The assignment

$$M \ni p \ \mapsto\ [\gamma_p] \in T_pM$$

is an example of a *vector field* on M in the sense of the following definition.

Definition 7.2. (Preliminary!) A vector field on a smooth manifold M is an assignment X that assigns to any point $p \in M$ a tangent vector $X(p) \in T_pM$.

Example 7.3. Let $f\colon M \to \mathbb{R}$ be a smooth function on a smooth manifold M. For $p \in M$, let

$$df_p = f_*\colon T_pM \longrightarrow T_{f(p)}\mathbb{R} = \mathbb{R}$$

be the differential of f at the point p (as usual we identify here the tangent space $T_q\mathbb{R}^n$ at $q \in \mathbb{R}^n$ with the vector space $\mathbb{R}^n$). We recall that for a vector space V the space $\mathrm{Hom}(V,\mathbb{R})$ is called the *dual vector space*and is denoted by V^*. In particular, df_p is an element of $\mathrm{Hom}(T_pM,\mathbb{R}) = (T_pM)^*$, which is called the *cotangent space of M at p* and is denoted by T_p^*M. The assignment df given by

$$M \ni p \ \mapsto\ df_p \in T_p^*M,$$

is called the *differential of f*. It is an example of a 1-form, defined as follows.

Definition 7.4. (Preliminary!) A 1-*form* α on a smooth manifold M is an assignment

$$M \ni p \ \mapsto\ \alpha_p \in T_p^*M.$$

Extracting the commonality of these examples, we make the following (preliminary!) definition.

Definition 7.5. (Preliminary!) A *vector bundle E of rank k over a smooth manifold M* is a family $\{E_p\}_{p\in M}$ of vector space E_p of dimension k parametrized by points p in M. The vector space E_p is called the *fiber over p*. A *section s* of E is an assignment

$$M \ni p \ \mapsto\ s(p) \in E_p.$$

Example 7.6. Let V be a vector space of dimension k. Then the vector bundle E given by $E_p = V$ for all $p \in M$ is called the *trivial vector bundle over M with fiber V*. We note that a section of this bundle is simply a map $M \to V$ with values in the vector space M. So a section s of a general vector bundle E should be viewed as generalization of vector-valued function on M, whose value $s(p)$ at a point $p \in M$ is a vector $s(p)$ in a vector space E_p, which for a general vector bundle *depends on the point p.*

Question. What is missing in the above definitions of vector field, 1-form and section?

To see what is missing, we revisit the examples 7.1 and 7.3 in the special case where the smooth manifold M is an open subset of $\mathbb{R}^n$.

- $\mathbb{R}^n \supset M \ni p \mapsto \gamma_p'(0) \in T_pM = \mathbb{R}^n$ is an $\mathbb{R}^n$-valued function. This is a *smooth* map, as can be seen by rewriting $\gamma_p'(0)$ in the form
$$\gamma_p'(0) = \frac{\partial \gamma}{\partial t}(0, p_1, \ldots, p_n)$$
- Similarly, $df_p \in T_p^*M = \mathrm{Hom}(T_pM, \mathbb{R}) = \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^* \cong \mathbb{R}^n$ using our standard identification $T_pM = \mathbb{R}^n$ for open subsets $M \subset \mathbb{R}^n$ and the isomorphism $(\mathbb{R}^n)^* \cong \mathbb{R}^n$, under which the standard basis vector $e_i \in \mathbb{R}^n$ corresponds to the i-th vector $e^i \in (\mathbb{R}^n)^*$ of the dual basis $\{e^i\}_{i=1,\ldots,n}$ of the dual space $(\mathbb{R}^n)^*$. It is easy to check that via this isomorphism, the cotangent vector $df_p \in T_p^*M$ corresponds to $(\mathrm{grad} f)_p \in \mathbb{R}^n$, which is a *smooth function* of p.

These examples hopefully show that the answer to the above question is that for a section
$$M \ni p \mapsto s(p) \in E_p$$
of a vector bundle $\{E_p\}_{p \in M}$ we should require that $s(p)$ depends "smoothly" on the point $p \in M$. The problem is that it is not clear what this should mean: the domain of s is a smooth manifold, but:

1. What is the codomain of s?
2. How can we ensure that the codomain is a manifold? (which would allow us to require that s is a smooth map)

The first question can be addressed as follows. Let E be the set given by the disjoint union of the vector spaces E_p. More explicitly, this disjoint union is the set of pairs (p, v) with $p \in M$ and $v \in E_p$:
$$E := \coprod_{p \in M} E_p = \{(p, v) \mid p \in M \text{ and } v \in E_p\}.$$

Then a section s of E determines a map $f\colon M \to E$ given by $p \mapsto (p, s(p))$. We note this construction gives a bijection

$$\{\text{sections } s \text{ of } \{E_p\}_{p\in M}\} \longleftrightarrow \{f\colon M \to E \mid \pi \circ f = \mathrm{id}_M\},$$

where $\pi\colon E \to M$ is the projection map $(p, v) \mapsto p$. The right hand side is a very convenient way to think about sections, and hence from now on, a section will be a map $s\colon M \to E$ with $\pi \circ s = \mathrm{id}_M$. In particular, this answers the first question above: the codomain of a section s is the set E.

With regards to the second question, the simplest solution is to *require* that E is a smooth manifold as part of the definition of a smooth vector bundle.

Definition 7.7. Let M be a smooth manifold. A *smooth vector bundle of rank k over M* consists of the following data:

1. A smooth manifold E, called the *total space* and a smooth map $\pi\colon E \to M$.
2. For each $p \in M$ the set $E_p := \pi^{-1}(p)$, called *the fiber over p*, has the structure of a k-dimensional vector space.

It is required that E is *locally trivial* in the sense that for each point $p \in M$, there is an open neighborhood U and a diffeomorphism Φ making the diagram

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

commutative, such that the map $E_p \to \{p\} \times \mathbb{R}^k = \mathbb{R}^k$ given by restriction of Φ is a vector space isomorphism for each $p \in U$. The map Φ is called a *local trivialization* of E. We note that this implies in particular that $\dim E = \dim(U \times \mathbb{R}^k) = \dim U + \dim \mathbb{R}^k = \dim M + k$.

A *section* is a map $s\colon M \to E$ with $\pi \circ s = \mathrm{id}_M$; in other words, $s(p)$ belongs to the fiber E_p for every $p \in M$. A section s is *smooth* if $s\colon M \to E$ is a smooth map. The vector space of smooth sections of E will be denoted $\Gamma(M, E)$.

Remark 7.8. The above definition is a particular variant, the "smooth" variant, of a general definition. Other variants are:

The "continuous" variant. Here M, E is a just topological spaces (instead of smooth manifolds), $\pi\colon E \to M$ is a continuous map (instead of smooth), and Φ is a homeomorphism (instead of a diffeomorphism). A *continuous section* is a continuous map $s\colon M \to E$ such that $\pi \circ s = \mathrm{id}_M$.

The "holomorphic" variant. Here M, E are complex manifolds, $\pi\colon E \to M$ is a holomorphic map, and Φ is a holomorphic map with a holomorphic inverse. A *holomorphic section* is a holomorphic map $s\colon M \to E$ with $\pi \circ s = \mathrm{id}_M$.

Since complex manifold are in particular smooth manifolds, and holomorphic maps are smooth maps, we can "forget" the holomorphic structure on a holomorphic vector bundle $E \to M$ and just regard it as a smooth vector bundle. Similarly, if $s\colon M \to E$ is a holomorphic section, it is in particular a smooth section of the underlying smooth vector bundle.

Analogously, a smooth vector bundle $E \to M$ can in particular be considered as a vector bundle over the topological space M, and a smooth section $s\colon M \to E$ is a continuous section.

Example 7.9. (Examples of smooth vector bundles).

1. Let M be a smooth manifold and let V be a finite dimensional vector space. Then $E = M \times V$ equipped with the projection map $\pi\colon E \to M$ is a smooth vector bundle over M called the *trivial vector bundle with fiber V*. It is clear that the product $E = M \times V$ is a smooth manifold, and that the projection map π is smooth. Each fiber $E_p = \pi^{-1}(p) = \{p\} \times V = V$ also has an obvious vector space structure. To show that E is locally trivial, we choose $U = M$, pick a vector space isomorphism $h\colon V \xrightarrow{\cong} \mathbb{R}^k$ (which always exists for $k = \dim V$) and define
$$\Phi\colon E_{|U} = M \times V \longrightarrow M \times \mathbb{R}^k \qquad (p, v) \mapsto (p, h(v))$$
This map satisfies all requirements of a locally trivialization.

2. Let U_1, U_2 be the open subsets of S^1 defined by $U_1 := S^1 \setminus \{-1\}$ and $U_2 := S^1 \setminus \{1\}$. Let E be the quotient of the disjoint union
$$U_1 \times \mathbb{R} \quad \amalg \quad U_2 \times \mathbb{R} \tag{7.10}$$
modulo the equivalence relation $\sim$ defined by
$$(1, z, x) \sim (2, z, \epsilon(z)x) \qquad \text{for } z \in U_1 \cap U_2 \text{ and} \quad \epsilon(z) := \begin{cases} +1 & \text{for } \mathrm{im}(z) > 0 \\ -1 & \text{for } \mathrm{im}(z) < 0 \end{cases}$$
Here $(1, z, x) \in U_1 \times \mathbb{R}$ and $(2, z, x) \in U_2 \times \mathbb{R}$, i.e., the number in the first component just indicates whether (z, x) is to be considered as an element of the first or the second summand in the disjoint union (7.10). Moreover, $\mathrm{im}(z)$ is the imaginary part of $z \in U_1 \cap U_2 \subset S^1 \subset \mathbb{C}$. The map $\pi\colon E \to S^1$ given by $[1, z, x] \mapsto z$ and $[2, z, x] \mapsto z$ is a well-defined continuous map (by the continuity property of maps out of quotients; the pre-composition of π with the projection map from $(U_1 \times \mathbb{R}) \amalg (U_2 \times \mathbb{R})$ to the quotient E is clearly continuous).

We construct bundle charts (U_1, Φ_1) and (U_2, Φ_2) for E as follows:

$$\Phi_i \colon E_{|U_i} \longrightarrow U_i \times \mathbb{R} \quad \text{given by} \quad \Phi_i[i, z, x] = (z, x) \tag{7.11}$$

Obviously, the restriction of Φ_i to each fiber E_z for $z \in U_i$ is a vector space isomorphism, and it is not hard to check that the maps Φ_i and their inverses are continuous, and so Φ_i is a homeomorphism. The only thing not clear is why Φ is a diffeomorphism; in fact, it is not even clear what that would mean, since we haven't constructed a smooth structure on E!

So our goal is to construct a smooth atlas for E in such a way that the maps Φ_i are diffeomorphisms. We observe that the homeomorphisms $\Phi_i \colon E_{|U_i} \xrightarrow{\approx} U_i \times \mathbb{R}$ can essentially be thought of as *charts* for E. This is not literally true, since U_i is an open subset of S^1 rather than an open subset of $\mathbb{R}$. However, $U_i \subset S^1$ is diffeomorphic to an open subset of $\mathbb{R}$, e.g. the map $(-1, 1) \to U_1$, $t \mapsto e^{\pi i t}$ is a diffeomorphism; similarly for U_2. Secretly composing with these diffeomorphisms, we will allow ourselves to think of Φ_1, Φ_2 as charts for E. To show that $\{(E_{U_1}, \Phi_1), (E_{U_2}, \Phi_2)\}$ is a smooth atlas, we need to check that the transition maps are smooth. For example, $\Phi_2 \circ \Phi_1^{-1}$ is given explicitly as follows:

$$\begin{array}{ccccc} (U_1 \cap U_2) \times \mathbb{R} & \xrightarrow{\Phi_1^{-1}} & E_{|U_1 \cap U_2} & \xrightarrow{\Phi_2} & (U_1 \cap U_2) \times \mathbb{R} \\ \\ (z, x) & \longmapsto & [1, z, x] = [2, z, \epsilon(z)x] & \longmapsto & (z, \epsilon(z)x) \end{array}$$

We note that this map is locally constant; in particular, it is smooth. It is equal to its own inverse inverse, and hence it is a diffeomorphism, which proves that $\{(E_{U_1}, \Phi_1), (E_{U_2}, \Phi_2)\}$ is a smooth atlas. As we have argued before, each chart of a smooth atlas is a diffeomorphisms between an open subset of the manifold and its image, which is an open subset of Euclidean space. In particular, the maps Φ_i of (7.11) are diffeomorphisms.

Lemma 7.12. (Vector Bundle Construction Lemma). *Let M be a smooth manifold of dimension n, and let $\{E_p\}$ be a collection of vector spaces parametrized by $p \in M$. Let E be the set given by the disjoint union of all these vector spaces, which we write as*

$$E := \coprod_{p \in M} E_p = \{(p, v) \mid p \in M,\ v \in E_p\}$$

and let $\pi \colon E \to M$ be the projection map defined by $\pi(p, v) = p$. Let $\{U_\alpha\}_{\alpha \in A}$ be an open cover of M, and let for each $\alpha \in A$, let $\Phi_\alpha \colon \pi^{-1}(U_\alpha) \longrightarrow U_\alpha \times \mathbb{R}^k$ be maps with the following properties

(i) *The diagram*

$$\begin{tikzcd} E_{|U_\alpha} := \pi^{-1}(U_\alpha) \arrow[rr, "\Phi_\alpha"] \arrow[dr, "\pi"'] & & U_\alpha \times \mathbb{R}^k \arrow[dl, "\pi_1"] \\ & U_\alpha & \end{tikzcd}$$

is commutative, where π_1 is the projection onto the first factor.

(ii) *For each $p \in U_\alpha$, the restriction of Φ_α to $E_p = \pi^{-1}(p)$ is a vector space isomorphism between E_p and $\{p\} \times \mathbb{R}^k = \mathbb{R}^k$ (which implies that Φ_α is a bijection).*

(iii) *For $\alpha, \beta \in A$, the composition*

$$(U_\alpha \cap U_\beta) \times \mathbb{R}^k \xrightarrow{\Phi_\alpha^{-1}} \pi^{-1}(U_\alpha \cap U_\beta) \xrightarrow{\Phi_\beta} (U_\alpha \cap U_\beta) \times \mathbb{R}^k$$

is smooth.

Then the total space E can be equipped with the structure of a smooth manifold of dimension $n+k$ such that $\pi\colon E \to M$ is a smooth vector bundle of rank k with local trivializations Φ_α.

Remark 7.13. We will apply this lemma to construct various smooth vector bundles over a smooth manifold M, in particular, the tangent bundle TM and the cotangent bundle T^*M (whose fiber over $p \in M$ is a dual of the tangent bundle). Later, we will be interested in other bundles, and bundles In some situations, e.g., when we use this lemma to construct the cotangent bundle of a smooth manifold M For the proof we will be using the Vector Bundle Construction Lemma 7.12, according to which it suffices to construct an atlas of bundle charts (aka local trivializations)

$$\begin{tikzcd} E_{|U_\alpha} \arrow[rr, "\Phi_\alpha"] \arrow[dr, "\pi"'] & & U_\alpha \times \mathbb{R}^k \arrow[dl, "\pi_1"] \\ & U_\alpha & \end{tikzcd}$$

that are smoothly compatible in the sense that all transition maps are smooth. Sometimes, for example for the proof of the lemma above, it will be convenient to replace the vector space $\mathbb{R}^k$ by some vector space V of dimension k, e.g., the dual space $(\mathbb{R}^k)^*$. Of course, any such vector space is isomorphic to $\mathbb{R}^k$, and we can just compose all local trivializations by the chosen isomorphism, but it is akward to carry that isomorphism through the calculations.

7.1 The tangent bundle of a smooth manifold

7.2 The cotangent bundle as the dual of the tangent bundle

Let M be a smooth manifold. We recall that the cotangent space of M at a point $p \in M$ is the vector space T_p^*M dual to the tangent space T_pM, i.e., $T_p^*M = \mathrm{Hom}(T_pM, \mathbb{R})$, the vector space of linear maps from T_pM to $\mathbb{R}$. The goal of this section is to show that the collection of cotangent spaces $\{T_p^*M\}_{p\in M}$ assemble to a smooth vector bundle over M, called the *cotangent bundle*. More generally, we will show that if $E \to M$ is a smooth vector bundle, then the collection $\{E_p^*\}_{p\in M}$ of the vectorspaces dual to the fibers E_p of E assemble into a smooth vector bundle $E^* \to M$.

Digression on dual vector spaces. Let V, W be vector spaces, and let $V^* := \mathrm{Hom}(V, \mathbb{R})$, $W^* = \mathrm{Hom}(W, \mathbb{R})$ be their dual vector spaces. Then a linear map $F \colon V \to W$ induces a *dual map*

$$F^* \colon W^* \longrightarrow V^* \qquad \text{given by} \qquad W^* \ni (W \xrightarrow{g} \mathbb{R}) \mapsto (V \xrightarrow{F} W \xrightarrow{g} \mathbb{R}) \in V^*;$$

in other words, $F^*(g) = g \circ F$. If $G \colon W \to X$ is a linear map, then the following diagram commutes:

$$\begin{tikzcd} V^* & W^* \arrow[l, "F^*"'] & X^* \arrow[l, "G^*"'] \arrow[ll, bend left, "(G\circ F)^*"] \end{tikzcd} ,$$

i.e., $(G \circ F)^* = F^* \circ G^*$.

We remark that this statement has a categorical interpretation: the assignment

$$V \mapsto V^* \qquad \text{and} \qquad (V \xrightarrow{F} W) \mapsto (V^* \xleftarrow{F^*} W^*)$$

is a contravariant function $* \colon \mathsf{Vect} \to \mathsf{Vect}$.

There is an *evaluation map*

$$\mathrm{ev} \colon V^* \times V \longrightarrow \mathbb{R} \qquad \text{given by} \qquad (V \xrightarrow{g} \mathbb{R}, v) \mapsto g(v).$$

Let $\{b_i\}_{i=1,\dots,n}$ be a basis for the vector space V. Then there is a *dual basis* $\{b^i\}_{i=1,\dots,m}$ for the dual vector space V^* characterized by the property

$$b^i(b_j) = \delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Let $F \colon V \to W$ be a linear map, and assume that $\{b_i\}_{i=1,\dots,m}$ is a basis for V, and $\{c_j\}_{j=1,\dots,n}$ is a basis for W. Then the linear map F corresponds to a matrix $M_F \in M_{n\times m}(\mathbb{R})$. Using the dual bases $\{b^i\}$ for V^* and $\{c_j\}$ for W^*, the dual map $F^* \colon W^* \to V^*$ corresponds to a matrix $M_{F^*} \in M_{m\times n}(\mathbb{R})$. An easy calculation shows $M_{F^*} = M_F^t$, the *transpose* of the matrix M_F. We note that for composable matrices A, B we have $(AB)^t = B^tA^t$, which reflects the property $(F \circ G)^* = G^* \circ F^*$ for linear maps.

Lemma 7.14. *Let $E \to M$ be a smooth vector bundle. Then there is a smooth vector bundle $E^* \to M$ whose fiber $(E^*)_p$ over p is the dual of the fiber E_p.*

The vector bundle $E^* \to M$ is called the *vector bundle dual to $E \to M$.*

Remark 7.15. For the proof we will be using the Vector Bundle Construction Lemma 7.12, according to which it suffices to construct an atlas of bundle charts (aka local trivializations)

$$\begin{tikzcd} E_{|U_\alpha} \arrow[rr, "\Phi_\alpha"] \arrow[dr, "\pi"'] & & U_\alpha \times \mathbb{R}^k \arrow[dl, "\pi_1"] \\ & U_\alpha & \end{tikzcd}$$

that are smoothly compatible in the sense that all transition maps are smooth. Sometimes, for example for the proof of the lemma above, it will be convenient to replace the vector space $\mathbb{R}^k$ by some vector space V of dimension k, e.g., the dual space $(\mathbb{R}^k)^*$. Of course, any such vector space is isomorphic to $\mathbb{R}^k$, and we can just compose all local trivializations by the chosen isomorphism, but it is akward to carry that isomorphism through the calculations.

Proof. Let E^* be the set defined by

$$E^* := \coprod_{p\in M} E_p^* = \{(p,g) \mid p \in M,\ g \in E_p^*\},$$

equipped with the obvious projection map π_{E^*} to M. Let $(U_\alpha, \Phi_\alpha)_{\alpha\in A}$ be a smooth bundle atlas for E, i.e.,

$$\Phi_\alpha \colon E_{|U_\alpha} \longrightarrow U_\alpha \times \mathbb{R}^k$$

is a diffeomorphism compatible with the projection map to U_α which restricts to a linear isomorphism $\Phi_{\alpha,p} \colon E_p \to \mathbb{R}^k$ on each fiber. $\square$

We will use the Vector Bundle Construction Lemma 7.12 to

7.3 Expressing vector fields and 1-forms in local coordinates

For calculations with vector spaces it is often convenient to choose a basis $\{b_i\}_{i=1,\dots,n}$ for the vector space in question and to express vectors $v \in V$ as linear combinations $v = \sum_{i=1}^n v^i b_i$

with $v^i \in \mathbb{R}$. The goal of this section is to show that similarly a vector field A or 1-form ω on a smooth n-manifold M can locally be expressed as linear combinations

$$A_{|U} = \sum_{i=1}^{n} A^i \frac{\partial}{\partial y^i} \qquad \omega = \sum_{i=1}^{n} \omega^i dy^i.$$

Here

(a) $U \subset M$ is an open subset which is the domain of a smooth chart $\phi\colon U \to \mathbb{R}^n$ of M,

(b) $y^i = \phi^* x^i \in C^\infty(U)$ are the smooth functions obtained by pulling back the smooth functions $x^i\colon \mathbb{R}^n \to \mathbb{R}$ given by $(x_1, \dots, x_n) \mapsto x_i$,

(c) $dy^i \in \Gamma(U, T^*M) = \Omega^1(U)$ is the 1-form given by the differential of the function $y^i \in C^\infty(U)$,

(d) $\frac{\partial}{\partial y^i} \in \Gamma(U, TM)$ is the collection of vector fields on U which are dual to the 1-forms dy^i (explained below), and

(e) A^i and ω^i are smooth functions on U.

We begin our discussion with introducing the usual notation for the standard basis of the tangent resp. cotangent space at a point q of an open subset $V \subset \mathbb{R}^n$.

The standard basis for T_qV.

The standard basis for T_q^*V.

7.4 Measurements in manifolds

Let U be an open subset of $\mathbb{R}^n$ and let $\gamma\colon [a, b] \to U$ be a smooth path. Then the length of the path γ is given by

$$\text{length}(\gamma) = \int_1^b ||\gamma'(t)|| dt, \tag{7.16}$$

where $||\gamma'(t)||$ is the norm of the tangent vector $\gamma'(t)$ of the path at the point $\gamma(t) \in U$. If γ is a smooth path in a manifold M, we would like to calculate the length of γ in a similar way. For each $t \in [a, b]$ the tangent vector $\gamma'(t)$ belongs to the tangent space $T_{\gamma(t)}M$, and so the question is how to make sense of the norm $||\gamma'(t)||$. We recall that the usual devise to make sense of the norm $||v|| \in [0, \infty)$ of a vector v of a vector space V is the following.

Definition 7.17. An *inner product* on a vector space V is a map $g\colon V \times V \to \mathbb{R}$ with the following properties:

(i) **multilinear:** g is a linear function in each of its two slots;

(ii) **symmetric:** $g(v, w) = g(w, v)$ for $v, w \in V$;

(iii) **positive definite:** $g(v, v) \geq 0$ for all $v \in V$ and $g(v, v) = 0$ if and only if $v = 0$.

A map $g\colon V \times V \to \mathbb{R}$ satisfying (i) and (ii) is called a *symmetric bilinear form on* V. The set of all symmetric bilinear forms is a vector space which is denoted $\mathsf{Sym}^2(V; \mathbb{R})$.

The usual scalar product on $\mathbb{R}^n$ given by $g(v, w) = v_1 w_1 + \cdots + v_n w_n$ for $v, w \in \mathbb{R}^n$, $v = (v_1, \ldots, v_n)$, $w = (w_1, \ldots, w_n)$, is an inner product on $\mathbb{R}^n$. The scalar product on $\mathbb{R}^n$ allows us to calculate the length $||v||$ of a vector $v \in \mathbb{R}^n$ or the angle $\alpha(v, w) \in [0, \pi]$ between vectors $v, w \in \mathbb{R}^n$. Similarly, an inner product g on a vector space V allows us to do the same for vectors $v, w \in V$ by defining:

$$||v|| := \sqrt{g(v, v)} \qquad \cos\alpha(v, w) := \frac{g(v, w)}{||v||||w||}.$$

So an inner product on a vector space V should be thought of as a "yard stick" making it possible to do measurements of lengths and angles in V. In particular in order to talk about the norm of tangent vectors of a smooth manifold M, we need an inner product g_p on the tangent space T_pM for all points $p \in M$. What we want to express is the desideratum that the inner product $g_p \in \mathsf{Sym}^2(T_pM; \mathbb{R})$ "depends smoothly on p". This is entirely analogous to asking how to make precise the statement that for a smooth function $f \in C^\infty(M)$ the differential $df_p \in T_p^*M = \operatorname{Hom}(T_pM, \mathbb{R})$ "depends smoothly on p".

Lemma 7.18. *Let E be a smooth vector bundle over a smooth manifold M. Then there is a smooth vector bundle $\mathsf{Sym}^2(E; \mathbb{R})$ whose fiber over $p \in M$ is the vector space $\mathsf{Sym}^2(E_p; \mathbb{R})$ of symmetric bilinear forms on the fiber E_p.*

The construction of the vector bundle $\mathsf{Sym}^2(E; \mathbb{R})$ is entirely analogous to the construction of the dual vector bundle E^*: from the local trivializations $E_{|U_\alpha} \cong U_\alpha \times \mathbb{R}^k$ of E we build maps

$$\coprod_{p \in U_\alpha} \mathsf{Sym}^2(E_p; \mathbb{R}) \longrightarrow U_\alpha \times \mathsf{Sym}^2(\mathbb{R}^k; \mathbb{R}) \cong U_\alpha \times \mathbb{R}^\ell, \qquad \ell = \dim \mathsf{Sym}^2(\mathbb{R}^k; \mathbb{R}).$$

which commute with the projection maps to U_α and are fiberwise isomorphisms of vector spaces. Then the Vector Bundle Construction Lemma 7.12 can be used to show that $\mathsf{Sym}^2(E; \mathbb{R}) = \{(p, v) \mid p \in M,\ v \in \mathsf{Sym}^2(E_p; \mathbb{R})\}$ has the structure of a smooth vector bundle.

Definition 7.19. Let M be a smooth manifold. A *Riemannian metric on* M is a smooth section $g\colon M \to \mathsf{Sym}^2(TM; \mathbb{R})$ of the vector bundle $\mathsf{Sym}^2(TM; \mathbb{R})$ such that for each $p \in M$ the symmetric bilinear form $g_p \in \mathsf{Sym}^2(T_pM; \mathbb{R})$ is positive definite (in particular, g_p is an inner product on the tangent space T_pM for every $p \in M$).

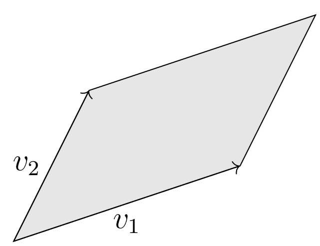
If M is a Riemannian manifold and $\gamma\colon [a, b] \to M$ is a smooth path in M, then the length of γ is defined by the formula (7.16), where the norm $||\gamma'(t)|| \in [0, \infty)$ of the tangent vector $\gamma'(t) \in T_pM$, $p = \gamma(t)$ is evaluated using the inner product g_p on T_pM.

7.4.1 Measuring volumes

Our eventual goal is to integrate over manifolds. When defining the Riemann integral of a function f over an open subset $U \subset \mathbb{R}^n$ we divide U into a bunch of small boxes and approximate the integral over f by the integral over a function which is constant on each small box, thus reducing the calculation of an integral to the calculation of the volume of rectangles. To define integration over manifolds we will use charts to reduce the calculation to open subsets of Euclidean space. However, it turns out to to be useful to not restrict ourselves to *rectangles* in $\mathbb{R}^n$, since the image of a rectangle in $\mathbb{R}^2$ under a linear map $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ is typically no box, but a parallelogram. More generally, the image of the standard n-cube $\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid 0 \leq x_i \leq 1\} \subset \mathbb{R}^n$ under a linear map T with $T(e_i) = v_i \in \mathbb{R}^n$ is the parallelepiped

$$P(v_1, \dots, v_n) := \{\sum_{i=1}^n x_i v_i \mid 0 \leq x_i \leq 1\} \subset \mathbb{R}^n.$$

For $n = 2$, a parallelepiped $P(v_1, v_2)$ is simply the parallelogram spanned by the vectors v_1, v_2 (which is a "degenerate" if v_1, v_2 are linearly dependent). Here is a picture of $P(v_1, v_2)$:



Lemma 7.20. *The volume of the (possibly degenerate) parallelepiped $P(v_1, \dots, v_n)$ spanned by $v_1, \dots, v_n \in \mathbb{R}^n$ is given by the formula*

$$\mathrm{vol}(P(v_1, \dots, v_n)) = |\det(v_1, \dots, v_n)|.$$

Here $\det$ *is interpreted as a map* $\det\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ *that sends an n-tupel $(v_1, \dots, v_n)$ of vectors $v_i \in \mathbb{R}^n$ to the determinant of the $n \times n$ matrix with column vectors v_1, …, v_n.*

We will prove this statement since the techniques going into that proof will be useful for us. Before doing so, we recall properties of the determinant function $\det\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$.

1. The determinant is a *multilinear* map, i.e., it is linear in each slot; explicitly,

   $$\det(v_1, \dots, av_i + bv_i', \dots, v_n) = a\det(v_1, \dots, v_i, \dots, v_n) + b\det(v_1, \dots, v_i', \dots, v_n)$$

   for $v_1, \dots, v_n, v_i' \in \mathbb{R}^n$, $a, b \in \mathbb{R}$.

2. The determinant is *alternating*, i.e., for any permutation $\sigma \in S_k$
$$\det(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \operatorname{sign}(\sigma) \det(v_1, \dots, v_n),$$
where $\operatorname{sign}(\sigma) \in \{\pm 1\}$ is the sign of the permutation σ. We recall that $\operatorname{sign}(\sigma) = 1$ if σ is the composition of an even number of transpositions; otherwise $\operatorname{sign}(\sigma) = -1$.

Definition 7.21. Let V be a vector space. A map
$$\omega \colon \underbrace{V \times \cdots \times V}_{k} \longrightarrow \mathbb{R}$$
is called

1. **multilinear** if ω is linear in each slot;
2. **alternating** if $\omega(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \operatorname{sign}(\sigma)\omega(v_1, \dots, v_k)$ for all $v_1, \dots, v_k$ and $\sigma \in S_k$.

Let $\mathrm{Alt}^k(V; \mathbb{R})$ denote the set of multilinear alternating maps $\omega \colon V \times \cdots \times V \to \mathbb{R}$. This is a vector space, since the sum of two multilinear alternating maps is again a multilinear alternating map; multiplying such a map by a constant $c \in \mathbb{R}$ again such a map.

To calculate the dimension of $\mathrm{Alt}^k(V; \mathbb{R})$, we want to construct a basis for this vector space. Let $\{e_i\}_{i=1,\dots,n}$ be a basis for V, and let $\{e^i\}_{i=1,\dots,n}$ be the dual basis for V^*. Given a multi-index $I = (i_1, \dots, i_k)$ with $i_j \in \{1, \dots, n\}$, it is evident that the map
$$\underbrace{V \times \cdots \times V}_{k} \longrightarrow \mathbb{R} \qquad \text{given by} \qquad (v_1, \dots, v_k) \mapsto e^{i_1}(v_1)e^{i_2}(v_2) \cdots e^{i_k}(v_k)$$
is multilinear. However, in general it is not alternating, since the value of this function on a k-tupel $(v_1, \dots, v_k)$ is unrelated to the value on the permuted k-tupel $(v_{\sigma(1)}, \dots, v_{\sigma(k)})$. For example, if $k = n$, $I = (1, \dots, n)$ and $v_i = e_i$, then
$$e^1(v_1) \dots e^n(v_n) = 1 \qquad \text{but} \qquad e^1(v_{\sigma(1)}) \dots e^n(v_{\sigma(n)}) = 0 \quad \text{for } \sigma \neq \mathrm{id}.$$
However, out of this non-alternating multilinear map we can can manufacture an alternating map e^I by a signed sum over permutations of the v_i:
$$e^I(v_1, \dots, v_k) := \sum_{\sigma \in S_k} \operatorname{sign}(\sigma) e^{i_1}(v_{\sigma(1)}) e^{i_2}(v_{\sigma(2)}) \cdots e^{i_k}(v_{\sigma(k)})$$
for $I = (i_1, \dots, i_k)$, $v_1, \dots, v_k \in V$. It is not hard to check that the multilinear map
$$e^I \colon V \times \cdots \times V \to \mathbb{R}$$
is in fact alternating and so $e^I \in \mathrm{Alt}^k(V; \mathbb{R})$. It is also straightforward to show that if $J = (i_{\sigma(1)}, \dots, i_{\sigma(k)})$ is a permutation of the multi-index $I = (i_1, \dots, i_k)$, then $e^J = \operatorname{sign}(\sigma) e^I$.

Lemma 7.22. *The collection* $\{e^I \mid I = (i_1, \ldots, i_k)$ *with* $i_1 < i_2 < \cdots < i_k\}$ *is a basis for* $\mathrm{Alt}^k(V; \mathbb{R})$.

A proof of this fact can be found in Lee's book. It follows that the dimension of the vector space $\mathrm{Alt}^k(V; \mathbb{R})$ is equal to the number of multi-indices $I = (i_1, \ldots, i_k)$ which are strictly increasing in the sense that $i_1 < \cdots < i_k$. Mapping a strictly increasing multi-index $I = (i_1, \ldots, i_k)$ to the subset $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ yields a bijection between the set of strictly increasing multi-indices and the set of cardinality k subsets of $\{1, \ldots, n\}$. In particular, we conclude:

Corollary 7.23. *If* V *is a vector space of dimension* n*, then* $\dim \mathrm{Alt}^k(V; \mathbb{R}) = \binom{n}{k}$.

Proof of Lemma 7.20. Our strategy to prove $\mathrm{vol}(P(v_1, \ldots, v_n)) = |\det(v_1, \ldots, v_n)|$ is to use the fact that det is an alternating multilinear map, i.e., an element of $\mathrm{Alt}^n(\mathbb{R}^n; \mathbb{R})$, and that the dimension of $\mathrm{Alt}^n(\mathbb{R}^n; \mathbb{R})$ is $\binom{n}{n} = 1$. The idea is that while $\mathrm{vol}(P(v_1, \ldots, v_n))$ is *not* an alternating multilinear map (e.g., its values are non-negative), it is the *absolute value* of an alternating multilinear map

$$\mathrm{svol}\colon \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n} \longrightarrow \mathbb{R}$$

called the *signed volume*, defined by $\mathrm{svol}(v_1, \ldots, v_n) := \epsilon(v_1, \ldots, v_n)\,\mathrm{vol}(P(v_1, \ldots, v_n))$, where

$$\epsilon(v_1, \ldots, v_n) := \begin{cases} +1 & \det(v_1, \ldots, v_n) > 0 \\ -1 & \det(v_1, \ldots, v_n) < 0 \\ 0 & \det(v_1, \ldots, v_n) = 0 \end{cases}$$

It is clear from the definition that $|\mathrm{svol}(v_1, \ldots, v_n)| = \mathrm{vol}(P(v_1, \ldots, v_n))$, and we claim that svol is an element of $\mathrm{Alt}^n(\mathbb{R}^n; \mathbb{R})$. Permuting the vectors $v_1, \ldots, v_n$ does not change the associated parallelepiped, but

$$\det(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = \mathrm{sign}(\sigma) \det(v_1, \ldots, v_n)$$

and hence

$$\epsilon(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = \mathrm{sign}(\sigma)\epsilon(v_1, \ldots, v_n).$$

It follows that svol is alternating. To show that svol is linear in each slot, let us first argue that

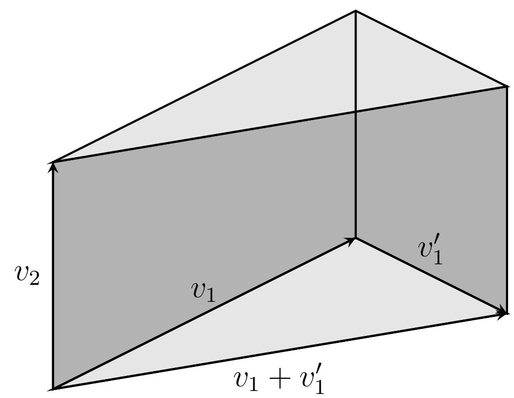
$$\mathrm{svol}(v_1, \ldots, cv_i, \ldots, v_n) = c\, \mathrm{svol}(v_1, \ldots, v_i, \ldots, v_n) \tag{7.24}$$

for $c \in \mathbb{R}$. If c is a positive integer, this is clear geometrically; for $c = -1$, again it is clear geometrically that the volume of the associated parallelepipeds $P(v_1, \ldots, -v_i, \ldots, v_n)$ and $P(v_1, \ldots, v_i, \ldots, v_n)$ is the same, but $\epsilon(v_1, \ldots, -v_i, \ldots, v_n) = -\epsilon(v_1, \ldots, v_i, \ldots, v_n)$. This implies equation (7.24) for $c \in \mathbb{Z}$ and hence for $c \in \mathbb{Q}$. Approximating a real number

$c \in \mathbb{R}$ by rational numbers c_ℓ and taking the limit of the equation $\operatorname{svol}(v_1, \dots, c_\ell v_i, \dots, v_n) = c_\ell \operatorname{svol}(v_1, \dots, v_i, \dots, v_n)$ as for $\ell \to \infty$ yields equation for a general $c \in \mathbb{R}$. The additivity property

$$\operatorname{svol}(v_1, \dots, v_i + v_i', \dots, v_n) = \operatorname{svol}(v_1, \dots, v_i, \dots, v_n) + \operatorname{svol}(v_1, \dots, v_i', \dots, v_n)$$

follows from a geometric argument which is illustrated by the following picture for $n = 2$: the area of the parallelogram $P(v_1 + v_1', v_2$ is equal to the sum of the areas of $P(v_1, v_2)$ and $P(v_1', v_2)$.



This shows that the signed volume svol is an alternating multilinear map, i.e., an element of $\operatorname{Alt}^n(\mathbb{R}^n; \mathbb{R})$. Since this vector space has dimension 1, the element svol must be a scalar multiple of the non-zero element $\det \in \operatorname{Alt}^n(\mathbb{R}^n; \mathbb{R})$, i.e., $\operatorname{svol} = c \det$ for some $c \in \mathbb{R}$. To determine c, we evaluate both sides on the n-tupel $(e_1, \dots, e_n)$, where $\{e_i\}_{i=1,\dots,n}$ is the standard basis of $\mathbb{R}^n$. The determinant $\det(e_1, \dots, e_n)$ is the determinant of the identity matrix and hence 1. The parallelepiped $P(e_1, \dots, e_n)$ is the standard cube which has volume 1 and hence $\operatorname{svol}(e_1, \dots, e_n) = \operatorname{vol}(e_1, \dots, e_n) = 1$. It follows that $c = 1$, and hence for every n-tupel $(v_, \dots, v_n)$ of vectors $v_1, \dots, v_n \in \mathbb{R}^n$ we have

$$\operatorname{svol}(v_1, \dots, v_n) = \det(v_1, \dots, v_n).$$

Taking the absolute value of both sides we conclude the statement of Lemma 7.20. □

I think of an element $\omega \in \operatorname{Alt}^k(V; \mathbb{R})$ as a little machine that takes an input of vectors $v_1, \dots, v_k \in V$ and produces as output the number $\omega(v_1, \dots, v_k) \in \mathbb{R}$; this output depends linearly on each v_i, and permuting the input vectors changes the output by a factor of ± 1, given by the sign of the permutation $\sigma \in S_k$. This suggests that we can multiply a machine ω with k inputs and η with ℓ inputs to obtain a machine typically denoted $\omega \otimes \eta$ with $k + \ell$ inputs by defining:

$$(\omega \otimes \eta)(v_1, \dots, v_{k+\ell}) := \omega(v_1, \dots, v_k)\eta(v_{k+1}, \dots, v_{k+\ell}).$$

It is clear that $\omega \otimes \eta$ is linear in each of its $k+\ell$ input slots, and changes by a factor of $\operatorname{sign}(\sigma) \in \{\pm 1\}$ when using a permutation $\sigma \in S_{k+\ell}$ to permute the $k+\ell$ input vectors, *provided* σ belongs to the subgroup $S_k \times S_\ell \subset S_{k+\ell}$. If we interchange one of the vectors $v_1, \dots, v_k$ with one of the vectors $v_{k+1}, \dots, v_{k+\ell}$, there is no reason that the output just is multiplied by -1. To produce an *alternating* multilinear map, we use the same method we applied before by using a signed sum over all permutations.

Definition 7.25. For $\omega \in \operatorname{Alt}^k(V;\mathbb{R})$ and $\eta \in \operatorname{Alt}^\ell(V;\mathbb{R})$ their *wedge product* is the alternating multilinear form $\omega \wedge \eta \in \operatorname{Alt}^{k+\ell}(V;\mathbb{R})$ defined by

$$(\omega \wedge \eta)(v_1, \dots, v_{k+\ell}) := \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} \operatorname{sign}(\sigma)\omega(v_{\sigma(1)}, \dots, v_{\sigma(k)})\eta(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)}).$$

We note that for $\sigma \in S_k \times S_\ell \subset S_{k+\ell}$ the summand

$$\operatorname{sign}(\sigma)\omega(v_{\sigma(1)}, \dots, v_{\sigma(k)})\eta(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)})$$

is equal to $\omega(v_1, \dots, v_k)\eta(v_{k+1}, \dots, v_{k+\ell})$. In particular, summing over this subgroup of order $k!\ell!$, we simply obtain $k!\ell!\omega(v_1, \dots, v_k)\eta(v_{k+1}, \dots, v_{k+\ell})$. This motivates the factor $\frac{1}{k!\ell!}$ in the definition of the wedge product.

Lemma 7.26. (Properties of the wedge product).

1. **Bilinearity:**
2. **Associativity:** $\omega \wedge (\eta \wedge \xi) = (\omega \wedge \eta) \wedge \xi$
3. **Graded Commutativity:** $\omega \wedge \eta = (-1)^{k\ell} \eta \wedge \omega$ *for* $\omega \in \operatorname{Alt}^k(V;\mathbb{R})$ *and* $\eta \in \operatorname{Alt}^\ell(V;\mathbb{R})$.
4. *For any multi-index* $I = (i_1, \dots, i_k)$, $e^I = e^{i_1} \wedge \cdots \wedge e^{i_k} \in \operatorname{Alt}^k(V;\mathbb{R})$. *In particular, an element* $\omega \in \operatorname{Alt}^k(V;\mathbb{R})$ *can uniquely be written as a sum*

$$\omega = \sum_{i_1 < \cdots < i_k} a_{i_1,\dots,i_k} e^{i_1} \wedge \cdots \wedge e^{i_k} \qquad a_{i_1,\dots,i_k} \in \mathbb{R}.$$

Lemma 7.27. *Let M be a smooth manifold and $E \to M$ a smooth vector bundle over M. Then there is a smooth vector bundle* $\operatorname{Alt}^k(E;\mathbb{R}) \to M$ *whose fiber at a point $p \in M$ is* $\operatorname{Alt}^k(E_p;\mathbb{R})$.

We note that $\operatorname{Alt}^1(E_p;\mathbb{R})$ is the dual space E_p^*, and $\operatorname{Alt}^1(E;\mathbb{R}) \to M$ is the dual vector bundle $E* \to M$. Like the construction of the dual vector bundle, the proof of the above statement uses the Vector Bundle Construction Lemma 7.12.

We recall that a 1-form on a smooth manifold M is a section of the cotangent bundle T^*M. Noting that T^*M is equal to the vector bundle $\operatorname{Alt}^1(TM;\mathbb{R})$, this suggests the following generalization of 1-forms.

Definition 7.28. Let M be a smooth manifold. A *k-form* or *differential form of degree k* on M is a smooth section of the vector bundle $\mathrm{Alt}^k(TM;\mathbb{R})$. The usual notation for the vector space of k-forms on M is
$$\Omega^k(M) := \Gamma(M;\mathrm{Alt}^k(TM;\mathbb{R})).$$

How do we do explicit calculations with differential forms? To do calculations with linear maps between vector spaces, it is often useful to choose a basis for the vector spaces involved. Thinking of a vector bundle as a collection of vector spaces parametrized by points $p \in M$, it is natural to ask how to generalize the notion of "basis" from vector spaces to vector bundles.

Definition 7.29. Let $E \to M$ be a smooth vector bundle of rank k. If $U \subset M$ is an open subset, a *local frame* for E over U is a collection $\{b^i\}_{i=1,\dots,k}$ of smooth sections of $E_{|U}$ such that $\{b^1_p, \dots, b^k_p\}$ is a basis of E_p for each $p \in U$.

Let $\{b^i\}_{i=1,\dots,k}$ be a local frame for E over U and let $s \in \Gamma(U;E)$, i.e., $s\colon U \to E$ is a smooth section of $E_{|U}$. Then for any $p \in U$, the element $s(p) \in E_p$ can be expanded in terms of the basis $\{b^i(p)\}$ to obtain
$$s(p) = \sum_{i=1,\dots,k} s_i(p) b^i(p) \qquad \text{with } s_i(p) \in \mathbb{R}.$$

It is not hard to see that $s_i(p)$ is a *smooth* function of p, since s and b^i are smooth sections of E. Hence we can write the section s as a linear combination
$$s = \sum_{i=1,\dots,k} s_i b^i$$
of the sections b^i whose coefficients s_i are smooth functions $U \to \mathbb{R}$.

Example 7.30. Let M be a smooth manifold of dimension n, and let $M \supset U \xrightarrow{\phi} \mathbb{R}^n$ be a smooth chart (i.e., (U,ϕ) belongs to the maximal smooth atlas defining the smooth structure on M). Let $x^1, \dots, x^n \in C^\infty(U)$ be the component functions of ϕ, i.e., $\phi(p) = (x^1(p), \dots, x^n(p))$. Then the differentials $dx^1_p, \dots, dx^n_p \in T^*_pM$ form a basis

7.5 Algebraic structures on differential forms

The goal of this section is to discuss the various algebraic structures on differential forms and their compatibility.

Definition 7.31. (Pullback). Let M, N be smooth manifolds and $F\colon M \to N$ a smooth map. Given a differential form $\omega \in \Omega^k(N)$, its *pullback* $F^*\omega \in \Omega^k(M)$ is defined by
$$(F^*\omega)_p(v_1, \dots, v_k) := \omega_p(F_*v_1, \dots, F_*v_k) \qquad \text{for } p \in M,\ v_1, \dots, v_k \in T_pM.$$

In more detail: the k-form $F^*\omega$ is a section of the vector bundle $\mathrm{Alt}^k(TM;\mathbb{R})$, and hence it can be evaluated at $p \in M$ to obtain an element $(F^*\omega)_p$ in the fiber of that vector bundle over p, which is $\mathrm{Alt}^k(T_pM;\mathbb{R})$. In other words, $(F^*\omega)_p$ is an alternating multilinear map

$$(F^*\omega)_p\colon \underbrace{T_pM \times \cdots \times T_pM}_{k} \longrightarrow \mathbb{R},$$

and hence it can be evaluated on the k tangent vectors $v_1, \ldots, v_k \in T_pM$ to obtain a real number $(F^*\omega)_p(v_1, \ldots, v_k)$. On the right hand side to the equation defining $F^*\omega$, the map $F_*\colon T_pM \to T_{F(p)}N$ is the differential of F. Hence the alternating multilinear map $\omega_{F(p)} \in \mathrm{Alt}^k(T_{F(p)}N;\mathbb{R})$ can be evaluated on $F_*v_1, \ldots, F_*v_k$ to obtain the real number $\omega_p(F_*v_1, \ldots, F_*v_k)$.

For $k = 1$, $\omega \in \Omega^0(N) = C^\infty(N)$ is a smooth function, and its pullback $F^*\omega$ is the previously defined pullback of functions, simply given by $(F^*\omega)(p) = \omega(F(p))$ for $p \in M$. We also previously defined the pullback of a 1-form $\omega \in \Omega^1(N)$, and we showed that the differentials are compatible with pullbacks in the sense that

$$d(F^*f) = F^*(df) \qquad \text{for } f \in C^\infty(N).$$

Definition 7.32. (Wedge products of differential forms. For $\omega \in \Omega^k(M)$ and $\eta \in \Omega^\ell(M)$, their *wedge product* $\omega \wedge \eta \in \Omega^{k+\ell}(M)$ is defined by

$$(\omega \wedge \eta)_p := \omega_p \wedge \eta_p \in \mathrm{Alt}^{k+\ell}(T_pM;\mathbb{R}) \qquad \text{for } p \in M.$$

Lemma 7.33. *The wedge product of differential forms has the following properties.*

(i) **Bilinearity:**

(ii) **Associativity:**

$$\omega \wedge (\eta \wedge \xi) = (\omega \wedge \eta) \wedge \boldsymbol{x} \qquad \text{for differential forms } \omega,\ \eta,\ \xi \text{ on a smooth manifold } M.$$

(iii) **Graded Commutativity:**

$$\omega \wedge \eta = (-1)^{k\ell}\eta \wedge \omega \qquad \text{for } \omega \in \Omega^k(M) \text{ and } \eta \in \Omega^\ell(M).$$

(iv) **Compatibility with pullbacks:** *If $F\colon M \to N$ is a smooth map and ω, η differential forms on N, then*

$$F^*(\omega \wedge \eta) = (F^*\omega) \wedge (F^*\eta).$$

Proof. The wedge product of differential form is defined *pointwise*, i.e., the wedge product of differential forms $\omega \in \Omega^k(M)$, $\eta \in \Omega^\ell(M)$ is defined by declaring for any point $p \in M$ the element $(\omega \wedge \eta)_p \in \mathrm{Alt}^{k+\ell}(T_pM;\mathbb{R})$ to be the wedge product $\omega_p \wedge \eta_a$ of the alternating multilinear maps $\omega_p \in \mathrm{Alt}^k(T_pM;\mathbb{R})$ and $\eta_p \in \mathrm{Alt}^\ell(T_pM;\mathbb{R})$. It follows that bilinearity, associativity and graded commutativity of the wedge product for alternating multlinear maps stated in Lemma **??** immediately imply these properties for the wedge product of differential forms.

To prove compatibility with pullbacks, let $p \in M$ and $v_1, \dots, v_{k+\ell} \in T_pM$. Then

$$\begin{aligned}
&(F^*(\omega \wedge \eta))_p(v_1, \dots, v_{k+\ell})\\
=&(\omega \wedge \eta)_{F(p)}(F_*(v_1), \dots, F_*(v_{k+\ell}))\\
=&(\omega_{F(p)} \wedge \eta_{F(p)})(F_*(v_1), \dots, F_*(v_{k+\ell}))\\
=&\frac{1}{k!\ell!}\sum_{\sigma \in S_{k+\ell}} (\omega_{F(p)}(F_*(v_{\sigma(1)}), \dots, F_*(v_{\sigma(\ell)}))(\eta_{F(p)}(F_*(v_{\sigma(k+1)}), \dots, F_*(v_{\sigma(k+\ell)}))\\
=&\frac{1}{k!\ell!}\sum_{\sigma \in S_{k+\ell}} (F^*\omega)_p(v_{\sigma(1)}, v_{\sigma(\ell)})(F^*\eta_p(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)})\\
=&((F^*\omega)_p \wedge (F^*\eta)_p)(v_1, \dots, v_{k+\ell})\\
=&(F^*\omega \wedge F^*\eta)_p(v_1, \dots, v_{k+\ell})
\end{aligned}$$

□

Definition 7.34. (Proposition/Definition). Let M be a smooth manifold. Then there is a unique map $d\colon \Omega^k(M) \to \Omega^{k+1}(M)$ called the *de Rham differential* with the following properties:

(i) d is linear;

(ii) for $f \in \Omega^0(M) = C^\infty(M)$, the de Rham differential $df \in \Omega^1(M)$ is the usual differential of f;

(iii) d is a *graded derivation*, i.e., it satisfies the following "product rule with signs":

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta. \tag{7.35}$$

(iv) $d^2 = 0$.

Remark 7.36. The signs appearing in the graded commutativity of the wedge product as well in the product rule for the de Rham differential are examples of the meta principle known as *Koszul sign rule*, according to which a good way to deal with objects with an integer degree (like differential forms) and signs, is to set up definitions such that *permuting* *objects of degree k and ℓ results in a sign of* $(-1)^{k\ell}$. This is satisfied for the wedge product (for alternating multilinear maps or for differential forms). This is also the case for the *graded derivation rule* (ii) above. We recall that a derivation of an algebra A is a linear map $D\colon A \to A$ satisfying the product rule

$$D(a \cdot b) = D(a) \cdot b + a \cdot D(b) \qquad \text{for } a, b \in A.$$

We note that on the left hand side of this equation, as well in the first term on the right hand side the symbols occur in the order D, a, b. By contrast, in the second summand the objects a and D switch occur in the opposite order, which according to the Koszul sign paradigm should involve the sign $(-1)^{\deg(D)\deg(a)}$ in a context where these objects have "degrees" $\deg(D), \deg(a) \in \mathbb{Z}$. For example, in equation (7.35), the differential forms ω, η have degrees $\deg(\omega) = k$, $\deg(\eta) = \ell$, and it is reasonable to declare the de Rham differential d to have degree $+1$, since applying it to a differential form of degree k results in a form of degree $k+1$. This shows that the "graded derivation property" (ii) conforms to the Koszul sign paradigm.

Lemma 7.37. (Additional Properties of the de Rham Differential).

1. **(Compatibility with pullbacks).** *If $F\colon M \to N$ is a smooth map, and $\omega \in \Omega^k(N)$, then $d(F^*\omega) = F^*(d\omega)$.*

2. **(Local Formula for d).** *Let (U, ϕ) be a smooth chart for an n-manifold M, and let $x^1, \dots, x^n \in C^\infty(U)$ be the local coordinate functions (the components of $\phi\colon U \to \mathbb{R}^n$). Then*

$$d(f dx^{i_1} \wedge \cdots \wedge dx^{i_k}) = df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

for $f \in C^\infty(U)$, $i_1, \dots, i_k \in \{1, \dots, n\}$.

We remark that the collection of sections $\{dx^{i_1} \wedge \cdots \wedge dx^{i_k}\}_{i_1 < \cdots < i_k}$ is a local frame for the vector bundle $\mathrm{Alt}^k(TM; \mathbb{R})$ restricted to $U \subset M$. Hence every k-form $\omega \in \Omega^k(U) = \Gamma(U; \mathrm{Alt}^k(TM; \mathbb{R}))$ can be written uniquely as linear combination

$$\omega = \sum_{i_1 < \cdots < i_k} f_{i_1, \dots, i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

for functions $f_{i_1, \dots, i_k} \in C^\infty(U)$. In particular, the local formula for the de Rham differential above allows us to calculate $d\omega$ for any k-form ω on $U \subset M$.

Lemma 7.38. *Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a smooth map and $f\ dx^1 \wedge \cdots \wedge dx^n \in \Omega^n(\mathbb{R}^n)$. Then*

$$F^*(f\ dx^1 \wedge \cdots \wedge dx^n) = F^*(f) \det(dF)\ dx^1 \wedge \cdots \wedge dx^n$$

Here $\det(dF)\colon \mathbb{R}^n \to \mathbb{R}$ *is the smooth function whose value at* $x \in \mathbb{R}^n$ *is the determinant of the differential* $dF_x\colon \mathbb{R}^n \to \mathbb{R}^n$ *of* F *at the point* $x \in \mathbb{R}^n$*; put differently, it is the determinant of the Jacobian matrix of* F *at* $x \in \mathbb{R}^n$*.*

Proof: homework.

7.6 Integration on manifolds

A smooth n-manifold is locally diffeomorphic to open subsets V of $\mathbb{R}^n$, and we know to integrate a smooth function $f \in C^\infty(U)$. More precisely, if we assume that

$$\mathrm{supp}(f) := \text{closure of } \{x \in U \mid f(x) \neq 0\},$$

the *support of* f, is compact, the integral can be defined by

$$\int_U f := \lim \sum_i f(p_i)\,\mathrm{vol}(R_i),$$

where U is divided into suitable small regions R_i (e.g., small n-cubes) whose union covers U, p_i is some point chosen in R_i, and $\mathrm{vol}(R_i)$ denotes the volume of R_i. The limit is taken over increasingly fine subdivisions of U into regions R_i. To make sense of integration over manifolds, we need to understand how integration behaves under diffeomorphisms, namely those coming from transition maps between charts. What we need is the statement of the *Change of Variables Theorem*.

We won't be proving this theorem, but before stating it, we would like to make its statement plausible. We first consider the effect of a linear map $F\colon \mathbb{R}^n \to \mathbb{R}^n$ on the volumes of parallelepipeds. Let $\mathrm{vol}(P(v_1, \dots, v_n))$ be the volume of the parallelepiped spanned by the vectors $v_1, \dots, v_n \in \mathbb{R}^n$. We recall that

$$\mathrm{vol}(P(v_1, \dots, v_n)) = |\det(v_1, \dots, v_n)|.$$

The image of $P(v_1, \dots, v_n)$ under the linear map F is the parallelepiped $P(F(v_1), \dots, F(v_n))$, and hence

$$\mathrm{vol}(F(P(v_1, \dots, v_n))) = |\det(F(v_1), \dots, F(v_n))|.$$

We note that the matrix with column vectors $(F(v_1), \dots, F(v_n))$ is $F \cdots V$, where V is the matrix with column vectors $(v_1, \dots, v_n)$. It follows that

$$\det(F(v_1), \dots, F(v_n)) = \det(F \cdots V) = \det(F)\det(v_1, \dots, v_n),$$

and hence

$$\mathrm{vol}(F(P(v_1, \dots, v_n))) = |\det(F)|\,\mathrm{vol}(P(v_1, \dots, v_n)). \tag{7.39}$$

Now suppose that $\mathbb{R}^n \supset U \xrightarrow{F} V \subset \mathbb{R}^n$ is a diffeomorphism. Let $p \in U$, and let $P(p; v_1, \dots, v_n)$ be the *affine parallelepiped* spanned by the vectors $v_1, \dots, v_n$ viewed as starting at the point p; in other words,

$$P(p; v_1, \dots, v_n) = \{p + \sum_{i=1}^{n} x_i v_i \mid 0 \leq x_i \leq 1\}.$$

Then the image of $P(p; v_1, \dots, v_n)$ under F is no longer a parallelepiped, since F is no longer linear. However, near the point p the map F is well-approximated by the linear map given by its differential $dF_p \colon \mathbb{R}^n \to \mathbb{R}^n$. Hence for small $v_1, \dots, v_n$, the image of $P(p; v_1, \dots, v_n)$ under F is close to the parallelepiped $P(F(p); dF_p(v_1), \dots, dF_p(v_n))$ and hence

$$\begin{aligned} \operatorname{vol}(F(P(p; v_1, \dots, v_n))) &\approx \operatorname{vol}(P(F(p); dF_p(v_1), \dots, dF_p(v_n))) \\ &= \operatorname{vol}(dF_p(v_1), \dots, dF_p(v_n))) \\ &= |\det(dF_p)| \operatorname{vol}(P(v_1, \dots, v_n)) \end{aligned}$$

Let $f \colon V \to \mathbb{R}$ be a smooth function with compact support. Let C_i, $i \in I$ be a collection of small cubes covering $U \subset \mathbb{R}^n$, and let $p_i \in C_i$ be the "lower left corner in C_i". Then the images $F(C_i)$ cover V and hence we can use this decomposition of V to approximate the integral of f over V as follows:

$$\begin{aligned} \int_V f &\approx \sum_{i \in I} f(F(p_i)) \operatorname{vol}(F(C_i)) \\ &\approx \sum_{i \in I} f(F(p_i)) |\det(dF_{p_i})| \operatorname{vol}(C_i) \\ &\approx \int_U f(F(p)) |\det(dF_p)| \end{aligned}$$

Taking the limit as the size of the cubes approaches zero, these approximations are become better and hence we conclude:

Theorem 7.40. (Change of Variables Theorem). *Let $\mathbb{R}^n \supset U \xrightarrow{F} V \subset \mathbb{R}^n$ be a diffeomorphism and let $f \colon V \to \mathbb{R}$ be a function with compact support. Then*

$$\int_V f = \int_U F^*(f) |\det(dF)|.$$

We notice the similarly of the integrand of the integral over U with the pullback of $f\, dx^1 \wedge \dots \wedge dx^n \in \Omega^n(V)$ via a map $F \colon U \to V$ which according to Lemma 7.38 is given by the formula.

$$F^*(f\ dx^1 \wedge \dots \wedge dx^n) = F^*(f) \det(dF)\ dx^1 \wedge \dots \wedge dx^n$$

This suggests:

Definition 7.41. Let $U \subset \mathbb{R}^n$ be an open subset, and let $f\colon U \to \mathbb{R}$ be a smooth function. Let $\mathrm{supp}(f)$, the *support of* f, be compact ($\mathrm{supp}(f)$ is by definition the closure of the subset $\{x \in U \mid f(x) \neq 0\}$). This assumption guarantees that the usual integral $\int_U f dx_1 \dots dx_n$ is finite. Then the *integral of* $f\ dx^1 \wedge \dots \wedge dx^n \in \Omega^n(U)$ *over* U is defined to be

$$\int_U f\ dx^1 \wedge \dots \wedge dx^n := \int_U f dx_1 \dots dx_n \tag{7.42}$$

More generally, any $\omega \in \Omega^n(U)$ can be written uniquely in the form $f\ dx^1 \wedge \dots \wedge dx^n$ for $f \in C^\infty(U)$. The support of f is equal to the support of ω defined to be the closure of the set $\{x \in U \mid \omega_x \neq 0\}$. Hence by the above equation defines the integral

$$\int_U \omega \in \mathbb{R} \qquad \text{for any } \omega \in \Omega^n(U) \text{ with compact support.}$$

It should be emphasized that despite the notional similarities of the integrand on both sides of equation (7.42), the flavor of the objects on each side is quite different: on the right hand side we are integrating over a function f, and the symbols $dx_1 \dots dx_n$ are not strictly speaking necessary – they are just a reminder which variables we are integrating over. By contrast, on the left hand side, the integrand is an n-form on U. Note that permuting say dx^1 and dx^2 on the left hand side, the n-form $\omega = f\ dx^1 \wedge \dots \wedge dx^n$ is replaced by $-\omega$. By contrast, permuting dx_1 and dx_2 on the right hand side doesn't change the value of that integral. It might seem that this is a problem, rendering the above definition of $\int_U \omega$ for $\omega \in \Omega^n(U)$ not well-defined. This in not the case: in order to evaluate $\int_U \omega$, you first have to write ω in the form $f\ dx^1 \wedge \dots \wedge dx^n$ for $f \in C^\infty(U)$, and then apply the definition above.

Corollary 7.43. (Corollary of the Change of Variables Theorem.) *Let $F\colon U \to V$ be a diffeomorphism between connected open subsets of $\mathbb{R}^n$, and let $\omega \in \Omega^n(V)$ be a differential form with compact support. Then*

$$\int_U F^*\omega = \epsilon_F \int_V \omega, \qquad \textit{where} \qquad \epsilon_F = \begin{cases} 1 & \textit{if } \det(dF_x) > 0 \textit{ for all } x \in U \\ -1 & \textit{if } \det(dF_x) < 0 \textit{ for all } x \in U \end{cases}$$

We note that the assumption that F is a diffeomorphism implies that $\det(dF_x) \neq 0$ for all $x \in U$. Since $\det(dF_x)$ depends continuously on $x \in U$, the assumption that U is connected implies that either $\det(dF_x) > 0$ for all $x \in U$, or $\det(dF_x) < 0$ for all $x \in U$.

Proof. Writing ω in the form $\omega = f\ dx^1 \wedge \dots \wedge dx^n$ we calculate:

$$\begin{aligned}
\int_U F^*\omega = \int_U F^*(f\ dx^1 \wedge \dots \wedge dx^n) &= \int_U F^*(f) \det(dF) dx^1 \wedge \dots \wedge dx^n \\
&= \int_U F^*(f) \det(dF) dx_1 \dots dx_n = \epsilon \int_U F^*(f) |\det(dF)| dx_1 \dots dx_n \\
&= \epsilon_F \int_V f dx_1 \dots dx_n = \epsilon_F \int_V f\ dx^1 \wedge \dots \wedge dx^n = \epsilon_F \int_V \omega
\end{aligned}$$

□

Definition 7.44. Let $F: U \to V$ be a diffeomorphism between open subsets of $\mathbb{R}^n$. Then F is called *orientation preserving* if $\det(dF_x) > 0$ for all $x \in U$, and *orientation reversing* if $\det(dF_x) < 0$ for all $x \in U$.

In particular, Corollary 7.43 implies that if $F: U \to V$ is an orientation preserving diffeomorphism between open subsets of $\mathbb{R}^n$ (not necessary connected), then $\int_U F^* = \int_V \omega$ for a compactly supported n-form $\omega \in \Omega^n(V)$. If F is orientation reversing, then $\int_U F^* = -\int_V \omega$. We note that if U is not connected, then the restriction of F to each connected component of U is either orientation preserving or reversing. However, if F is orientation preserving on some component of U, and reversing on another, then $F: U \to V$ is neither orientation preserving nor orientation reversing.

Let M be a smooth n-manifold, (U_α, ϕ_α) a smooth chart, and $\omega \in \Omega^n(M)$ whose support $\operatorname{supp}(\omega)$ is compact and contained in U_α. Then ω can be pulled back via the diffeomorphism

$$\mathbb{R}^n \supset \phi_\alpha(U_\alpha) \xrightarrow{\phi_\alpha^{-1}} U_\alpha \subset M$$

to obtain the n-form $(\phi_\alpha)^*\omega \in \Omega^n(\phi_\alpha(U_\alpha))$. This n-form has compact support and hence can be integrated:

$$\int_{\phi_\alpha(U_\alpha)} (\phi_\alpha^{-1})^*\omega \in \mathbb{R}. \tag{7.45}$$

Does this integral depend on the choice of the chart (U_α, ϕ_α)? Let (U_β, ϕ_β) be another smooth chart of M such that $\operatorname{supp}(\omega) \subset U_\beta$, and let F be the transition map between these charts, given by the composition

$$\phi_\alpha(U_\alpha \cap U_\beta) \xrightarrow{\phi_\alpha^{-1}} U_\alpha \cap U_\beta \xrightarrow{\phi_\beta} \phi_\beta(U_\alpha \cap U_\beta).$$

Then $\phi_\alpha^{-1} = \phi_\beta^{-1} \circ F$, and hence $(\phi_\alpha^{-1})^*\omega = (\phi_\beta \circ F^{-1})^*\omega = F^*(\phi_\beta^{-1})^*\omega$. This implies

$$\begin{aligned}\int_{\phi_\alpha(U_\alpha \cap U_\beta)} (\phi_\alpha^{-1})^*\omega &= \int_{\phi_\alpha(U_\alpha \cap U_\beta)} F^*(\phi_\beta^{-1})^*\omega \\ &= \begin{cases} \int_{\phi_\beta(U_\alpha \cap U_\beta)} (\phi_\beta^{-1})^*\omega & F \text{ is orientation preserving} \\ -\int_{\phi_\beta(U_\alpha \cap U_\beta)} (\phi_\beta^{-1})^*\omega & F \text{ is orientation reversing} \end{cases}\end{aligned} \tag{7.46}$$

These considerations show that the integral (7.45) depends on the choice of the chart used up to sign.

Definition 7.47. (Orientations on smooth manifolds). Let $\mathcal{A}_M$ be a smooth atlas (not necessarily maximal) for a for smooth manifold M of dimension $n > 0$. An *orientation for* $\mathcal{A}_M$ is a map

$$\epsilon: \mathcal{A}_M \longrightarrow \{\pm 1\}$$

such that the transition map $\phi_\beta \circ \phi_\alpha^{-1}$ is orientation preserving if $\epsilon(\phi_\alpha) = \epsilon(\phi_\beta)$, and orientation reversing otherwise. If ϵ is an orientation for $\mathcal{A}_M$, then $-\epsilon$ is also an orientation for $\mathcal{A}_M$, called the *opposite* of ϵ. An *oriented smooth atlas* is an atlas equipped with an orientation. An *oriented smooth n-manifold* is a topological n-manifold M, $n > 0$ equipped with a maximal smooth oriented atlas $\mathcal{A}_M$. Not every smooth manifold M has an oriented smooth atlas; if it does, then M is called *orientable*.

If M is a manifold of dimension 0, i.e., a discrete countable set, then an orientation for M is a map $\epsilon\colon M \to \{\pm 1\}$.

Now we are ready to define the integral $\int_M \omega \in \mathbb{R}$ of a compactly supported n-form $\omega \in \Omega^n(M)$ over a smooth manifold M of dimension $n > 0$ equipped with an orientation given by a maximal oriented smooth atlas $\mathcal{A}_M$. First we assume that $\operatorname{supp}(\omega)$, the support of ω is contained in the domain of a chart $\phi\colon U \xrightarrow{\approx} \phi(U) \subset \mathbb{R}^n$ belonging to $\mathcal{A}_M$. In that case we define

$$\int_M \omega := \epsilon(\phi) \int_{\phi(U)} (\phi^{-1})^* \omega. \tag{7.48}$$

We observe that this is *independent* of the choice of the chart (U, ϕ). To see this, let $(V, \psi) \in \mathcal{A}_M$ be another chart with $\operatorname{supp}(\omega) \subset V$. Let $F := \psi \circ \phi^{-1}$ be the transition map, and let $\epsilon(F) = 1$ (resp. $\epsilon(F) = -1$) if F is orientation preserving (resp. orientation reversing). Then by definition of an oriented smooth atlas, the sign $\epsilon(F)$ of the transition map F is equal to the product $\epsilon(\phi)\epsilon(\psi)$ of the signs of the two charts involved. Hence

$$\epsilon(\phi) \int_{\phi(U\cap V)} (\phi^{-1})^* \omega \underset{(7.46)}{=} \epsilon(F)\epsilon(\phi) \int_{\psi(U\cap V)} (\psi^{-1})^* \omega = \epsilon(\psi) \int_{\psi(U\cap V)} (\psi^{-1})^* \omega.$$

In general, the assumption that $\operatorname{supp}(\omega)$ is compact guarantees that there are smooth charts $(U_1, \phi_1), \dots, (U_k, \phi_k)$ belonging to the maximal oriented smooth atlas $\mathcal{A}_M$ such that $\operatorname{supp}(\omega) \subset U_1 \cup \dots \cup U_k$.

Lemma 7.49. (Smooth partition of unity). *Let K be a compact subset of a smooth manifold M, and let $K \subset U_1 \cup \dots \cup U_k$, where $U_1, \dots, U_k$ are open subsets of M. Then there exist functions $\lambda_i \in C^\infty(M)$ such that*

1. $\operatorname{supp}(\lambda_i) \subset U_i$, *and*

2. $\sum_{i=1}^k \lambda_i(p) = 1$ *for* $p \in K$.

Applying the lemma to $K = \operatorname{supp}(\omega)$, allows us to write $\omega \in \Omega^n(M)$ as a sum

$$\omega = \sum_{i=1}^k \lambda_i \omega.$$

We remark that $\omega_p = \sum_{i=1}^k \lambda_i(p)\omega_p$ for $p \in K$ by construction of the functions λ_i; for $p \notin K$ it holds since $\omega_p = 0$ in that case. This allows us to define

$$\int_M \omega := \sum_{i=1}^k \int_{U_i} \lambda_i \omega, \tag{7.50}$$

where the integrals $\int_{U_i} \lambda_i \omega$ are defined using equation (7.48). It is straightforward to show that the integral is independent of the choices made in definition, i.e., the choice of the charts (U_i, ϕ_i) and of the partition of unity $\{\lambda_i\}$.

If M is a manifold of dimension 0, equipped with an orientation $\epsilon\colon M \to \{\pm 1\}$, and $\omega \in \Omega^0(M) = C^\infty(M)$ is a compactly supported function $\omega\colon M \to \mathbb{R}$, we define

$$\int_M \omega := \sum_{p \in M} \epsilon(p)\omega(p).$$

We note that this is a finite sum, due to the assumption that ω is compactly supported, which means that $\omega(p) = 0$ for all but finitely many $p \in M$.

Our next goal is to extend the integration of forms from manifolds to manifolds with boundary.

Definition 7.51. A *topological n-manifold with boundary* is a second countable Hausdorff space M with a subspace $\partial M \subset M$, such that the pair $(M, \partial M)$ is locally homeomorphic to the pair $(\mathbb{H}^n, \partial \mathbb{H}^n)$, where $\mathbb{H}^n \subset \mathbb{R}^n$ is the halfspace

$$\mathbb{H}^n := \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_1 \leq 0\},$$

and $\partial \mathbb{H}^n = \{(0, x_2, \dots, x_n)\} \subset \mathbb{H}^n$. In more detail, this means that there is a collection of open subsets $U_\alpha \subset M$, $\alpha \in A$, covering M, and homeomorphisms

$$M \underset{\text{open}}{\supset} U_\alpha \xrightarrow[\approx]{\phi_\alpha} V_\alpha \underset{\text{open}}{\subset} \mathbb{H}^n; \qquad \text{such that} \qquad \phi_\alpha(U_\alpha \cap \partial M) = V_\alpha \cap (\partial \mathbb{H}^n). \tag{7.52}$$

Such a pair (U_α, ϕ_α) is called a *chart* for $(M, \partial M)$, and the collection the charts $\{(U_\alpha, \phi_\alpha)\}$ is called an *atlas* for $(M, \partial M)$.

The requirement $\phi_\alpha(U_\alpha \cap \partial M) = V_\alpha \cap (\partial \mathbb{H}^n)$ for a chart (U_α, ϕ_α) guarantees that ϕ_α^∂, the restriction of ϕ_α gives a homeomorphism

$$\partial M \underset{\text{open}}{\supset} U_\alpha^\partial \xrightarrow[\approx]{\phi_\alpha^\partial} V_\alpha^\partial \underset{\text{open}}{\subset} \partial \mathbb{H}^n = \mathbb{R}^{n-1}.$$

In particular, ∂M is a topological manifold of dimension $n-1$ with atlas $\{(U_\alpha^\partial, \phi_\alpha^\partial)\}_{\alpha \in A}$.

Definition 7.53. (Smooth manifold with boundary.) Let $(M, \partial M)$ be a topological manifold with boundary. An atlas $\{(U_\alpha, \phi_\alpha)_{\alpha \in A}$ for $(M, \partial M)$ (in the sense of Definition 7.51) is *smooth* if for any $\alpha, \beta \in A$ the transition map

$$\mathbb{H}^n \underset{\text{open}}{\supset} \phi_\alpha(U_\alpha \cap U_\beta) \xrightarrow[\approx]{\phi_\alpha^{-1}} U_\alpha \cap U_\beta \xrightarrow[\approx]{\phi_\beta} \phi_\beta(U_\alpha \cap U_\beta) \underset{\text{open}}{\subset} \mathbb{H}^n$$

is smooth. This means that this map is the restriction of a smooth map defined on an open subset of $\mathbb{R}^n$ (unlike the transition maps for charts for manifolds without boundary, the domain $\phi_\alpha(U_\alpha \cap U_\beta)$ of the transition map above might *not* be an open subset of $\mathbb{R}^n$; it is an open subset of $\mathbb{H}^n$).

A *smooth n-manifold with boundary* is a topological manifold with boundary ∂M equipped with a maximal smooth atlas.

Definition 7.54. (The boundary of a smooth manifold). Let M be a smooth n-manifold with boundary ∂M, and with the smooth structure given by a smooth atlas $\mathcal{A}_M = \{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$. Then the homeomorphism $\phi_\alpha \colon U_\alpha \xrightarrow{\approx} V_\alpha$ restricts to a homeomorphism

$$\partial M \underset{\text{open}}{\supset} U_\alpha^\partial \xrightarrow[\approx]{\phi_\alpha^\partial} V_\alpha^\partial \underset{\text{open}}{\subset} \partial \mathbb{H}^n = \mathbb{R}^{n-1}.$$

where $U_\alpha^\partial := U_\alpha \cap \partial M$ and $V_\alpha^\partial := V_\alpha \cap \partial \mathbb{H}^n$. It is clear that $\{U_\alpha^\partial\}_{\alpha \in A}$ is an open cover of ∂M, and hence $\mathcal{A}_{\partial M} := \{(U_\alpha^\partial, \phi_\alpha^p)\}_{\alpha \in A}$ is an atlas for ∂M, showing that ∂M is a topological manifold (∂M is Hausdorff and second countable as a subspace of M which is Hausdorff and second countable). This atlas is in fact smooth, since the transition map

$$\phi_\beta^\partial \circ (\phi_\alpha^\partial)^{-1} \colon \phi_\alpha^\partial(U_\alpha^\partial \cap U_\beta^\partial) \longrightarrow \phi_\beta^\partial(U_\alpha^\partial \cap U_\beta^\partial)$$

is the restriction of the smooth transition map $\phi_\beta \circ \phi_\alpha^{-1}$ to $\phi_\alpha(U_\alpha \cap U_\beta) \cap \partial \mathbb{H}^n = \phi_\alpha^\partial(U_\alpha^\partial \cap U_\beta^\partial)$ and hence smooth. This equips the topological space ∂M with the structure of a smooth manifold.

An orientation on a smooth n-manifold M with boundary ∂M is defined exactly as in Definition 7.47, namely as a maximal oriented smooth atlas for $(M, \partial M)$. The only difference is that now the atlas consists of charts (U_α, ϕ_α) appropriate for manifolds with boundary as spelled out in (7.52).

Let M be an oriented smooth manifold of dimension $n > 1$ with boundary ∂M and let $\epsilon \colon \mathcal{A}_M \to \{\pm 1\}$ be the maximal oriented smooth atlas for $(M, \partial M)$. Let $\mathcal{A}_{\partial M}$ be the associated smooth atlas for ∂M described in Definition 7.54. We recall that for every chart $(U_\alpha, \phi_\alpha) \in \mathcal{A}_M$, there is a chart $(U_\alpha^\partial, \phi_\alpha^\partial)$, where $U_\alpha^\partial = U_\alpha \cap \partial M$, and ϕ_α^∂ is the restriction of ϕ_α to $U_\alpha \cap \partial M$.

Lemma 7.55. *Let U, V be open subsets of $\mathbb{H}^n$, and let $F\colon U \to V$ be a diffeomorphism which restricts to a diffeomorphism $F^\partial\colon U \cap \partial\mathbb{H}^n \to \partial\mathbb{H}^n$. Then* $\det(dF_p)$ *and* $\det(dF^\partial_p)$ *have the same sign for any point $p \in U \cap \partial\mathbb{H}$.*

Proof. Let $p \in U \cap \partial\mathbb{H}^n$. For $i = 2, \dots, n$, the tangent vector $\frac{\partial}{\partial x^i} \in T_p\mathbb{H}^n$ belongs to the subspace $T_p\partial\mathbb{H}^n$ and hence $F_*(\frac{\partial}{\partial x^i}) = F^\partial_*(\frac{\partial}{\partial x^i}) \in T_p\partial\mathbb{H}^n$. The tangent vector $\frac{\partial}{\partial x^1} \in T_p\mathbb{H}^n$ is represented by a path $\gamma\colon (-\epsilon, \epsilon) \to \mathbb{R}^n \supset \mathbb{H}^n$ with $\gamma(t) \in \mathbb{H}^n$ for $t \leq 0$ and $\gamma(t) \notin \mathbb{H}^n$ for $t > 0$; hence the same is true for the image $F(\gamma(t))$ (this makes sense for $0 < t < \epsilon$ for a sufficiently small $\epsilon > 0$, since F is the restriction of a map defined on an open neighborhood of p). This implies that the first component of the tangent vector $F_*(\frac{\partial}{\partial x^1})$ is positive.

The vector $F_*(\frac{\partial}{\partial x^i})$ is the i^{th} column vector of the Jacobian matrix dF_p at the point p. Hence the Jacobian matrix dF_p has the following block matrix form:

$$dF_p = \begin{pmatrix} > 0 & 0 \\ * & dF^\partial_p \end{pmatrix},$$

where dF^∂_p is the Jacobian $(n-1) \times (n-1)$ matrix of F^∂ at the point p. In particular, $\det(F_p)$ is a positive multiple of $\det(F^\partial_p)$ which proves the lemma. $\square$

Definition 7.56. (Induced orientation on the boundary.) Let M be an oriented smooth manifold of dimension $n > 1$ with boundary ∂M and let $\epsilon\colon \mathcal{A}_M \to \{\pm 1\}$ be the maximal oriented smooth atlas for $(M, \partial M)$. Let $\mathcal{A}_{\partial M} = \{(U^\partial_\alpha, \phi^\partial_\alpha)\}_{\alpha \in A}$ be the associated smooth atlas for ∂M. Let $\epsilon^\partial\colon \mathcal{A}_{\partial M} \to \{\pm 1\}$ be the map defined by $\epsilon^\partial(\phi^\partial_\alpha) := \epsilon(\phi_\alpha)$. The Lemma above shows that this is indeed an orientation for the smooth atlas $\mathcal{A}_{\partial M}$.

Theorem 7.57. (Stokes' Theorem). *Let M be a smooth oriented manifold of dimension n with boundary ∂M and let $\omega \in \Omega^{n-1}(M)$ be a differential form with compact support. Then*

$$\int_M d\omega = \int_{\partial M} \omega_{|\partial M}.$$

Proof. Let $\epsilon\colon \mathcal{A}_M \to \{\pm 1\}$ be the maximal oriented smooth atlas for $(M, \partial M)$. Using partitions of unity it suffices to prove Stokes' Theorem in the case where the support of ω is contained in the domain $U \subset M$ of a chart (U, ϕ) belonging to the atlas $\mathcal{A}_M$. Using the chart (U, ϕ) to compute the integral, we obtain

$$\int_M d\omega = \epsilon(\phi) \int_V (\phi^{-1})^* d\omega = \epsilon(\phi) \int_V d(\phi^{-1})^*\omega = \epsilon(\phi) \int_V d\eta,$$

where $\eta := (\phi^{-1})^*\omega$. Similarly, to calculate the integral $\int_{\partial M} \omega$ we use the chart $(U^\partial, \phi^\partial)$ to obtain

$$\int_{\partial M} \omega = \epsilon(\phi^\partial) \int_{V^\partial} ((\phi^\partial)^{-1})^* \omega_{|\partial M}.$$

This can be simplified observing that $((\phi^\partial)^{-1})^*\omega_{|\partial M} = ((\phi^{-1})^*\omega)_{V^\partial} = \eta_{|V^\partial}$ and hence

$$\int_{\partial M} \omega = \epsilon(\phi^\partial) \int_{V^\partial} \eta_{|V^\partial}.$$

Using that $\epsilon(\phi^\partial) = \epsilon(\phi)$ by the definition of the induced orientation on the boundary ∂M of the oriented manifold M, we conclude that it suffices to show the equality

$$\int_V d\eta = \int_{V^\partial} \eta_{|V^\partial}. \tag{7.58}$$

The assumption that ω has compact support implies that η has compact support, and hence by Heine-Borel the support $\mathrm{supp}(\eta)$ is bounded, and consequently contained in the box shaped subset

$$B := [-R, 0] \times \underbrace{[-R, R] \times \cdots \times [-R, R]}_{n-1} \subset \mathbb{H}^n$$

for sufficiently large R. From now on, we will be thinking of η as an $(n-1)$ form defined in this box (extending it by zero outside its original domain $V \subset \mathbb{H}^n$). Like any $(n-1)$-form on $V \subset \mathbb{H}^n \subset \mathbb{R}^n$ the form η can be written as a linear combination

$$\eta = \sum_{i=1}^n f_i dx^1 \wedge \ldots \widehat{dx^i} \cdots \wedge dx^n,$$

where the coefficients f_i are smooth functions on V, and as usual the hat over dx^i indicates that this term should be skipped. Restricting η to $V^\partial \subset \partial\mathbb{H}^n$ yields zero for all but the first summand of η, since the restriction of dx^1 to $\partial\mathbb{H}^n$ is trivial (since the restriction of the coordinate function x^1 to $\partial\mathbb{H}^n$ is trivial). Hence $\eta_{|V^\partial} = (f_1 dx_2 \wedge \cdots \wedge dx^n)_{|V^\partial}$. Then

$$\int_{V^\partial} \eta_{|V^\partial} = \int_{|V^\partial} (f_1 dx_2 \wedge \cdots \wedge dx^n)_{|V^\partial} = \int_{-R}^R dx^2 \cdots \int_{-R}^R dx^n f_1(0, x_2, \ldots, x_n). \tag{7.59}$$

To calculate the left hand side of (7.58), we first compute $d\eta$ using the graded derivation property of the de Rham differential:

$$\begin{aligned} d\eta &= \sum_{i=1}^n df_i \wedge dx^1 \wedge \ldots \widehat{dx^i} \cdots \wedge dx^n \\ &= \sum_{i=1}^n \left(\sum_{j=1}^n \frac{\partial f_i}{\partial x^j} dx^j \right) \wedge dx^1 \wedge \ldots \widehat{dx^i} \cdots \wedge dx^n \\ &= \sum_{i=1}^n \frac{\partial f_i}{\partial x^i} dx^i \wedge dx^1 \wedge \ldots \widehat{dx^i} \cdots \wedge dx^n \\ &= \sum_{i=1}^n (-1)^{i-1} \frac{\partial f_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n \end{aligned} \tag{7.60}$$

The third equality holds since the terms with $i \neq j$ vanish due to the multiple occurrence of the factor dx^j in the wedge product. The last equality holds due to the graded commutativity of the wedge product: the factor dx^i is moved past the $i-1$ factors $dx^1, \ldots, dx^{i-1}$; moving the 1-form dx^i past each one of these factors results in a minus sign.

Integrating the summand $\frac{\partial f_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n$ of the decomposition (7.60) of $d\eta$ we obtain

$$\begin{aligned}
\int_V \frac{\partial f_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n &= \int_B \frac{\partial f_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n \\
&= \int_B \frac{\partial f_i}{\partial x^i} \\
&= \int_{-R}^{0} dx_1 \int_{-R}^{R} dx_2 \cdots \int_{-R}^{R} dx_n \frac{\partial f_i}{\partial x^i}.
\end{aligned}$$

In the last line the symbols dx_i are not 1-forms, but rather the usual calculus notation used to indicate the integration intervals of the variables involved (e.g., $\int_{-R}^{R} dx^2$ indicates to integrate the variable x_2 from $-R$ to R). The order in which we integrate over the variables $x_1, \ldots, x_n$ doesn't matter, so let us evaluate first the integral over x_i, $i \neq 1$. For $x = (x_1, \ldots, x_n) \in B$ we obtain

$$\begin{aligned}
\int_{-R}^{R} dx_i \frac{\partial f_i}{\partial x^i} =& f_i(x)|_{x_i=-R}^{x_i=R} \\
=& f_i(x_1, \ldots, x_{i-1}, R, x_{i+1}, \ldots, x_n) - f_i(x_1, \ldots, x_{i-1}, -R, x_{i+1}, \ldots, x_n)
\end{aligned}$$

We note that the points $(x_1, \ldots, x_{i-1}, \pm R, x_{i+1}, \ldots, x_n)$ both belong to the boundary of the box B. Since the support of η is in the interior of the box, the support of the coefficient functions f_i of η is also in the interior of the box. Hence the above integral vanishes for all $i = 2, \ldots, n$.

For $i = 1$, we have

$$\int_{-R}^{0} dx_1 \frac{\partial f_1}{\partial x^1} = f_1(x)|_{x_1=-R}^{x_1=0} = f_1(0, x_2, \ldots, x_n) - f_1(-R, x_2, \ldots, x_n) = f_1(0, x_2, \ldots, x_n)$$

Putting these calculations together, we obtain

$$\int_V d\eta = \int_{-R}^{R} dx^2 \cdots \int_{-R}^{R} dx^n f_1(0, x_2, \ldots, x_n)$$

which agrees the expression (7.59) for the right hand side of (7.58). This completes the proof of Stokes' Theorem. □