Manifolds

Bruno Martelli

Dipartimento di Matematica, Largo Pontecorvo 5, 56127 Pisa, Italy
E-mail address: martelli at dm dot unipi dot it

version: June 29, 2023

Contents

Introduction

Part 1

Preliminaries

CHAPTER 1

Preliminaries

We state here some basic notions of topology and analysis that we will use in this book. The proofs of some theorems are omitted and can be found in many excellent sources.

1.1. General topology

1.1.1. Topological spaces. A *topological space* is a pair (X, τ) where X is a set and τ is a collection of subsets of X called *open subsets*, satisfying the following axioms:

- $\varnothing$ and X are open subsets;
- the arbitrary union of open subsets is an open subset;
- the finite intersection of open subsets is an open subset.

If this holds we say that τ is a *topology* for the set X. The complement $X \setminus U$ of an open subset $U \in \tau$ is by definition a *closed* subset. Of course the open subsets determine the closed subsets and viceversa.

Every set X has many different topologies. At the two extremes we have the following:

- the *trivial* topology $\tau = \{X, \varnothing\}$, and
- the *discrete* topology where τ consists of all subsets of X.

Informally, in the trivial topology all points are undistinguishable, while in the discrete topology all the points are neatly separated from each other. The topologies that are of interest for us in are of neither of these extremal types and lie somehow in the middle.

When we denote a topological space, we often write X instead of (X, τ) for simplicity.

1.1.2. Continuous maps. A map $f\colon X \to Y$ between topological spaces is *continuous* if the inverse image of every open subset of Y is an open subset of X. The map f is a *homeomorphism* if it has an inverse $f^{-1}\colon Y \to X$ which is also continuous.

Two topological spaces X and Y are *homeomorphic* if there is a homeomorphism $f\colon X \to Y$ relating them. Being homeomorphic is clearly an equivalence relation. Informally, two homeomorphic spaces have the same kind of topological structure and should share the same topological properties.

A *neighbourhood* of a point $x \in X$ is any subset $N \subset X$ containing an open set U that contains x, that is $x \in U \subset N \subset X$. Here is an equivalent notion of continuity that is closer to the one introduced in analysis.

Exercise 1.1.1. A function $f : X \to Y$ is continuous if and only if for every $x \in X$ the inverse image $f^{-1}(N)$ of any neighbourhood N of $f(x)$ is a neighbourhood of x.

1.1.3. Examples. There are many ways to construct topological spaces and we summarise them here very briefly.

Metric spaces. Every metric space (X, d) is also naturally a topological space: by definition, a subset $U \subset X$ is open $\iff$ for every $x_0 \in U$ there is an $r > 0$ such that the open ball

$$B(x_0, r) = \big\{ x \in X \bigm| d(x, x_0) < r \big\}$$

is entirely contained in U.

In particular $\mathbb{R}^n$ is a topological space, whose topology is induced by the euclidean distance between points:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x^i - y^i)^2}.$$

Exercise 1.1.2. Every open ball $B(x_0, r) \subset \mathbb{R}^n$ is homeomorphic to $\mathbb{R}^n$ itself. More generally, the open convex subsets of $\mathbb{R}^n$ are all homeomorphic.

Product topology. The cartesian product $X = \prod_{i \in I} X_i$ of two or more topological spaces is a topological space: by definition, a subset $U \subset X$ is open $\iff$ it is a (possibly infinite) union of products $\prod_{i \in I} U_i$ of open subsets $U_i \subset X_i$, where $U_i \neq X_i$ only for finitely many i.

Exercise 1.1.3. This is the coarsest topology (that is, the topology with the fewest open sets) on X such that the projections $X \to X_i$ are all continuous.

Subspace topology. Every subset $S \subset X$ of a topological space X is also naturally a topological space: by definition a subset $U \subset S$ is open $\iff$ there is an open subset $V \subset X$ such that $U = V \cap S$.

Exercise 1.1.4. This is the coarsest topology on S such that the inclusion $i : S \hookrightarrow X$ is continuous.

In particular every subset $S \subset \mathbb{R}^n$ is naturally a topological space. It is quite remarkable that a topological structure on a set X induces one on any subset $S \subset X$, with no requirement on S.

Quotient topology. Let $f\colon X \to Y$ be a surjective map. A topology on X induces one on Y as follows: by definition a set $U \subset Y$ is open $\iff$ its counterimage $f^{-1}(U)$ is open in X.

Exercise 1.1.5. This is the finest topology (that is, the one with the most open subsets) on Y such that the map $f\colon X \to Y$ is continuous.

A typical situation is when Y is the quotient space $Y = X/_\sim$ for some equivalence relation $\sim$ on X, and $X \to Y$ is the induced projection.

1.1.4. Connected spaces. A topological space X is *connected* if it is not the disjoint union $X = X_1 \sqcup X_2$ of two non-empty open subsets X_1, X_2.

Exercise 1.1.6. The space $\mathbb{R}$ is connected. A product of connected spaces is connected. Hence $\mathbb{R}^n$ is also connected.

Exercise 1.1.7. Every topological space X is partitioned canonically into maximal connected subsets, called *connected components*.

Given the canonical decomposition into connected components, it is typically harmless to restrict our attention to connected spaces.

Exercise 1.1.8. Let $f\colon X \to Y$ be a continuous map between topological spaces. If X is connected, then also $f(X)$ is.

A slightly stronger notion is that of path-connectedness. A space X is *path-connected* if for every $x, y \in X$ there is a *path* connecting them, that is a continuous map $\alpha\colon [0,1] \to X$ with $\alpha(0) = x$ and $\alpha(1) = y$. Every path-connected space is connected, and the converse is also true if the space fulfills some reasonable requirement. A space is *locally path-connected* if every point has a path-connected neighbourhood. Every topological space we will encounter in this book will be locally path-connected.

Exercise 1.1.9. A locally path-connected topological space X is connected $\iff$ it is path-connected.

Exercise 1.1.10. The Euclidean space $\mathbb{R}^n$ is path-connected. Products and quotients of path-connected spaces are path-connected.

1.1.5. Compact spaces. Let X be a topological space. An *open cover* for X is a collection $\{U_i\}_{i\in I}$ of open sets whose union is X. A *subcover* is any subcollection of U_i's that still form a cover. An open cover is finite if it consists of finitely many open sets.

Definition 1.1.11. A topological space X is *compact* if every open cover for X contains a finite subcover.

We are not merely requiring that X has a finite open cover, since every X has one, with X itself as a unique open set. The definition is more subtle and says that every open cover, no matter how complicated, should contain a finite one.

Exercise 1.1.12. The closed segment $[0, 1]$ is compact.

On metric spaces the notion of compactness may be expressed in a different, and maybe more familiar, analytic way.

Exercise 1.1.13. A metric space X is compact $\iff$ the following holds: every sequence of points in X contains a converging subsequence.

On $\mathbb{R}^n$ there is a still more familiar formulation.

Exercise 1.1.14. A subspace of $\mathbb{R}^n$ is compact $\iff$ it is closed and bounded.

We already know that continuous maps send connected spaces to connected spaces, and they do the same with compact spaces.

Exercise 1.1.15. Let $f : X \to Y$ be a continuous map between topological spaces. If X is compact, then $f(X)$ also is.

Finally, compactness is preserved under some operations.

Exercise 1.1.16. Products and quotients of compact spaces are compact. A closed subspace in a compact space is also compact.

1.1.6. Reasonable assumptions. A topological space can be very wild, but most of the spaces encountered in this book will satisfy some reasonable assumptions, that we now list.

Hausdorff. A topological space X is *Hausdorff* if every two distinct points $x, y \in X$ have disjoint open neighbourhoods U_x and U_y, that is $U_x \cap U_y = \varnothing$.

Proposition 1.1.17. *Every metric space has a Hausdorff topology.*

Proof. Two distinct points $x, y \in X$ are at some strictly positive distance $d = d(x, y) > 0$. The balls $B(x, d/2)$ and $B(y, d/2)$ are disjoint thanks to the triangular inequality. □

In particular the euclidean space $\mathbb{R}^n$ is Hausdorff.

Exercise 1.1.18. Products and subspaces of Hausdorff spaces are also Hausdorff. The quotient of a Hausdorff space needs not to be Hausdorff!

Countable base. A *base* for a topological space X is a set of open subsets $\{U_i\}$ such that every open set in X is a union of these. Here are some examples:

- On a metric space X, pick all the balls $B(x, r)$ with varying $x \in X$ and $r > 0$. These form a base.
- If $\mathcal{B}$ and $\mathcal{B}'$ are bases for X and X' respectively, the products $U \times U'$ as $U \in \mathcal{B}$ and $U' \in \mathcal{B}'$ vary form a base for $X \times X'$.

At some point we will only consider spaces that have a countable basis. This amounts informally to requiring that X be not too large. For instance, the euclidean space $\mathbb{R}^n$ has a countable base: we can take all the open balls $B(x, r)$ where x has rational coordinates and $r > 0$ is a rational number.

Exercise 1.1.19. Countable products and subspaces of spaces with a countable basis also have a countable basis.

Locally compact. A topological space X is *locally compact* if every point $x \in X$ has a compact neighbourhood. The euclidean space $\mathbb{R}^n$ is locally compact.

1.1.7. Reasonable consequences. The reasonable assumptions listed in the previous section have some nice and reasonable consequences.

Countable base with compact closure. We first note the following.

Proposition 1.1.20. *If a topological space X is Hausdorff and locally compact, every $x \in X$ has an open neighbourhood $U(x)$ with compact closure.*

Proof. Every $x \in X$ has a compact neighbourhood $V(x)$, that is closed since X is Hausdorff. The neighbourhood $V(x)$ contains an open neighbourhood $U(x)$ of x, whose closure is contained in $V(x)$ and hence compact. □

Proposition 1.1.21. *Every locally compact Hausdorff space X with a countable base has a countable base made of open sets with compact closure.*

Proof. Let $\{U_i\}$ be a countable base. For every open set $U \subset X$ and $x \in U$, there is an open neighbourhood $U(x) \subset U$ of x with compact closure, which contains a U_i that contains x. Therefore the U_i with compact closure suffice as a base for X. □

Exhaustion by compact sets. Let X be a topological space. An *exhaustion by compact subsets* is a countable family $K_1, K_2, \ldots$ of compact subsets such that $K_i \subset \text{int}(K_{i+1})$ for all i and $\cup_i K_i = X$.

The standard example is the exhaustion of $\mathbb{R}^n$ by closed balls

$$K_i = \overline{B(0,i)} = \left\{x \in \mathbb{R}^n \;\middle|\; \|x\| \leq i\right\}.$$

Proposition 1.1.22. *Every locally compact Hausdorff space X with a countable base has an exhaustion by compact subsets.*

Proof. The space X has a countable base $U_1, U_2, \ldots$ of open sets with compact closures. Define $K_1 = \overline{U_1}$ and

$$K_{i+1} = \overline{U_1} \cup \ldots \cup \overline{U_k}$$

where k is the smallest natural number such that $K_i \subset \cup_{j=1}^k U_j$. □

Paracompactness. An *open cover* for a topological space X is a set $\{U_i\}$ of open sets whose union is the whole of X. An open cover $\{U_i\}$ is *locally finite* if every point in X has a neighbourhood that intersects only finitely many U_i. A *refinement* of an open cover $\{U_i\}$ is another open cover $\{V_j\}$ such that every V_j is contained in some U_i.

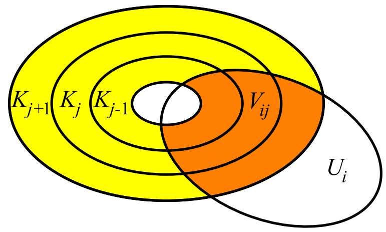

Figure 1.1. A locally compact Hausdorff space with countable base is paracompact: how to construct a locally finite refinement using an exhaustion by compact subsets.

Definition 1.1.23. A topological space X is *paracompact* if every open cover $\{U_i\}$ has a locally finite refinement $\{V_j\}$.

Of course a compact space is paracompact, but the class of paracompact spaces is much larger.

Proposition 1.1.24. *Every locally compact Hausdorff space X with countable base is paracompact.*

Proof. Let $\{U_i\}$ be an open covering: we now prove that there is a locally finite refinement. We know that X has an exhaustion by compact subsets $\{K_j\}$, and we set $K_0 = K_{-1} = \varnothing$. For every i, j we define $V_{ij} = \big(\mathrm{int}(K_{j+1}) \setminus K_{j-2}\big) \cap U_i$ as in Figure 1.1. The family $\{V_{ij}\}$ is an open cover and a refinement of $\{U_i\}$, but it may not be locally finite.

For every fixed $j = 1, 2, \ldots$ only finitely many V_{ij} suffice to cover the compact set $K_j \setminus \mathrm{int}(K_{j-1})$, so we remove all the others. The resulting refinement $\{V_{ij}\}$ is now locally finite. $\square$

In particular the Euclidean space $\mathbb{R}^n$ is paracompact, and more generally every subspace $X \subset \mathbb{R}^n$ is paracompact. The reason for being interested in paracompactness may probably sound obscure at this point, and it will be unveiled in the next chapters.

1.1.8. Topological manifolds.

Recall that the *open unit ball* in $\mathbb{R}^n$ is

$$B^n = \big\{ x \in \mathbb{R}^n \ \big|\ \|x\| < 1 \big\}.$$

A *topological manifold* of dimension n is a reasonable topological space locally modelled on B^n.

Definition 1.1.25. A *topological manifold* of dimension n (shortly, a *topological n-manifold*) is a Hausdorff topological space M with countable base such that every point x has an open neighbourhood U_x homeomorphic to B^n.

In other words, a Hausdorff topological space M with countable base is a manifold $\Longleftrightarrow$ it has an open covering $\{U_i\}$ such that each U_i is homeomorphic

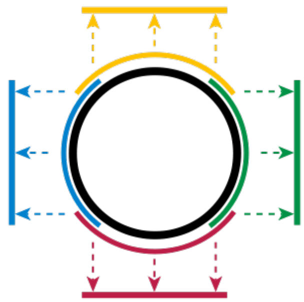

Figure 1.2. A topological manifold is covered by open subsets, each homeomorphic to B^n. Here the manifold is a circle, and is covered by four open arcs, each homeomorphic to the open interval B^1.

to B^n. A schematic picture in Figure 1.2 shows that the circle is a topological 1-manifold: a more rigorous proof will be given in the next chapters.

Example 1.1.26. Every open subset of $\mathbb{R}^n$ is a topological n-manifold. In general, any open subset of a topological n-manifold is a topological n-manifold.

1.1.9. Pathologies. The two reasonability hypothesis in Definition 1.1.25 are there only to discard some spaces that are usually considered as pathological. Here are two examples. The impressionable reader may skip this section.

Exercise 1.1.27 (The double point). Consider two parallel lines $Y = \{y = \pm 1\} \subset \mathbb{R}^2$ and their quotient $X = Y/_\sim$ where $(x, y) \sim (x', y') \Longleftrightarrow x = x'$ and ($y = y'$ or $x \neq 0$). Prove that every point in X has an open neighbourhood homeomorphic to B^1, but X is not Hausdorff.

The following is particularly crazy.

Exercise 1.1.28 (The long ray). Let α be an ordinal, and consider $X = \alpha \times [0, 1)$ with the lexicographic order. Remove from X the first element $(0, 0)$, and give X the *order topology*, having the intervals $(a, b) = \{a < x < b\}$ as a base. If α is countable, then X is homeomorphic to $\mathbb{R}$. If $\alpha = \omega_1$ is the first non countable ordinal, then X is the *long ray*: every point in X has an open neighbourhood homeomorphic to B^1, but X is not separable (it contains no countable dense subset) and hence does not have a countable base. However, the long ray X is path-connected!

1.1.10. Homotopy. Let X and Y be two topological spaces. A *homotopy* between two continuous maps $f, g\colon X \to Y$ is another continuous map $F\colon X \times [0, 1] \to Y$ such that $F(\cdot, 0) = f$ and $F(\cdot, 1) = g$. Two maps f and g are *homotopic* if there is a homotopy between them, and we may write $f \sim g$.

Two topological spaces X and Y are *homotopically equivalent* if there are two continuous maps $f\colon X \to Y$ and $g\colon Y \to X$ such that $f \circ g \sim \mathrm{id}_Y$ and $g \circ f \sim \mathrm{id}_X$.

Two homeomorphic spaces are homotopically equivalent, but the converse may not hold. For instance, the euclidean space $\mathbb{R}^n$ is homotopically equivalent to a point for every n. A topological space that is homotopically equivalent to a point is called *contractible*.

1.2. Algebraic topology

1.2.1. Fundamental group. Let X be a topological space and $x_0 \in X$ a base point. The *fundamental group* of the pair (X, x_0) is a group

$$\pi_1(X, x_0)$$

defined by taking all *loops*, that is all paths starting and ending at x_0, considered up to homotopies with fixed endpoints. Loops may be concatenated, and this operation gives a group structure to $\pi_1(X, x_0)$.

If x_1 is another base point, every arc from x_0 to x_1 defines an isomorphism between $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$. Therefore if X is path-connected the fundamental group is base point independent, at least up to isomorphisms, and we write it as $\pi_1(X)$. If $\pi_1(X)$ is trivial we say that X is *simply connected*.

Every continuous map $f\colon X \to Y$ between topological spaces induces a homomorphism

$$f_*\colon \pi_1(X, x_0) \longrightarrow \pi_1\big(Y, f(x_0)\big).$$

The transformation from f to f_* is a *functor* from the category of pointed topological spaces to that of groups. This means that $(f \circ g)_* = f_* \circ g_*$ and $(\mathrm{id}_X)_* = \mathrm{id}_{\pi_1(X,x_0)}$. It implies in particular that homeomorphic spaces have isomorphic fundamental groups.

Exercise 1.2.1. Every topological connected manifold M has a countable fundamental group.

Hint. Since M has a countable base, we may find an open covering of M that consists of countably many open sets homeomorphic to open balls called *islands*. Every pair of such sets intersect in an open set that has at most countably many connected components called *bridges*. Every loop in $\pi_1(M, x_0)$ may be determined by a (non unique!) finite sequence of symbols saying which islands and bridges it crosses. There are only countably many sequences. □

Two maps $f, g\colon (X, x_0) \to (Y, y_0)$ that are homotopic, via a homotopy that sends x_0 to y_0 at each time, induce the same homomorphisms $f_* = g_*$ on fundamental groups. This implies that homotopically equivalent path-connected spaces have isomorphic fundamental groups, so in particular every contractible topological space is simply connected.

There are simply connected manifolds that are not contractible, as we will discover in the next chapters.

1.2.2. Coverings. Let $\tilde{X}$ and X be two path-connected topological spaces. A continuous surjective map $p\colon \tilde{X} \to X$ is a *covering map* if every $x \in X$ has an open neighbourhood U such that

$$p^{-1}(U) = \bigsqcup_{i \in I} U_i$$

where U_i is open and $p|_{U_i}\colon U_i \to U$ is a homeomorphism for all $i \in I$.

A *local homeomorphism* is a continuous map $f: X \to Y$ where every $x \in X$ has an open neighbourhood U such that $f(U)$ is open and $f|_U\colon U \to f(U)$ is a homeomorphism. A covering map is always a local homeomorphism, but the converse may not hold.

The *degree* of a covering $p\colon \tilde{X} \to X$ is the cardinality of a fibre $p^{-1}(x)$ of a point x, a number which does not depend on x.

Two coverings $p\colon \tilde{X} \to X$ and $p'\colon \tilde{X}' \to X$ of the same space X are *isomorphic* if there is a homeomorphism $f: \tilde{X} \to \tilde{X}'$ such that $p = p' \circ f$.

1.2.3. Coverings and fundamental group. One of the most beautiful aspects of algebraic topology is the exceptionally strong connection between fundamental groups and covering maps.

Let $p\colon \tilde{X} \to X$ be a covering map. We fix a basepoint $x_0 \in X$ and a lift $\tilde{x}_0 \in p^{-1}(x_0)$ in the fibre of x_0. The induced homomorphism

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

is always injective. If we modify $\tilde{x}_0$ in the fibre of x_0, the image subgroup $\operatorname{Im} p_*$ changes only by a conjugation inside $\pi_1(X, x_0)$. The degree of p equals the index of $\operatorname{Im} p_*$ in $\pi_1(X, x_0)$.

A topological space Y is *locally contractible* if every point $y \in Y$ has a contractible neighbourhood. This is again a very reasonable assumption: every topological space considered in this book will be of this kind.

We now consider a connected and locally contractible topological space X and fix a base-point $x_0 \in X$.

Theorem 1.2.2. *By sending p to $\operatorname{Im} p_*$ we get a bijective correspondence*

$$\left\{ \begin{array}{c} \text{coverings } p\colon \tilde{X} \to X \\ \text{up to isomorphism} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{subgroups of } \pi_1(X, x_0) \\ \text{up to conjugacy} \end{array} \right\}$$

The covering corresponding to the trivial subgroup is called the *universal covering*. In other words, a covering $\tilde{X} \to X$ is *universal* if $\tilde{X}$ is simply connected, and we have just discovered that this covering is unique up to isomorphism.

Exercise 1.2.3. Let $p\colon \tilde{X} \to X$ be a covering map. If X is a topological manifold, then $\tilde{X}$ also is.

Hint. To lift a countable base from X to $\tilde{X}$, use that $\pi_1(X)$ is countable by Exercise 1.2.1 and hence p has countable degree. $\square$

1.2.4. Deck transformations. Let $p\colon \tilde{X} \to X$ be a covering map. A *deck transformation* or *automorphism* for p is a homeomorphism $f : \tilde{X} \to \tilde{X}$ such that $p \circ f = p$. The deck transformations form a group $\operatorname{Aut}(p)$ called the *deck transformation group* of p.

If $\operatorname{Im} p_*$ is a normal subgroup, the covering map is called *regular*. For instance, the universal cover is regular. Regular covering maps behave nicely in many aspects: for instance we have a natural isomorphism

$$\operatorname{Aut}(p) \cong \pi_1(X)/_{\pi_1(\tilde{X})}.$$

To be more specific, we need to recall some basic notions on group actions.

1.2.5. Group actions. An *action* of a group G on a set X is a group homomorphism

$$\rho\colon G \longrightarrow S(X)$$

where $S(X)$ is the group of all the bijections $X \to X$. We denote $\rho(g)$ simply by g, and hence write $g(x)$ instead of $\rho(g)(x)$. We note that

$$g(h(x)) = (gh)(x), \qquad e(x) = x$$

for every $g, h \in G$ and $x \in X$. In particular if $g(x) = y$ then $g^{-1}(y) = x$.

The *stabiliser* of a point $x \in X$ is the subgroup $G_x < G$ consisting of all the elements g such that $g(x) = x$. The *orbit* of a point $x \in X$ is the subset

$$O(x) = \big\{ g(x) \bigm| g \in G \big\} \subset X.$$

Exercise 1.2.4. We have $x \in O(x)$. Two orbits $O(x)$ and $O(y)$ either coincide or are disjoint. They coincide $\iff \exists g \in G$ such that $g(x) = y$.

Therefore the set X is partitioned into orbits. The action is:

- *transitive* if for every $x, y \in X$ there is a $g \in G$ such that $g(x) = y$;
- *faithful* if ρ is injective;
- *free* if the stabiliser of every point is trivial, that is $g(x) \neq x$ for every $x \in X$ and every non-trivial $g \in G$.

Exercise 1.2.5. The stabilisers G_x and G_y of two points x, y lying in the same orbit are conjugate subgroups of G.

Exercise 1.2.6. There is a natural bijection between the left cosets of G_x in G and the elements of $O(x)$. In particular the cardinality of $O(x)$ equals the index $[G : G_x]$ of G_x in G.

The space of all the orbits is denoted by $X/_G$. We have a natural projection $\pi\colon X \to X/_G$.

1.2.6. Topological actions. If X is a topological space, a *topological action* of a group G on X is a homomorphism

$$G \longrightarrow \mathrm{Homeo}(X)$$

where $\mathrm{Homeo}(X)$ is the group of all the self-homeomorphisms of X. We have a natural projection $\pi \colon X \to X/_G$ and we equip the quotient set $X/_G$ with the quotient topology. The action is:

- *properly discontinuous* if any two points $x, y \in X$ have neighbourhoods U_x and U_y such that the set

  $$\left\{ g \in G \;\middle|\; g(U_x) \cap U_y \neq \varnothing \right\}$$

  is finite.

Example 1.2.7. The action of a finite group G is always properly discontinuous.

This definition is relevant mainly because of the following remarkable fact.

Proposition 1.2.8. *Let G act on a Hausdorff path-connected space X. The following are equivalent:*

(1) *G acts freely and properly discontinuously;*
(2) *the quotient $X/_G$ is Hausdorff and $X \to X/_G$ is a regular covering.*

Every regular covering between Hausdorff path-connected spaces arises in this way.

Concerning the last sentence: if $\tilde{X} \to X$ is a regular covering, the deck transformation group G acts transitively on each fibre, and we get $X = \tilde{X}/_G$. This does not hold for non-regular coverings.

We have here a formidable and universal tool to construct plenty of regular coverings and of topological spaces: it suffices to have X and a group G acting freely and properly discontinously on it.

Since every universal cover is regular, we also get the following.

Corollary 1.2.9. *Every path-connected locally contractible Hausdorff topological space X is the quotient $\tilde{X}/_G$ of its universal cover by the action of some group G acting freely and properly discontinuously.*

Note that the group G is isomorphic to $\pi_1(X)$. There are plenty of examples of this phenomenon, but in this introductory chapter we limit ourselves to a very basic one. More will come later.

Example 1.2.10. Let $G = \mathbb{Z}$ act on $X = \mathbb{R}$ as translations, that is $g(v) = v + g$. The action is free and properly discontinuous; hence we get a covering $\mathbb{R} \to \mathbb{R}/_\mathbb{Z}$. The quotient $\mathbb{R}/_\mathbb{Z}$ is in fact homeomorphic to S^1 (exercise).

In principle, one could now think of classifying all the (locally contractible, path-connected, Hausdorff) topological spaces by looking only at the simply connected ones and then studying the groups acting freely and properly discontinuously on them. It is of course impossible to carry on this too ambitious strategy in this wide generality, but the task becomes more reasonable if one restricts the attention to spaces of some particular kind like – as we will see – the riemannian manifolds having constant curvature.

Recall that a continuous map $f\colon X \to Y$ is *proper* if $f^{-1}(K)$ is compact for every compact $K \subset Y$.

Exercise 1.2.11. Let a group G act on a locally compact space X. Assign to G the discrete topology. The following are equivalent:

- the action is properly discontinuous;
- for every compact $K \subset X$, the set $\{g \mid g(K) \cap K \neq \varnothing\}$ is finite;
- the map $G \times X \to X \times X$ that sends (g, x) to $(g(x), x)$ is proper.

1.3. Multivariable analysis

It will be important in this book to use superscripts and subscripts in a globally coherent way, and to obey this rule (to be explained later on) we will employ superscripts $x^1, \ldots, x^n$ to indicate the coordinates of a vector $x \in \mathbb{R}^n$. At some points we will break this rule and use subscripts $x_1, \ldots, x_n$ only to avoid cumbersome formulas.

1.3.1. Smooth maps. A map $f\colon U \to V$ between two open sets $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ is C^∞ or *smooth* if it has partial derivatives of any order. All the maps considered in this book will be smooth.

In particular, for every $p \in U$ we have a *differential*

$$df_p\colon \mathbb{R}^n \longmapsto \mathbb{R}^m$$

which is the linear map that best approximates f near p, that is we get

$$f(x) = f(p) + df_p(x - p) + o\big(\|x - p\|\big).$$

If we see df_p as a $m \times n$ matrix, it is called the *Jacobian* and we get

$$df_p = \left(\frac{\partial f}{\partial x^1} \cdots \frac{\partial f}{\partial x^n}\right) = \begin{pmatrix} \frac{\partial f_1}{\partial x^1} & \cdots & \frac{\partial f_1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x^1} & \cdots & \frac{\partial f_m}{\partial x^n} \end{pmatrix}.$$

A fundamental property of differentials is the *chain rule*: if we are given two smooth functions

$$U \xrightarrow{f} V \xrightarrow{g} W$$

then for every $p \in U$ we have

$$d(g \circ f)_p = dg_{f(p)} \circ df_p.$$

1.3.2. Taylor Theorem. A *multi-index* is a vector $\alpha = (\alpha_1, \ldots, \alpha_n)$ of non-negative integers $\alpha_i \geq 0$. We set

$$|\alpha| = \alpha_1 + \ldots + \alpha_n, \qquad \alpha! = \alpha_1! \cdots \alpha_n!, \qquad x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

Let $f: U \to \mathbb{R}$ be a smooth map defined on some open set $U \subset \mathbb{R}^n$. For every multi-index α we define the corresponding combination of partial derivatives:

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

We recall Taylor's Theorem:

Theorem 1.3.1. *Let $f: U \to \mathbb{R}$ be a smooth map defined on some open convex set $U \subset \mathbb{R}^n$. For every point $x_0 \in U$ and integer $k \geq 0$ we have*

$$f(x) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x_0)}{\alpha!}(x - x_0)^\alpha + \sum_{|\alpha| = k+1} h_\alpha(x)(x - x_0)^\alpha$$

where $h_\alpha: U \to \mathbb{R}$ is a smooth map that depends on α.

1.3.3. Diffeomorphisms. A smooth map $f: U \to V$ between two open sets $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ is a *diffeomorphism* if it is invertible and its inverse $f^{-1}: V \to U$ is also smooth.

Proposition 1.3.2. *If f is a diffeomorphism, then df_p is invertible for every $p \in U$. In particular we get $n = m$.*

Proof. The chain rule gives

$$\mathrm{id}_{\mathbb{R}^n} = d(\mathrm{id}_U)_p = d(f^{-1} \circ f)_p = df^{-1}_{f(p)} \circ df_p,$$

$$\mathrm{id}_{\mathbb{R}^m} = d(\mathrm{id}_V)_{f(p)} = d(f \circ f^{-1})_{f(p)} = df_p \circ df^{-1}_{f(p)}.$$

Therefore the linear map df_p is invertible. □

We now show that a weak converse of this statement holds.

1.3.4. Local diffeomorphisms. We say that a smooth map $f: U \to V$ is a *local diffeomorphism* at a point $p \in U$ if there is an open neighbourhood $U' \subset U$ of p such that $f(U')$ is open and $f|_{U'}: U' \to f(U')$ is a diffeomorphism.

Here is an important theorem, that we will use frequently.

Theorem 1.3.3 (Inverse Function Theorem). *A smooth map $f: U \to V$ is a local diffeomorphism at $p \in U \iff$ its differential df_p is invertible.*

We say that a smooth map $f: U \to V$ is a *local diffeomorphism* if it is so at every point $p \in U$. A diffeomorphism is always a local diffeomorphism, but the converse does not hold as the following example shows.

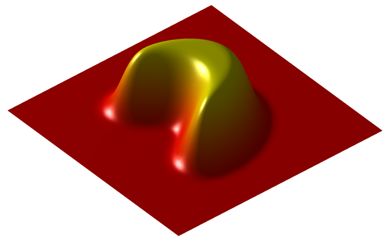

Figure 1.3. A smooth bump function $f\colon \mathbb{R}^2 \to \mathbb{R}$.

Example 1.3.4. The smooth map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}$$

has Jacobian

$$df_{(x,y)} = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}$$

with determinant e^{2x} and hence everywhere invertible. By the Inverse Function Theorem, the map f is a local diffeomorphism. The map f is however not injective, hence it is not a diffeomorphism.

1.3.5. Bump functions. A *smooth bump function* is a smooth function $\rho\colon \mathbb{R}^n \to \mathbb{R}$ that has compact support (the *support* is the closure of the set of points $x \in \mathbb{R}^n$ where $\rho(x) \neq 0$). See Figure 1.3.

The existence of bump functions is a peculiar feature of the smooth environment that has many important consequences in differential topology. The main tool is the smooth function

$$h(t) = \begin{cases} e^{-\frac{1}{t}} & \text{if } t \geq 0, \\ 0 & \text{if } t \leq 0. \end{cases}$$

We may use it to build a bump function $\rho\colon \mathbb{R}^n \to \mathbb{R}$ as follows:

$$\rho(x) = h\big(1 - \|x\|^2\big).$$

The support of ρ is the closed unit disc $\|x\| \leq 1$, and it has a unique maximum at the origin $x = 0$.

Note that a bump function is never analytic (unless it is constantly zero). Sometimes it is useful to have a bump function that looks like a *plateau*, for instance consider $\eta\colon \mathbb{R}^n \to \mathbb{R}$ defined as follows:

$$\eta(x) = \frac{h\big(1 - \|x\|^2\big)}{h\big(1 - \|x\|^2\big) + h\big(\|x\|^2 - \frac{1}{4}\big)}.$$

Here $\eta(x) = 1$ for all $\|x\| \leq \frac{1}{2}$ and $\eta(x) = 0$ for all $\|x\| \geq 1$, while $\eta(x) \in (0, 1)$ for all $\frac{1}{2} < \|x\| < 1$.

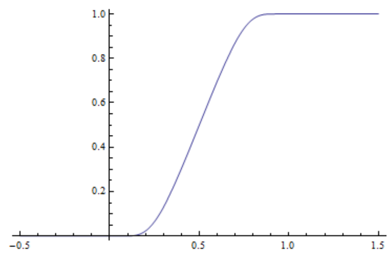

Figure 1.4. A smooth transition function Ψ.

1.3.6. Transition function. Another important smooth non-analytic functions is the *transition function* $\Psi\colon \mathbb{R} \to \mathbb{R}$ defined as

$$\Psi(x) = \frac{h(x)}{h(x) + h(1-x)}$$

where $h(x)$ is the function defined above. The function Ψ is smooth and non-decreasing, and we have $\Psi(x) = 0$ for all $x \leq 0$ and $\Psi(x) = 1$ for all $x \geq 1$. See Figure 1.4.

1.3.7. Cauchy–Lipschitz Theorem. The Cauchy–Lipschitz Theorem certifies the existence and uniqueness of solutions of a system of first-order differential equations, and also the smooth dependence on its initial values, when the given equations are smooth.

Let $f\colon I \times \mathbb{R}^n \to \mathbb{R}^n$ be a smooth map, with $I \subset \mathbb{R}$ some interval.

Theorem 1.3.5. *The Cauchy problem*

$$\begin{cases} x'(t) &= f(t, x(t)), \\ x(0) &= x_0 \end{cases}$$

has a unique solution $x(t)$, defined on some maximal open interval $J \subset I$. The point $x(t)$ depends smoothly on both t and $x_0 \in \mathbb{R}^n$.

If we have a higher order differential equation

$$x^{(n)}(t) = f(t, x'(t), x''(t), \ldots, x^{(n-1)}(t))$$

we can reduce it to a system of first-order equations as above, with variables $x_1 = x, x_2, \ldots, x_n$ and equations $x_i'(t) = x_{i+1}(t)$ and $x_n' = f(t, x_1, \ldots, x_n)$. Therefore we have again a unique smooth solution $x(t)$ for any arbitrarily fixed initial values of $x(0)$, $x'(0), \ldots, x^{(n-1)}(0)$.

If the solution $x(t)$ is defined on some maximal interval $J = (a, b)$ and $b < +\infty$, then $x(t)$ must diverge (that is, exit from any compact set) as $t \to b$, otherwise (one can prove that) the solution could be prolonged on some bigger open interval and J would not be maximal.

1.3.8. Integration. A *Borel set* $V \subset \mathbb{R}^n$ is any subset constructed from the open and closed sets by countable unions and intersections.

If $V \subset \mathbb{R}^n$ is a Borel set and $f : V \to \mathbb{R}$ is a non-negative measurable function, we may consider its *Lebesgue integral*

$$\int_V f.$$

If $\varphi : U \to V$ is a diffeomorphism between two open subsets of $\mathbb{R}^n$, then we get the following *changes of variables* formula

$$\int_{V'} f = \int_{U'} |\det d\varphi| f \circ \varphi$$

for any Borel subsets $U' \subset U$ and $V' = \varphi(U')$.

Remark 1.3.6. A diffeomorphism of course does not preserve the measure of Borel sets, but it sends zero-measure sets to zero-measure sets.

1.3.9. The Sard Lemma. Let $f : U \to \mathbb{R}^n$ be a smooth map defined on some open subset $U \subset \mathbb{R}^m$. We say that a point $p \in U$ is *regular* if the differential df_p is surjective, and *singular* otherwise. A value $q \in \mathbb{R}^n$ is a *regular value* if all its counterimages $p \in f^{-1}(q)$ are regular points, and *singular* otherwise.

Here is an important fact on smooth maps.

Lemma 1.3.7 (Sard's Lemma). *The singular values of f form a zero-measure subset of $\mathbb{R}^n$.*

Corollary 1.3.8. *If $m < n$, the image of f is a zero-measure subset.*

Recall that a *Peano curve* is a continuous surjection $\mathbb{R} \to \mathbb{R}^2$. Maps of this kind are forbidden in the smooth world.

1.3.10. Complex analysis. Let $U, V \subset \mathbb{C}$ be open subsets. Recall that a function $f : U \to V$ is *holomorphic* if for every $z_0 \in U$ the limit

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. The limit $f'(z_0)$ is a complex number called the *complex derivative* of f at z_0.

Quite surprisingly, a homolorphic function satisfies a wealth of very good properties: if we identify $\mathbb{C}$ with $\mathbb{R}^2$ in the usual way, we may interpret f as a function between open sets of $\mathbb{R}^2$, and it turns out that f is smooth (and even analytic) and its Jacobian at z_0 is such that

$$\det(df_{z_0}) = |f'(z_0)|^2.$$

It is indeed a remarkable fact that the presence of the complex derivative alone guarantees the existence of partial derivatives of any order.

1.4. Projective geometry

1.4.1. Projective spaces. Let $\mathbb{K}$ be any field: we will be essentially interested in the cases $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Let V be a finite-dimensional vector space on $\mathbb{K}$. The *projective space* of V is

$$\mathbb{P}(V) = \big(V \setminus \{0\}\big)/_{\sim}$$

where $v \sim w \iff v = \lambda w$ for some $\lambda \neq 0$. In particular we write

$$\mathbb{KP}^n = \mathbb{P}(\mathbb{K}^{n+1}).$$

Every non-zero vector $v = (x_0, \ldots, x_n) \in \mathbb{K}^{n+1}$ determines a point in $\mathbb{KP}^n$ which we denote as

$$[x_0, \ldots, x_n].$$

These are the *homogeneous coordinates* of the point. Of course not all the x_i are zero, and $[x_0, \ldots, x_n] = [\lambda x_0, \ldots, \lambda x_n]$ for all $\lambda \neq 0$.

1.4.2. Topology. When $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, the space $\mathbb{KP}^n$ inherits the quotient topology from $\mathbb{K}^{n+1}$ and is a Hausdorff compact connected topological space. A convenient way to see this is to consider the projections

$$\pi \colon S^n \longrightarrow \mathbb{RP}^n, \qquad \pi \colon S^{2n+1} \longrightarrow \mathbb{CP}^n$$

obtained by restricting the projections from $\mathbb{R}^n \setminus \{0\}$ and $\mathbb{C}^n \setminus \{0\}$. Note that

$$S^{2n+1} = \big\{ z \in \mathbb{C}^{n+1} \bigm| |z_0|^2 + \ldots + |z_n|^2 = 1 \big\}.$$

Exercise 1.4.1. Show that the projections are surjective and deduce that the projective spaces are connected and compact.

Exercise 1.4.2. We have the following homeomorphisms

$$\mathbb{RP}^1 \cong S^1, \qquad \mathbb{CP}^1 \cong S^2.$$

The fundamental group of $\mathbb{RP}^n$ is $\mathbb{Z}$ when $n = 1$ and $\mathbb{Z}/_{2\mathbb{Z}}$ when $n > 1$. On the other hand the complex projective space $\mathbb{CP}^n$ is simply connected for every n.

CHAPTER 2

Tensors

2.1. Multilinear algebra

2.1.1. The dual space. In this book we will be concerned mostly with real finite-dimensional vector spaces. Given two such spaces V, W of dimension m, n, we denote by $\mathrm{Hom}(V, W)$ the set of all the linear maps $V \to W$. The set $\mathrm{Hom}(V, W)$ is itself naturally a vector space of dimension mn.

A space that will be quite relevant here is the *dual space* $V^* = \mathrm{Hom}(V, \mathbb{R})$, that consists of all the linear functionals $V \to \mathbb{R}$, also called *covectors*. The spaces V and V^* have the same dimension, but there is no canonical way to choose an isomorphism $V \to V^*$ between them: this fact will have important consequences in this book.

A basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ for V induces a *dual basis* $\mathcal{B}^* = \{v^1, \ldots, v^n\}$ for V^* by requiring that $v^i(v_j) = \delta_{ij}$. (Recall that the *Kronecker delta* δ_{ij} equals 1 if $i = j$ and 0 otherwise.) We can construct an isomorphism $V \to V^*$ by sending v_i to v^i, but it heavily depends on the chosen basis $\mathcal{B}$.

On the other hand, a canonical isomorphism $V \to V^{**}$ exists between V and its *bidual space* $V^{**} = (V^*)^*$. The isomorphism is the following:

$$v \longmapsto \big(v^* \longmapsto v^*(v)\big).$$

Exercise 2.1.1. This is indeed an isomorphism. If V had infinite dimension, it would be injective and not surjective.

For that reason, the bidual space V^{**} will play no role here and will always be identified with V. In fact, it is useful to think of V and V^* as related by a bilinear pairing

$$V \times V^* \longrightarrow \mathbb{R}$$

that sends (v, v^*) to $v^*(v)$. Not only the vectors in V^* act on V, but also the vectors in V act on V^*.

Every linear map $L \colon V \to W$ induces an *adjoint* linear map $L^* \colon W^* \to V^*$ that sends f to $f \circ L$. Of course we get $L^{**} = L$.

2.1.2. Multilinear maps. Given some vector spaces $V_1, \ldots, V_k, W$, a map

$$F \colon V_1 \times \cdots \times V_k \longrightarrow W$$

is *multilinear* if it is linear on each component.

Let $\mathcal{B}_i = \{v_{i,1}, \ldots, v_{i,m_i}\}$ be a basis of V_i and $\mathcal{C} = \{w_1, \ldots, w_n\}$ a basis of W. The *coefficients* of F with respect to these basis are the numbers

$$F^j_{j_1,\ldots,j_k}$$

with $1 \leq j_i \leq m_i$ and $1 \leq j \leq n$ such that

$$F(v_{1,j_1}, \ldots, v_{k,j_k}) = \sum_{j=1}^{n} F^j_{j_1,\ldots,j_k} w_j.$$

Exercise 2.1.2. Every multilinear F is determined by its coefficients, and every choice of coefficients determines a multilinear F.

We denote by $\operatorname{Mult}(V_1, \ldots, V_k; W)$ the space of all the multilinear maps $V_1 \times \cdots \times V_k \to W$. This is naturally a vector space.

Corollary 2.1.3. *We have*

$$\dim \operatorname{Mult}(V_1, \ldots, V_k; W) = \dim V_1 \cdots \dim V_k \dim W.$$

When $W = \mathbb{R}$ we omit it from the notation and write $\operatorname{Mult}(V_1, \ldots, V_k)$. In that case of course we have

$$\dim \operatorname{Mult}(V_1, \ldots, V_k) = \dim V_1 \cdots \dim V_k.$$

In fact, every space $\operatorname{Mult}(V_1, \ldots, V_k; W)$ may be transformed canonically into a similar one where the target vector space is $\mathbb{R}$, thanks to the following:

Exercise 2.1.4. There is a canonical isomorphism

$$\operatorname{Mult}(V_1, \ldots, V_k; W) \longrightarrow \operatorname{Mult}(V_1, \ldots, V_k, W^*)$$

defined by sending $F \in \operatorname{Mult}(V_1, \ldots, V_k; W)$ to the map

$$(v_1, \ldots, v_k, w^*) \longmapsto w^*\big(F(v_1, \ldots, v_k)\big).$$

Hint. The spaces have the same dimension and the map is injective. $\square$

2.1.3. Sum and product of spaces. We now introduce a couple of operations $\oplus$ and $\otimes$ on vector spaces. Let $V_1, \ldots, V_k$ be some real finite-dimensional vector spaces.

Sum. The *sum* $V_1 \oplus \cdots \oplus V_k$ is just the cartesian product with componentwise vector space operations. That is:

$$V_1 \oplus \cdots \oplus V_k = \big\{(v_1, \ldots, v_k) \bigm| v_1 \in V_1, \ldots, v_k \in V_k\big\}$$

and the vector space operations are

$$\begin{aligned}(v_1, \ldots, v_k) + (w_1, \ldots, w_k) &= (v_1 + w_1, \ldots, v_k + w_k),\\ \lambda(v_1, \ldots, v_k) &= (\lambda v_1, \ldots, \lambda v_k).\end{aligned}$$

Let $\mathcal{B}_i = \{v_{i,1}, \ldots, v_{i,m_i}\}$ be a basis of V_i, for all $i = 1, \ldots, k$.

Exercise 2.1.5. A basis for $V_1 \oplus \cdots \oplus V_k$ is

$$\left\{(v_{1,j_1}, 0, \ldots, 0), \ldots, (0, \ldots, 0, v_{i,j_i}, 0, \ldots, 0), \ldots, (0, \ldots, 0, v_{k,j_k})\right\}$$

where $1 \leq j_i \leq m_i$ varies for each $i = 1, \ldots, k$.

We deduce that

$$\dim(V_1 \oplus \cdots \oplus V_k) = \dim V_1 + \ldots + \dim V_k.$$

Tensor product. The *tensor product* $V_1 \otimes \cdots \otimes V_k$ is defined (a bit more obscurely...) as the space of all the multilinear maps $V_1^* \times \cdots \times V_k^* \to \mathbb{R}$, *i.e.*

$$V_1 \otimes \cdots \otimes V_k = \mathrm{Mult}(V_1^*, \ldots, V_k^*).$$

We already know that

$$\dim(V_1 \otimes \cdots \otimes V_k) = \dim V_1 \cdots \dim V_k.$$

Any k vectors $v_1 \in V_1, \ldots, v_k \in V_k$ determine an element

$$v_1 \otimes \cdots \otimes v_k \in V_1 \otimes \cdots \otimes V_k$$

which is by definition the multilinear map

$$(v_1^*, \ldots, v_k^*) \longmapsto v_1^*(v_1) \cdots v_k^*(v_k).$$

As opposite to the sum operation, it is important to note that *not* all the elements of $V_1 \otimes \cdots \otimes V_k$ are of the form $v_1 \otimes \cdots \otimes v_k$. The elements of this type (sometimes called *pure* or *simple*) can however generate the space, as the next proposition shows. Let $\mathcal{B}_i = \{v_{i,1}, \ldots, v_{i,m_i}\}$ be a basis of V_i for all $1 \leq i \leq k$.

Proposition 2.1.6. *A basis for the tensor product* $V_1 \otimes \cdots \otimes V_k$ *is*

$$\{v_{1,j_1} \otimes \cdots \otimes v_{k,j_k}\}$$

where $1 \leq j_i \leq m_i$ *varies for each* $i = 1, \ldots, k$.

Proof. This is a consequence of Exercise 2.1.2. If we use the dual basis for V_i^*, the element $v_{1,j_1} \otimes \cdots \otimes v_{k,j_k}$ corresponds to the multilinear map whose coefficients $F^1_{i_1,\ldots,i_k}$ equal 1 if $(i_1, \ldots, i_k) = (j_1, \ldots, j_k)$ and 0 otherwise. $\square$

Example 2.1.7. A basis for $\mathbb{R}^2 \otimes \mathbb{R}^2$ is given by the elements

$$\begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix}, \quad \begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix}, \quad \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix}, \quad \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix}.$$

Exercise 2.1.8. The following relations hold in $V \otimes W$:

$$(v + v') \otimes w = v \otimes w + v' \otimes w, \qquad v \otimes (w + w') = v \otimes w + v \otimes w',$$
$$\lambda(v \otimes w) = (\lambda v) \otimes w = v \otimes (\lambda w),$$
$$v \otimes w = 0 \iff v = 0 \text{ or } w = 0.$$

Exercise 2.1.9. Let $v, v' \in V$ and $w, w' \in W$ be non-zero vectors. If v and v' are independent, then $v \otimes w$ and $v' \otimes w'$ also are.

Exercise 2.1.10. Let $v, v' \in V$ and $w, w' \in W$ be two pairs of independent vectors. Show that

$$v \otimes w + v' \otimes w' \in V \otimes W$$

is not a pure element.

2.1.4. Canonical isomorphisms. We now introduce some canonical isomorphisms, that may look quite abstract at a first sight, but that will help us a lot to simplify many situations: two spaces that are canonically isomorphic may be harmlessly considered as the same space.

We start with the following easy:

Proposition 2.1.11. *The map $v \mapsto v \otimes 1$ defines a canonical isomorphism*

$$V \longrightarrow V \otimes \mathbb{R}.$$

Proof. The spaces have the same dimension and the map is linear and injective by Exercise 2.1.8. $\square$

Let $V_1, \ldots, V_k, Z$ be any vector spaces.

Proposition 2.1.12. *The linear map*

$$\mathrm{Hom}(V_1 \otimes \cdots \otimes V_k, Z) \longrightarrow \mathrm{Mult}(V_1, \ldots, V_k; Z)$$

that sends F to F' via $F'(v_1, \ldots, v_k) = F(v_1 \otimes \cdots \otimes v_k)$ is an isomorphism.

Proof. The spaces have the same dimension and the map is injective (exercise: use Proposition 2.1.6). $\square$

This canonical isomorphism is called the *universal property* of $\otimes$ and one can also show that it characterises the tensor product uniquely. This is typically stated by drawing a commutative diagram like this:

$$\begin{tikzcd} V_1 \times \cdots \times V_k \arrow[r] \arrow[dr, "F'"'] & V_1 \otimes \cdots \otimes V_k \arrow[d, dotted, "F"] \\ & Z \end{tikzcd} \tag{1}$$

Given a multilinear F' there is a unique linear F so that the diagram commutes. The universal property is very useful to construct maps. For instance, we may use it to construct more canonical isomorphisms:

Proposition 2.1.13. *There are canonical isomorphisms*

$$V \oplus W \cong W \oplus V, \qquad (V \oplus W) \oplus Z \cong V \oplus W \oplus Z \cong V \oplus (W \oplus Z),$$

$$V \otimes W \cong W \otimes V, \qquad (V \otimes W) \otimes Z \cong V \otimes W \otimes Z \cong V \otimes (W \otimes Z),$$

$$V \otimes (W \oplus Z) \cong (V \otimes W) \oplus (V \otimes Z),$$

$$(V_1 \oplus \cdots \oplus V_k)^* \cong V_1^* \oplus \cdots \oplus V_k^*, \qquad (V_1 \otimes \cdots \otimes V_k)^* \cong V_1^* \otimes \cdots \otimes V_k^*.$$

Proof. The isomorphisms in the first line are

$$(v,w)\mapsto(w,v),\qquad (v,w,z)\mapsto\big((v,w),z\big),\qquad (v,w,z)\mapsto\big(v,(w,z)\big).$$

Those in the second line are uniquely determined by the conditions

$$v\otimes w\mapsto w\otimes v,\qquad v\otimes w\otimes z\mapsto(v\otimes w)\otimes z,\qquad v\otimes w\otimes z\mapsto v\otimes(w\otimes z)$$

thanks to the universal property of the tensor products. Analogously the isomorphism of the third line is determined by

$$v\otimes(w,z)\mapsto(v\otimes w,v\otimes z).$$

Concerning the last line, the first isomorphism is straightforward. For the second, we have

$$(V_1\otimes\cdots\otimes V_k)^*=\mathrm{Hom}(V_1\otimes\cdots\otimes V_k,\mathbb{R})=\mathrm{Mult}(V_1,\ldots,V_k)=V_1^*\otimes\cdots\otimes V_k^*.$$

More concretely, every element $v^1\otimes\cdots\otimes v^k\in V_1^*\otimes\cdots\otimes V_k^*$ is naturally an element of $(V_1\otimes\cdots\otimes V_k)^*$ as follows:

$$(v^1\otimes\cdots\otimes v^k)(w_1\otimes\cdots\otimes w_k)=v^1(w_1)\cdots v^k(w_k).$$

The proof is complete. □

There are yet more canonical isomorphisms to discover! The following is a consequence of Exercise 2.1.4 and is particularly useful.

Corollary 2.1.14. *There is a canonical isomorphism*

$$\mathrm{Hom}(V,W)\cong V^*\otimes W.$$

In particular we have $\mathrm{End}(V)\cong V^*\otimes V=\mathrm{Mult}(V,V^*)$. In this canonical isomorphism, the identity endomorphism id_V corresponds to the bilinear map $V\times V^*\to\mathbb{R}$ that sends (v,v^*) to $v^*(v)$.

Exercise 2.1.15. Given $v^*\in V^*$ and $w\in W$, the element $v^*\otimes w$ corresponds via the canonical isomorphism $\mathrm{Hom}(V,W)\cong V^*\otimes W$ to the homomorphism $v\mapsto v^*(v)w$. Deduce that the pure elements in $V^*\otimes W$ correspond precisely to the homomorphisms $V\to W$ of rank ≤ 1.

2.1.5. The Segre embedding. We briefly show a geometric application of the algebra introduced in this section. Let U,V be vector spaces. The natural map $U\times V\to U\otimes V$ induces an injective map on projective spaces

$$\mathbb{P}(U)\times\mathbb{P}(V)\hookrightarrow\mathbb{P}(U\otimes V)$$

called the *Segre embedding*. The map is injective thanks to Exercise 2.1.9.

We have just discovered a simple method for embedding a product of projective spaces in a bigger projective space. If $U=\mathbb{R}^{m+1}$ and $V=\mathbb{R}^{n+1}$ we have an isomorphism $U\otimes V\cong\mathbb{R}^{(m+1)(n+1)}$ and we get an embedding

$$\mathbb{RP}^m\times\mathbb{RP}^n\hookrightarrow\mathbb{RP}^{mn+m+n}.$$

Example 2.1.16. When $m = n = 1$ we get $\mathbb{RP}^1 \times \mathbb{RP}^1 \hookrightarrow \mathbb{RP}^3$. Note that $\mathbb{RP}^1 \times \mathbb{RP}^1$ is topologically a torus. The Segre map is

$$\big([x_0, x_1], [y_0, y_1]\big) \longmapsto \left[\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \otimes \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}\right]$$

and the right member equals

$$\left[x_0 y_0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_0 y_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} + x_1 y_0 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_1 y_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right].$$

In coordinates with respect to the canonical basis the Segre embedding is

$$\big([x_0, x_1], [y_0, y_1]\big) \longmapsto [x_0 y_0, x_0 y_1, x_1 y_0, x_1 y_1].$$

It is now an exercise to show that the image is precisely the quadric $z_0 z_3 = z_1 z_2$ in $\mathbb{RP}^3$. We recover the well-known fact that such a quadric is a torus.

2.1.6. Infinite-dimensional spaces. In very few points in this book we will be concerned with infinite dimensional real vector spaces. We summarise briefly how to extend some of the operations introduced above to an infinite-dimensional context.

The *dual* V^* of a vector space V is always the space of all functionals $V \to \mathbb{R}$. There is a canonical injective map $V \hookrightarrow V^{**}$ which is surjective if and only if V has finite dimension.

Let $V_1, V_2, \ldots$ be vector spaces. The *direct product* and the *direct sum*

$$\prod_i V_i, \qquad \bigoplus_i V_i$$

are respectively the space of all sequences $(v_1, v_2, \ldots)$ with $v_i \in V_i$, and the subspace consisting of sequences with only finitely many non-zero elements. In the latter case, when the spaces V_i are clearly distinct, one may write every sequence simply as a sum

$$v_{i_1} + \ldots + v_{i_h}$$

of the non-zero elements in the sequence. There is a canonical isomorphism

$$(\oplus_i V_i)^* = \textstyle\prod_i V_i^*.$$

The *tensor product* $V \otimes W$ of two vector spaces of arbitrary dimension may be defined as the unique vector space that satisfies the universal property (1). Uniqueness is easy to prove, but existence is more involved: the space $\mathrm{Mult}(V^*, W^*)$ does not work here, it is too big because $V \neq V^{**}$. Instead we may define $V \otimes W$ as a quotient

$$V \otimes W = F(V \times W)/_{\sim}$$

where $F(S)$ is the *free vector space* generated by the set S, that is the abstract vector space with basis S, and $\sim$ is the equivalence relation generated by equivalences of this type:

$$(v_1, w) + (v_2, w) \sim (v_1 + v_2, w),$$
$$(v, w_1) + (v, w_2) \sim (v, w_1 + w_2),$$
$$(\lambda v, w) \sim \lambda(v, w) \sim (v, \lambda w).$$

The equivalence class of (v, w) is indicated as $v \otimes w$. More concretely, if $\{v_i\}$ and $\{w_j\}$ are basis of V and W, then $\{v_i \otimes w_j\}$ is a basis of $V \otimes W$, and this is the most important thing to keep in mind.

The tensor product is distributive with respect to direct sum, that is there are canonical isomorphisms

$$V \otimes \big(\oplus_i W_i\big) \cong \oplus_i (V \otimes W_i)$$

but the tensor product is *not* distributive with respect to the direct product in general! We need $\dim V < \infty$ for that:

Exercise 2.1.17. If V has finite dimension, there is a canonical isomorphism

$$V \otimes \big(\textstyle\prod_i W_i\big) \cong \textstyle\prod_i (V \otimes W_i).$$

Dimostrare?

2.2. Tensors

We have defined the operations $\oplus, \otimes, *$ in full generality, and we now apply them to a single finite-dimensional real vector space V.

2.2.1. Definition. Let V be a real vector space of dimension n and $h, k \geq 0$ some integers. A *tensor* of type (h, k) is an element T of the vector space

$$\mathcal{T}_h^k(V) = \underbrace{V \otimes \cdots \otimes V}_{h} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{k}.$$

In other words T is a multilinear map

$$T\colon \underbrace{V^* \times \cdots \times V^*}_{h} \times \underbrace{V \times \cdots \times V}_{k} \longrightarrow \mathbb{R}.$$

This elegant definition gathers many well-known notions in a single word:

- a tensor of type $(0, 0)$ is by convention an element of $\mathbb{R}$, a *scalar*;
- a tensor of type $(1, 0)$ is an element of V, a *vector*;
- a tensor of type $(0, 1)$ is an element of V^*, a *covector*;
- a tensor of type $(0, 2)$ is a *bilinear form* $V \times V \to \mathbb{R}$;
- a tensor of type $(1, 1)$ is an element of $V \otimes V^*$ and hence may be interpreted as an *endomorphism* $V \to V$, by Corollary 2.1.14;

More generally, every tensor T of type (h, k) may be interpreted as a multilinear map

$$T'\colon \underbrace{V \times \cdots \times V}_{k} \longrightarrow \underbrace{V \otimes \cdots \otimes V}_{h}$$

by writing

$$T'(v_1,\ldots,v_k)(v_1^*,\ldots,v_h^*) = T(v_1^*,\ldots,v_h^*,v_1,\ldots,v_k).$$

In particular a tensor of type $(1,k)$ can be interpreted as a multilinear map

$$T\colon \underbrace{V\times\cdots\times V}_{k} \longrightarrow V.$$

Example 2.2.1. The *euclidean scalar product* in $\mathbb{R}^n$ is defined as

$$(x^1,\ldots,x^n)\cdot(y^1,\ldots,y^n) = x^1y^1+\ldots+x^ny^n.$$

It is a bilinear map $\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ and hence a tensor of type $(0,2)$.

Example 2.2.2. The *cross product* in $\mathbb{R}^3$ is defined as

$$(x,y,z)\wedge(x',y',z') = (yz'-zy', zx'-xz', xy'-yx').$$

It is a bilinear map $\mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}^3$ and hence a tensor of type $(1,2)$.

Example 2.2.3. The *determinant* may be interpreted as a multilinear map

$$\underbrace{\mathbb{R}^n\times\cdots\times\mathbb{R}^n}_{n} \longrightarrow \mathbb{R}$$

that sends $(v_1,\ldots,v_n)$ to $\det(v_1\cdots v_n)$. As such, it is a tensor of type $(0,n)$.

2.2.2. Coordinates. Every abstract and ethereal object in linear algebra transforms into a more reassuring multidimensional array of numbers, called *coordinates*, as soon as we choose a basis.

Let $\mathcal{B}=\{v_1,\ldots,v_n\}$ be a basis of V, and $\mathcal{B}^*=\{v^1,\ldots,v^n\}$ be the dual basis of V^*. A basis of the tensor space $\mathcal{T}_h^k(V)$ consists of all the vectors

$$v_{i_1}\otimes\cdots\otimes v_{i_h}\otimes v^{j_1}\otimes\cdots\otimes v^{j_k}$$

where $1\leq i_1,\ldots,i_h,j_1,\ldots,j_k\leq n$. Overall, this basis consists of n^{h+k} vectors. Every tensor T of type (h,k) can be written uniquely as

$$T = T^{i_1,\ldots,i_h}_{j_1,\ldots,j_k} v_{i_1}\otimes\cdots\otimes v_{i_h}\otimes v^{j_1}\otimes\cdots\otimes v^{j_k}. \tag{2}$$

We are using here the *Einstein summation convention*: every index that is repeated at least twice should be summed over the values of the index. Therefore in (2) we sum over all the indices $i_1,\ldots,i_h,j_1,\ldots,j_k$. The following proposition shows how to compute the coordinates of T directly.

Proposition 2.2.4. *The coordinates of T are*

$$T^{i_1,\ldots,i_h}_{j_1,\ldots,j_k} = T\big(v^{i_1},\ldots,v^{i_h},v_{j_1},\ldots,v_{j_k}\big).$$

Proof. Apply both members of (2) to $\big(v^{i_1},\ldots,v^{i_h},v_{j_1},\ldots,v_{j_k}\big)$. $\square$

Example 2.2.5. The coordinates of the Euclidean scalar product g on $\mathbb{R}^n$ with respect to an orthonormal basis are $g_{ij}=\delta_{ij}$.

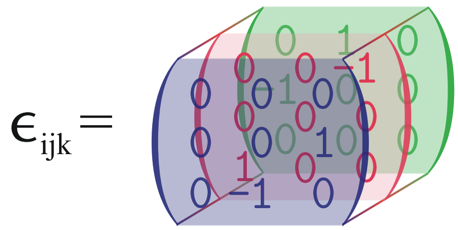

Figure 2.1. The coordinates of the cross product tensor with respect to the canonical basis of $\mathbb{R}^3$ (or any positive orthonormal basis) form the *Levi-Civita symbol* ϵ_{ijk}.

Example 2.2.6. The coordinates of $\mathrm{id} \in \mathrm{Hom}(V, V) = V \otimes V^*$ with respect to *any* basis are $\mathrm{id}^i_j = \delta^i_j$. This is again the Kronecker delta, written as δ^i_j for convenience.

Exercise 2.2.7. The coordinates of the cross product tensor in $\mathbb{R}^3$ with respect to any positive orthonormal basis are

$$T^i_{jk} = \epsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) \text{ is } (1,2,3), (2,3,1), \text{ or } (3,1,2), \\ -1 & \text{if } (i,j,k) \text{ is } (3,2,1), (1,3,2), \text{ or } (2,1,3), \\ 0 & \text{if } i = j, \text{ or } j = k, \text{ or } k = i. \end{cases}$$

The three-dimensional array ϵ_{ijk} is called the *Levi-Civita symbol* and is shown in Figure 2.1.

Exercise 2.2.8. The determinant in $\mathbb{R}^3$ may be interpreted as a tensor of type $(0, 3)$. Show that its coordinates with respect to any positive orthonormal basis are also ϵ_{ijk}.

2.2.3. Coordinates manipulation. The coordinates and the Einstein convention are powerful tools that enable us to describe complicated tensor manipulations in a very concise way, and the reader should familiarise with them. We start by exhibiting some simple examples. We fix a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ for V and consider coordinates with respect to this basis. We write the coordinates of a generic vector v as v^i, that is we have

$$v = v^i v_i.$$

Note that v^i is a number while v_i is a vector. If $v \in V$ is a vector and $T \colon V \to V$ is an endomorphism, that is $T \in \mathcal{T}^1_1(V)$, we may write $w = T(v)$ directly in coordinates as follows:

$$w^j = T^j_i v^i$$

where v^i, w^j, T^j_i are the coordinates of v, w, T. The trace of T is simply

$$T^i_i.$$

If $v, w \in V$ are vectors and $g\colon V \times V \to \mathbb{R}$ is a bilinear form, that is $g \in \mathcal{T}_0^2(V)$, it has coordinates g_{ij} and we may write the scalar $g(v, w)$ as follows:

$$v^i g_{ij} w^j.$$

The expressions $w^j = T^j_i v^i$ and $v^i g_{ij} w^j$ are just the usual products matrix-times-vector(s) that describe endomorphisms and bilinear forms in coordinates: we are only rewriting them using the Einstein convention.

Let T be the tensor of type $(1, 2)$ that describes the cross product in $\mathbb{R}^3$. The equality $z = v \wedge w$ can be written in coordinates as

$$z^i = T^i_{jk} v^j w^k.$$

Note that in all the cases described so far the Einstein convention is applied to pairs of indices where one is a superscript and the other is a subscript. This is in fact a general phenomenon: all the notation is designed to get this in *any* possible situation, with the purpose of limiting considerably the possibilities of errors and the amount of information that one has to remember by heart.

Example 2.2.9. We prove the well-known equalities

$$(v \wedge w) \cdot z = v \cdot (w \wedge z) = \det(v\, w\, z)$$

using coordinates. The three members may be written as

$$v^j T^i_{jk} w^k g_{il} z^l, \qquad v^l g_{li} w^j T^i_{jk} z^k, \qquad \det_{ijk} v^i w^j z^k.$$

Now we take an orthonormal basis $\mathcal{B}$, so that $g_{ij} = \delta_{ij}$ and $T^i_{jk} = \epsilon_{ijk} = \det_{ijk}$. The three members simplify as

$$\epsilon_{ijk} v^j w^k z^i, \qquad \epsilon_{ijk} v^i w^j z^k, \qquad \epsilon_{ijk} v^i w^j z^k$$

and they represent the same number thanks to the symmetries of ϵ.

Remark 2.2.10. In accordance with the previous discussion, the coordinates of a vector $x \in \mathbb{R}^n$ with respect to the canonical basis should be indicated with superscripts $x^1, \ldots, x^n$, and we will try to stick to this convention as much as possible; at few points we will break this rule and use subscripts to avoid cumbersome formulas like $(x^1)^2/(x^2)^1$.

2.2.4. Change of basis. If $\mathcal{C} = \{w_1, \ldots, w_n\}$ is another basis of V then

$$w_j = A^i_j v_i$$

for some invertible $n \times n$ matrix A of coefficients. Here “invertible” means of course that there is a $n \times n$ matrix B, called the *inverse* of A, such that

$$A^i_k B^k_j = \delta^i_j = B^i_k A^k_j$$

where δ^i_j is the Kronecker delta.

Proposition 2.2.11. *The dual basis changes as follows:*

$$w^i = B^i_j v^j.$$

Proof. We check that the proposed w^i form the dual basis of w_i:

$$w^i(w_j) = (B^i_k v^k)(A^l_j v_l) = B^i_k A^l_j v^k(v_l) = B^i_k A^l_j \delta^k_l = B^i_k A^k_j = \delta^i_j.$$

It is a useful exercise to fully understand each of the previous equalities! In the fourth one we removed the Kronecker delta and set $k = l$. □

Let T be a tensor as in (2). We now want to determine the coordinates $\hat{T}^{i_1,\dots,i_h}_{j_1,\dots,j_k}$ of T in the new basis $\mathcal{C}$, in terms of the coordinates $T^{i_1,\dots,i_h}_{j_1,\dots,j_k}$ in the old basis $\mathcal{B}$ and of the matrices A and B.

Proposition 2.2.12. *We have*

$$\hat{T}^{i_1\dots i_h}_{j_1\dots j_k} = B^{i_1}_{l_1}\cdots B^{i_h}_{l_h} A^{m_1}_{j_1}\cdots A^{m_k}_{j_k} T^{l_1\dots l_h}_{m_1\dots m_k} \tag{3}$$

This complicated-looking equation may be memorised by noting that we need one A for every lower index of T, and one B for every upper index.

Proof. By Proposition 2.2.4 we have

$$\begin{aligned}
\hat{T}^{i_1,\dots,i_h}_{j_1,\dots,j_k} &= T\big(w^{i_1},\dots,w^{i_h},w_{j_1},\dots,w_{j_k}\big) \\
&= T\big(B^{i_1}_{l_1}v^{l_1},\dots,B^{i_h}_{l_h}v^{l_h},A^{m_1}_{j_1}v_{m_1},\dots,A^{m_k}_{j_k}v_{m_k}\big) \\
&= B^{i_1}_{l_1}\cdots B^{i_h}_{l_h} A^{m_1}_{j_1}\cdots A^{m_k}_{j_k} T\big(v^{l_1},\dots,v^{l_h},v_{m_1},\dots,v_{m_k}\big) \\
&= B^{i_1}_{l_1}\cdots B^{i_h}_{l_h} A^{m_1}_{j_1}\cdots A^{m_k}_{j_k} T^{l_1\dots l_h}_{m_1\dots m_k}.
\end{aligned}$$

The proof is complete. □

The reader should appreciate the generality of the formula (3): it describes in a single equality the coordinate changes of vectors, covectors, endomorphisms, bilinear forms, the cross product in $\mathbb{R}^3$, the determinant, and some more complicated tensors that we will encounter in this book. We write some of them:

$$\hat{v}^i = B^i_l v^l, \quad \hat{v}_j = A^m_j v_m, \quad \hat{T}^i_j = B^i_l A^m_j T^l_m, \quad \hat{g}_{ij} = A^m_i A^n_j g_{mn}.$$

The formula (3) contains many indices and may look complicated at a first glance, but in fact it only says that the lower indices $j_1,\dots,j_k$ change through the matrix A, while the upper indices $i_1,\dots,i_h$ change via the inverse matrix B. For that reason, the lower and upper indices are also called respectively *covariant* and *contravariant*.

Remark 2.2.13. In some physics and engineering text books, the formula (3) is used as a *definition* of tensor: a tensor is simply a multi-dimensional array, that changes as prescribed by that formula if one modifies the basis of the vector space.

We now introduce some operations with tensors.

2.2.5. Tensor product. It follows from the definitions that

$$\mathcal{T}_h^k(V) \otimes \mathcal{T}_m^n(V) = \mathcal{T}_{h+m}^{k+n}(V).$$

In particular, given two tensors $S \in \mathcal{T}_h^k(V)$ and $T \in \mathcal{T}_m^n(V)$, their *product* $S \otimes T$ is an element of $\mathcal{T}_{h+m}^{k+n}(V)$. In coordinates with respect to some basis $\mathcal{B}$, it may be written as

$$(S \otimes T)^{i_1 \cdots i_h i_{h+1} \cdots i_{h+m}}_{j_1 \cdots j_k j_{k+1} \cdots j_{k+n}} = S^{i_1 \cdots i_h}_{j_1 \cdots j_k} T^{i_{h+1} \cdots i_{h+m}}_{j_{k+1} \cdots j_{k+n}}.$$

2.2.6. The tensor algebra. The *tensor algebra* of V is

$$\mathcal{T}(V) = \bigoplus_{h,k \geq 0} \mathcal{T}_h^k(V).$$

The product $\otimes$ is defined on every pair of tensors, and it extends distributively on the whole of $\mathcal{T}(V)$. With this operation $\mathcal{T}(V)$ is an associative algebra and an infinite-dimensional vector space (if V is not trivial). Recall that

$$\mathcal{T}_0^0(V) = \mathbb{R}, \qquad \mathcal{T}_1^0(V) = V, \qquad \mathcal{T}_0^1(V) = V^*.$$

Exercise 2.2.14. If $\dim V \geq 2$ the algebra is not commutative: if $v, w \in V$ are independent vectors, then $v \otimes w \neq w \otimes v$.

We denote for simplicity

$$\mathcal{T}_h(V) = \mathcal{T}_h^0(V), \qquad \mathcal{T}^k(V) = \mathcal{T}_0^k(V).$$

The vector spaces

$$\mathcal{T}_*(V) = \bigoplus_{h \geq 0} \mathcal{T}_h(V), \qquad \mathcal{T}^*(V) = \bigoplus_{k \geq 0} \mathcal{T}^k(V)$$

are both subalgebras of $\mathcal{T}(V)$ and are sometimes called the *contravariant* and *covariant tensor algebras*, respectively.

Exercise 2.2.15. The algebras $\mathcal{T}_*(\mathbb{R})$ and $\mathbb{R}[x]$ are isomorphic.

Remark 2.2.16. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V. The elements $v_1, \ldots, v_n \in \mathcal{T}_1(V)$ generate $\mathcal{T}_*(V)$ as a *free algebra*. This means that every element of $\mathcal{T}_*(V)$ may be written as a polynomial in the variables $v_1, \ldots, v_n$ in a unique way up to permuting its addenda. Note that $\otimes$ is not commutative, hence the ordering in each monomial is important. As an example:

$$3 + v_1 - 7v_2 + v_1 \otimes v_2 - 3v_2 \otimes v_1.$$

2.2.7. Contractions. We now introduce a general important operation on tensors called *contraction* that generalises the trace of endomorphisms.

The trace is an operation that picks as an input an endomorphism, that is a $(1,1)$-tensor, and produces as an output a number, that is a $(0,0)$-tensor. More generally, a contraction is an operation that transforms a (h,k)-tensor into a $(h-1,k-1)$-tensor, and is defined for all $h,k\geq 1$. It depends on the choice of two integers $1\leq a\leq h$ and $1\leq b\leq k$ and results in a linear map

$$C\colon \mathcal{T}_h^k(V)\longrightarrow \mathcal{T}_{h-1}^{k-1}(V).$$

The contraction is defined as follows. Recall that

$$\mathcal{T}_h^k(V)=\underbrace{V\otimes\cdots\otimes V}_{h}\otimes\underbrace{V^*\otimes\cdots\otimes V^*}_{k}.$$

The indices a and b indicate which factors V and V^* need to be "contracted". After a canonical isomorphism we may put these factors at the end and write

$$\mathcal{T}_h^k(V)=\underbrace{V\otimes\cdots\otimes V}_{h-1}\otimes\underbrace{V^*\otimes\cdots\otimes V^*}_{k-1}\otimes V\otimes V^*=\mathcal{T}_{h-1}^{k-1}(V)\otimes V\otimes V^*.$$

The *contraction* is the linear map

$$C\colon \mathcal{T}_{h-1}^{k-1}(V)\otimes V\otimes V^*\longrightarrow \mathcal{T}_{h-1}^{k-1}(V)$$

determined by the condition

$$C(w\otimes v\otimes v^*)=v^*(v)w.$$

Recall that C is well-defined because $(w,v,v^*)\mapsto v^*(v)w$ is multilinear and hence the universal property applies.

Example 2.2.17. The contraction of a pure tensor is

$$C(v_1\otimes\cdots\otimes v_h\otimes v^1\otimes\cdots\otimes v^k)=$$
$$v^b(v_a)v_1\otimes\cdots\otimes\widehat{v_a}\otimes\cdots\otimes v_h\otimes v^1\otimes\cdots\otimes\widehat{v^b}\otimes\cdots\otimes v^k$$

where $\widehat{w}$ indicates that the factor w is omitted.

2.2.8. In coordinates. The definition of a contraction may look abstruse, but we now see that everything is pretty simple in coordinates. Let $\mathcal{B}=\{v_1,\ldots,v_n\}$ be a basis for V.

Proposition 2.2.18. *If T has coordinates $T^{i_1,\ldots,i_h}_{j_1,\ldots,j_k}$, then $C(T)$ has*

$$C(T)^{i_1,\ldots,i_{h-1}}_{j_1,\ldots,j_{k-1}}=T^{i_1,\ldots,l,\ldots,i_{h-1}}_{j_1,\ldots,l,\ldots,j_{k-1}}$$

where l is inserted at the positions a above and b below.

Proof. We write the coordinates of T as $T^{i_1,\dots,i,\dots,i_{h-1}}_{j_1,\dots,j,\dots,j_{k-1}}$ for convenience, where i and j occupy the places a and b. We have

$$\begin{aligned}
C(T) &= C\big(T^{i_1,\dots,i,\dots,i_{h-1}}_{j_1,\dots,j,\dots,j_{k-1}} v_{i_1}\otimes\cdots\otimes v_i\otimes\cdots\otimes v_{i_{h-1}}\otimes v^{j_1}\otimes\cdots\otimes v^j\otimes\cdots\otimes v^{j_{k-1}}\big)\\
&= T^{i_1,\dots,i,\dots,i_{h-1}}_{j_1,\dots,j,\dots,j_{k-1}}\delta^j_i v_{i_1}\otimes\cdots\otimes v_{i_{h-1}}\otimes v^{j_1}\otimes\cdots\otimes v^{j_{k-1}}\\
&= T^{i_1,\dots,l,\dots,i_{h-1}}_{j_1,\dots,l,\dots j_{k-1}} v_{i_1}\otimes\cdots\otimes v_{i_{h-1}}\otimes v^{j_1}\otimes\cdots\otimes v^{j_{k-1}}.
\end{aligned}$$

The proof is complete. $\square$

This shows in particular that, as promised, the contraction of an endomorphism whose coordinates are T^i_j is indeed its trace T^i_i.

Contractions are handled very easily in coordinates. As an example, a tensor T of type $(1,2)$ has coordinates T^i_{jk} and can be contracted in two ways, producing two (typically distinct) covectors v and v' with coordinates

$$v_k = T^i_{ik}, \quad v'_j = T^i_{ji}.$$

It is important to remember that the coordinates depend on the choice of a basis $\mathcal{B}$, but the covectors v and v' obtained by contracting T do *not* depend on $\mathcal{B}$. Likewise, a tensor of type T^{ij}_{kl} has four types of contractions, producing four (possibly distinct) tensors of type (1,1), that is endomorphisms.

It is convenient to manipulate a tensor using its coordinates as we just did: remember however that we must always contract a covariant index together with a contravariant one! The "contraction" of two covariant (or contravariant) indices makes no sense because it is not basis-independent. This should not be surprising: the trace T^i_i of an endomorphism is basis-independent, but the trace g_{ii} of a bilinear form is notoriously not. Said with other words: there is a canonical homomorphism $V\otimes V^*\to\mathbb{R}$, but there is no canonical homomorphism $V\otimes V\to\mathbb{R}$.

Exercise 2.2.19. The tensor T that expresses the cross product in $\mathbb{R}^3$ has two contractions. Prove that they both give rise to the null covector.

Hint. This can be done by calculation, or abstractly: since T is invariant under orientation-preserving isometries, its contractions also are. $\square$

Example 2.2.20. Let $T, \det, g$ be the tensors in $\mathbb{R}^3$ that represent the cross product, the determinant, and the Euclidean scalar product. They are of type $(1,2)$, $(0,3)$, and $(0,2)$ respectively. The tensor $T\otimes g$ is of type $(1,4)$ and may be written in coordinates as $T^k_{ij}g_{lm}$. It has four contractions $C(T\otimes g)$, that are all of type $(0,3)$. These are

$$T^k_{kj}g_{lm}, \quad T^k_{ik}g_{lm}, \quad T^k_{ij}g_{km}, \quad T^k_{ij}g_{lk}.$$

The first two are null by the previous exercise. The last two, expressed on a orthonormal basis, become ϵ_{ijm} and ϵ_{ijl}. Therefore for these two contractions we get $C(T\otimes g) = \det$.

Every time we sum over a pair of covariant and contravariant indices, we are doing a contraction. So for instance each of the operations

$$w^j = T^j_i v^i, \qquad v^i g_{ij} w^j$$

described in Section 2.2.3 may be interpreted as two-steps operations, where we first multiply some tensors and then we contract the result. Contractions and tensor products are everywhere.

2.3. Scalar products

We now study vector spaces V equipped with a *scalar product* g. We investigate in particular the effects of g on the tensor algebra $\mathcal{T}(V)$. We start by recalling some basic facts on scalar products.

2.3.1. Definition. A *scalar product* on V is a symmetric bilinear form g that is *not degenerate*, that is

$$g(v, w) = 0 \ \forall v \in V \iff w = 0.$$

Recall that the scalar product is

- *positive definite* if $g(v, v) > 0 \ \forall v \neq 0$,
- *negative definite* if $g(v, v) < 0 \ \forall v \neq 0$,
- *indefinite* in the other cases.

Every scalar product g has a *signature* (p, m) where p (respectively, m) is the maximum dimension of a subspace $W \subset V$ such that the restriction $g|_W$ is positive definite (respectively, negative definite). We have $p+m = n = \dim V$. The scalar product is positive definite (respectively, negative definite) $\iff$ its signature is $(n, 0)$ (respectively, $(0, n)$).

A scalar product g is a tensor of type $(0, 2)$ and its coordinates with respect to some basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ are written as g_{ij}. The basis $\mathcal{B}$ is *orthonormal* if $g_{ij} = \pm\delta_{ij}$ for all i, j. In particular $g_{ii} = \pm 1$, and the sign $+1$ and -1 occur p and m times as i varies. Every scalar product has an orthonormal basis.

We are mostly interested in positive definite scalar products, but indefinite scalar product also arise in some interesting contexts – notably in Einstein's general relativity.

2.3.2. Isometries. Let V and W be equipped with some scalar products g and h. A linear map $T \colon V \to W$ is an *isometry* if $g(u, v) = h\big(T(u), T(v)\big)$ for all $u, v \in V$. This condition can be expressed in coordinates as

$$u^i g_{ij} v^j = u^i T^k_i h_{kl} T^l_j v^j$$

and since it must be satisfied for all u, v we get

$$g_{ij} = T^k_i h_{kl} T^l_j.$$

The isometries from V to itself form a group that we denote by $\mathrm{O}(V)$. After fixing a basis, the group $\mathrm{O}(V)$ can be represented as the subgroup of

$\mathrm{GL}(n, \mathbb{R})$ formed by the matrices A such that ${}^{t}AgA = g$. In particular Binet's formula yields $\det A = \pm 1$. Therefore every isometry $f \in \mathrm{O}(V)$ has $\det f = \pm 1$ and the positive isometries (that is, those with $\det = 1$) form an index-two normal subgroup $\mathrm{SO}(V) < \mathrm{O}(V)$.

When g is positive-definite and the basis is orthonormal we get the usual orthogonal group $\mathrm{O}(n) < \mathrm{GL}(n, \mathbb{R})$ formed by all the matrices A such that ${}^{t}AA = I$. More generally, if g has signature (p, m) we can find an orthonormal basis $v_1, \ldots, v_n$ where $g_{ii} = -1$ for $i = 1, \ldots, m$ and $g_{ii} = 1$ for $i = m + 1, \ldots, n$ and in this basis $\mathrm{O}(V)$ can be represented as the subgroup

$$\mathrm{O}(p, m) = \left\{A \in \mathrm{GL}(n, \mathbb{R}) \mid {}^{t}AJA = J\right\}, \quad J = \begin{pmatrix} -I_m & 0 \\ 0 & I_p \end{pmatrix}.$$

2.3.3. The identification of V and V^*. Let V be equipped with a scalar product g. Our aim is now to show that g enriches the tensor algebra $\mathcal{T}(V)$ with some new interesting structures.

We first discover that g induces an isomorphism

$$V \longrightarrow V^*$$

that sends $v \in V$ to the functional $v^* \in V^*$ defined by $v^*(w) = g(v, w)$. (This is an isomorphism because g is non-degenerate!) This is an important point: as we know, the spaces V and V^* are not canonically identified, but we can identify them once we have fixed a scalar product g.

Exercise 2.3.1. The isomorphism $V \to V^*$ sends a vector with coordinates v^i to the covector with coordinates

$$v_j = g_{ij} v^i.$$

The scalar product g induces a scalar product on V^*, that we lazily still name g, as follows:

$$g(v^*, w^*) = g(v, w)$$

where $v^*, w^* \in V^*$ are the images of $v, w \in V$ along the isomorphism $V \to V^*$ defined above. The scalar product g on V^* is a tensor of type $(2, 0)$ and its coordinates are denoted by g^{ij}.

Proposition 2.3.2. *The matrix g^{ij} is the inverse of g_{ij}.*

Proof. Note that g_{ij} is invertible because g is non-degenerate. The equality defining g^{ij} may be rewritten in coordinates as

$$v^i g_{ik} g^{kl} g_{lj} w^j = v_k g^{kl} w_l = v^i g_{ij} w^j.$$

Since this holds for every $v, w \in V$ we get

$$g_{ik} g^{kl} g_{lj} = g_{ij}.$$

Read as a matrices multiplication, this is $GHG = G$ that implies $GH = HG = I$ because G is invertible and hence $H = G^{-1}$. The proof is complete. $\square$

Note that the proposition holds for every choice of a basis $\mathcal{B}$.

2.3.4. Raising and lowering indices. We may use the scalar product g on V to "raise" and "lower" the indices of any tensor at our pleasure. That is, the isomorphism $V \to V^*$ induces an isomorphism

$$\mathcal{T}_h^k(V) \longrightarrow \mathcal{T}_{h+k}(V)$$

for all $h, k \geq 0$. In coordinates, the isomorphism sends a tensor $T^{i_1,\ldots,i_h}_{j_1,\ldots,j_k}$ to

$$U^{i_1,\ldots,i_h,j_1,\ldots,j_k} = T^{i_1,\ldots,i_h}_{l_1,\ldots,l_k} g^{l_1 j_1} \cdots g^{l_k j_k}.$$

We can use g^{ij} to raise the indices of a tensor, and in the opposite direction we can use g_{ij} to lower them. This operation may be encoded efficiently and unambiguously by assigning different indices to distinct columns in the notation. So for instance we start with a tensor like

$$T_i{}^j{}_{kl}$$

and then we may raise or lower some indices to produce a new tensor that we may lazily indicate with the same letter; for instance we can move the indices i and j and get a new tensor

$$T^i{}_{jkl}.$$

If $g_{ij} = \delta_{ij}$, then $g^{ij} = \delta^{ij}$ and the coordinates of the two different tensors are just the same, that is $T_i{}^j{}_{kl} = T^i{}_{jkl}$ for every i, j, k, l. In general we have

$$T^i{}_{jkl} = T_a{}^b{}_{kl} g^{ai} g_{bj}.$$

2.3.5. Scalar product on the tensor spaces. A scalar product g on V induces a scalar product on each vector space $\mathcal{T}_h^k(V)$, still boringly denoted by g. This is done as follows: if $S, T \in \mathcal{T}_h^k(V)$, then $g(S, T)$ is the scalar

$$T^{i_1,\ldots,i_h}_{j_1,\ldots,j_k} g_{i_1 l_1} \cdots g_{i_h l_h} g^{j_1 m_1} \cdots g^{j_k m_k} S^{l_1,\ldots,l_h}_{m_1,\ldots,m_k}.$$

This number is clearly basis-independent because it is obtained by multiple contractions of a product of tensors.

Exercise 2.3.3. If $\mathcal{B} = \{v_1, \ldots, v_n\}$ is an orthonormal basis of V, then

$$\{v_{i_1} \otimes \cdots \otimes v_{i_h} \otimes v^{j_1} \otimes \cdots \otimes v^{j_k}\}$$

is an orthonormal basis of $\mathcal{T}_h^k(V)$. If g is positive-definite on V then it is so also on $\mathcal{T}_h^k(V)$.

More generally, the following holds. We denote the scalar product as $\langle , \rangle$.

Exercise 2.3.4. For any choice of $v_i, w_j \in V$ and $v^k, w^l \in V^*$ we have

$$\langle v_1 \otimes \cdots v_h \otimes v^1 \otimes \cdots \otimes v^k, w_1 \otimes \cdots w_h \otimes w^1 \otimes \cdots \otimes w^k \rangle = \prod_{i=1}^h \langle v_i, w_i \rangle \prod_{j=1}^k \langle v^j, w^j \rangle.$$

2.4. The symmetric and exterior algebras

Symmetric and antisymmetric matrices play an important role in linear algebra: both concepts can be generalised to tensors.

2.4.1. Symmetric and antisymmetric tensors. We now introduce two special types of covariant tensors.

Definition 2.4.1. A tensor $T \in \mathcal{T}^k(V)$ is *symmetric* if

$$T(u_1, \ldots, u_k) = T\big(u_{\sigma(1)}, \ldots, u_{\sigma(k)}\big) \tag{4}$$

for every vectors $u_1, \ldots, u_k \in V$ and every permutation $\sigma \in S_k$. On the other hand T is *antisymmetric* if

$$T(u_1, \ldots, u_k) = \operatorname{sgn}(\sigma) T\big(u_{\sigma(1)}, \ldots, u_{\sigma(k)}\big)$$

for every vectors $u_1, \ldots, u_k \in V$ and every permutation $\sigma \in S_k$. Here $\operatorname{sgn}(\sigma) = \pm 1$ is the sign of the permutation σ.

Both conditions are very easily expressed in coordinates. As usual we fix any basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ on V and consider the coordinates of T with respect to $\mathcal{B}$.

Proposition 2.4.2. *A tensor $T \in \mathcal{T}^k(V)$ is*

- *symmetric* $\iff T_{i_1,\ldots,i_k} = T_{i_{\sigma(1)},\ldots,i_{\sigma(k)}} \quad \forall i_1, \ldots, i_k,\ \forall \sigma;$
- *antisymmetric* $\iff T_{i_1,\ldots,i_k} = \operatorname{sgn}(\sigma) T_{i_{\sigma(1)},\ldots,i_{\sigma(k)}} \ \forall i_1, \ldots, i_k, \forall \sigma.$

Proof. We prove the first sentence, the second is analogous. Recall that

$$T_{i_1,\ldots,i_k} = T\big(v_{i_1}, \ldots, v_{i_k}\big).$$

Therefore we must prove that (4) holds for all vectors $\iff$ it holds for the vectors in the basis $\mathcal{B}$. This is left as an exercise. □

For instance a tensor T_{ij} is symmetric if $T_{ij} = T_{ji}$ and antisymmetric if $T_{ij} = -T_{ji}$, for all $1 \leq i, j \leq n$.

Example 2.4.3. Every scalar product on V is a symmetric tensor $g \in \mathcal{T}^2(V)$. The determinant is an antisymmetric tensor $\det \in \mathcal{T}^n(\mathbb{R}^n)$.

Remark 2.4.4. If T is antisymmetric and the indices $i_1, \ldots, i_k$ are not all distinct, then $T_{i_1,\ldots,i_k} = 0$.

2.4.2. Symmetrisation and antisymmetrisation of tensors. If a tensor T is not (anti-)symmetric, we can transform it by brute force into an (anti-)symmetric one.

Let $T \in \mathcal{T}^k(V)$ be a covariant tensor. The *symmetrisation* of T is the tensor $S(T) \in \mathcal{T}^k(V)$ defined by averaging T on permutations as follows:

$$S(T)(v_1, \ldots, v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} T\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big).$$

Analogously, the *antisymmetrisation* of T is the tensor

$$A(T)(v_1, \ldots, v_k) = \frac{1}{k!}\sum_{\sigma\in S_k} \mathrm{sgn}(\sigma) T\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big).$$

Exercise 2.4.5. The tensors $S(T)$ and $A(T)$ are indeed symmetric and antisymmetric, respectively. We have $S(T) = T \iff T$ is symmetric and $A(T) = T \iff T$ is antisymmetric.

In coordinates with respect to some basis we have

$$S(T)_{i_1,\ldots,i_k} = \frac{1}{k!}\sum_{\sigma\in S_k} T_{i_{\sigma(1)},\ldots,i_{\sigma(k)}},$$

$$A(T)_{i_1,\ldots,i_k} = \frac{1}{k!}\sum_{\sigma\in S_k} \mathrm{sgn}(\sigma) T_{i_{\sigma(1)},\ldots,i_{\sigma(k)}}.$$

The members on the right can be written more concisely as

$$T_{(i_1,\ldots,i_k)}, \qquad T_{[i_1,\ldots,i_k]}.$$

The round or square brackets indicate that we symmetrise or antisymmetrise by summing along all permutations on the indices (and dividing by $k!$).

2.4.3. The symmetric and antisymmetric algebras. We now introduce two more algebras associated to V. For every $k \geq 0$ we denote by

$$S^k(V), \qquad \Lambda^k(V)$$

the vector subspace of $\mathcal{T}^k(V)$ consisting of all the symmetric or antisymmetric tensors, respectively. We now define

$$S^*(V) = \bigoplus_{k\geq 0} S^k(V), \qquad \Lambda^*(V) = \bigoplus_{k\geq 0} \Lambda^k(V).$$

These are both vector subspaces of the covariant tensor algebra $\mathcal{T}^*(V)$. These are *not* subalgebras of $\mathcal{T}^*(V)$, because they are not closed under $\otimes$. Note that

$$S^1(V) = \Lambda^1(V) = \mathcal{T}^1(V) = V^*$$

but $S^2(V)$ and $\Lambda^2(V)$ are strictly smaller than $\mathcal{T}^2(V)$ if $\dim V \geq 2$, because of the following:

Exercise 2.4.6. If $v^*, w^* \in V^*$ are independent, then $v^* \otimes w^*$ is neither symmetric nor antisymmetric. Moreover

$$S(v^*\otimes w^*) = \frac{1}{2}(v^*\otimes w^* + w^*\otimes v^*), \quad A(v^*\otimes w^*) = \frac{1}{2}(v^*\otimes w^* - w^*\otimes v^*).$$

The spaces $S^*(V)$ and $\Lambda^*(V)$ are actually algebras, but with some products different from $\otimes$, that we now define. The *symmetrised product* of some covariant tensors $T^1 \in \mathcal{T}^{k_1}(V), \ldots, T^m \in \mathcal{T}^{k_m}(V)$ is

$$T^1 \odot \cdots \odot T^m = \frac{(k_1 + \ldots + k_m)!}{k_1! \cdots k_m!} S(T^1 \otimes \cdots \otimes T^m)$$

while their *antisymmetrised product*

$$T^1 \wedge \cdots \wedge T^m = \frac{(k_1 + \cdots + k_m)!}{k_1! \cdots k_m!} A(T^1 \otimes \cdots \otimes T^m).$$

For instance if $v^*, w^* \in V^*$ then

$$v^* \odot w^* = v^* \otimes w^* + w^* \otimes v^*, \qquad v^* \wedge w^* = v^* \otimes w^* - w^* \otimes v^*.$$

Note that

$$v^* \odot w^* = w^* \odot v^*, \qquad v^* \wedge w^* = -w^* \wedge v^*.$$

More generally, if $v^1, \ldots, v^m \in V^*$ then

$$\begin{aligned} v^1 \odot \cdots \odot v^m &= \sum_{\sigma \in S_m} v^{\sigma(1)} \otimes \cdots \otimes v^{\sigma(m)}, \\ v^1 \wedge \cdots \wedge v^m &= \sum_{\sigma \in S_m} \operatorname{sgn}(\sigma) v^{\sigma(1)} \otimes \cdots \otimes v^{\sigma(m)}. \end{aligned}$$

Using coordinates with respect to some basis $\mathcal{B}$ of V and the brackets $()$, $[]$ to denote symmetrisation and antisymmetrisation, we may write

$$\begin{aligned} (T \odot U)_{i_1,\ldots,i_{p+q}} &= \frac{(p+q)!}{p!q!} T_{(i_1,\ldots,i_p} U_{i_{p+1},\ldots,i_{p+q})}, \\ (T \wedge U)_{i_1,\ldots,i_{p+q}} &= \frac{(p+q)!}{p!q!} T_{[i_1,\ldots,i_p} U_{i_{p+1},\ldots,i_{p+q}]}. \end{aligned}$$

Proposition 2.4.7. *The vector spaces $S^*(V)$ and $\Lambda^*(V)$ form two associative algebras with the products $\odot$ and $\wedge$ respectively.*

Proof. Everything is immediate except associativity. We prove it for Λ, the other is analogous. Pick $S \in \Lambda^p, T \in \Lambda^q$, and $U \in \Lambda^r$. In coordinates

$$\begin{aligned} \big((S \wedge T) \wedge U\big)_{i_1,\ldots,i_{p+q+r}} &= \frac{1}{(p+q)!r!} (S \wedge T)_{[i_1,\ldots,i_{p+q}} U_{i_{p+q+1},\ldots,i_{p+q+r}]} \\ &= \frac{1}{(p+q)!p!q!r!} S_{[[i_1,\ldots,i_p} T_{i_{p+1},\ldots,i_{p+q}]} U_{i_{p+q+1},\ldots,i_{p+q+r}]} \\ &= \frac{1}{p!q!r!} S_{[i_1,\ldots,i_p} T_{i_{p+1},\ldots,i_{p+q}} U_{i_{p+q+1},\ldots,i_{p+q+r}]} \\ &= (S \wedge T \wedge U)_{i_1,\ldots,i_{p+q+r}}. \end{aligned}$$

The third equality follows from the fact that the same permutation in the symmetric group S_{p+q+r} is obtained $(p+q)!$ times. Analogously we can prove that $S \wedge (T \wedge U) = S \wedge T \wedge U$. The proof is complete. $\square$

The two algebras $S^*(V)$ and $\Lambda^*(V)$ are called the *covariant symmetric algebra* and the *covariant exterior algebra*. The products $\odot$ and $\wedge$ are called the *symmetric* and *exterior product*.

2.4.4. Dimensions. We now construct some standard basis for $S^k(V)$ and $\Lambda^k(V)$ and calculate their dimensions. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for V and $\mathcal{B}^* = \{v^1, \ldots, v^n\}$ the dual basis of V^*.

Proposition 2.4.8. *A basis for $S^k(V)$ is*

$$\left\{v^{i_1} \odot \cdots \odot v^{i_k}\right\}$$

where $1 \leq i_1 \leq \ldots \leq i_k \leq n$ vary. A basis for $\Lambda^k(V)$ is

$$\left\{v^{i_1} \wedge \cdots \wedge v^{i_k}\right\}$$

where $1 \leq i_1 < \ldots < i_k \leq n$ vary.

Proof. This is a consequence of Propositions 2.4.2 and Remark 2.4.4. □

Example 2.4.9. Let e_1, e_2 be the canonical basis for $\mathbb{R}^2$ and e^1, e^2 be the dual basis. The following is a basis for $S^2(\mathbb{R}^2)$:

$$e^1 \odot e^1, \quad e^1 \odot e^2, \quad e^2 \odot e^2.$$

The following is a basis for $\Lambda^2(\mathbb{R}^3)$:

$$e^1 \wedge e^2, \quad e^1 \wedge e^3, \quad e^2 \wedge e^3.$$

Corollary 2.4.10. *We have*

$$\dim S^k(V) = \binom{n+k-1}{k},$$
$$\dim \Lambda^k(V) = \binom{n}{k}.$$

Corollary 2.4.11. *The algebra $S^*(V)$ is commutative, while $\Lambda^*(V)$ is* anticommutative, *that is*

$$T \wedge U = (-1)^{pq} U \wedge T$$

for every $T \in \Lambda^p(V)$ and $U \in \Lambda^q(V)$.

Proof. We prove anticommutativity. It suffices to prove this when T, U are basis elements, that is we must show that

$$(v^{i_1} \wedge \cdots \wedge v^{i_p}) \wedge (v^{j_1} \wedge \cdots \wedge v^{j_q}) = (-1)^{pq} (v^{j_1} \wedge \cdots \wedge v^{j_q}) \wedge (v^{i_1} \wedge \cdots \wedge v^{i_p}).$$

This equality follows from applying pq times the simple equality

$$v^* \wedge w^* = -w^* \wedge v^*.$$

The proof is complete. □

Corollary 2.4.12. *If $T \in \Lambda^k(V)$ with odd k then $T \wedge T = 0$.*

Corollary 2.4.13. *We have* $\dim S^*(V) = \infty$ *and* $\dim \Lambda^*(V) = 2^n$.

Exercise 2.4.14. The algebras $S^*(V)$ and $\mathbb{R}[x_1, \ldots, x_n]$ are isomorphic.

In the rest of this section we will focus mostly on the exterior algebra $\Lambda^*(V)$, that will be a fundamental tool in this book.

2.4.5. The determinant line. One of the most important aspect of the theory, that will have important applications in the next chapters, is the following – apparently innocuous – fact:

$$\dim \Lambda^n(V) = 1.$$

The space $\Lambda^n(V)$ is called the *determinant line*. If $v^1, \ldots, v^n$ is a basis of V^*, then a generator for $\Lambda^n(V)$ is the tensor

$$v^1 \wedge \cdots \wedge v^n.$$

In fact, we already know that there is only one alternating n-linear form in V up to rescaling – this is exactly where the determinant comes from. When $V = \mathbb{R}^n$, we get

$$\det = e^1 \wedge \cdots \wedge e^n$$

where $e^1, \ldots, e^n$ is the canonical basis of $(\mathbb{R}^n)^* = \mathbb{R}^n$. Note however that $\Lambda^n(V)$ has no canonical generator unless we make some choice, like for instance a basis of V.

Let now $v^1, \ldots, v^n$ and $w^1, \ldots, w^n$ be two bases of V^*, and let A the change of basis matrix, so that $v^i = A^i_j w^j$.

Proposition 2.4.15. *The following equality holds:*

$$v^1 \wedge \cdots \wedge v^n = \det A \cdot w^1 \wedge \cdots \wedge w^n.$$

Proof. We have

$$\begin{aligned} v^1 \wedge \cdots \wedge v^n &= A^1_{j_1} \cdots A^n_{j_n} w^{j_1} \wedge \cdots \wedge w^{j_n} \\ &= \sum_{\sigma \in S_n} A^1_{\sigma(1)} \cdots A^n_{\sigma(n)} w^{\sigma(1)} \wedge \cdots \wedge w^{\sigma(n)} \\ &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) A^1_{\sigma(1)} \cdots A^n_{\sigma(n)} w^1 \wedge \cdots \wedge w^n \\ &= \det A \cdot w^1 \wedge \cdots \wedge w^n. \end{aligned}$$

The proof is complete. □

We have discovered here another important fact: the equality looks like the formula in the change of variables in multiple integrals, see Section 1.3.8. This will allow us to connect alternating tensors with integration and volume.

2.4.6. Contractions. Let $v \in V$ be a fixed vector. By contracting with v we may define a linear map

$$\iota_v \colon \Lambda^k(V) \longrightarrow \Lambda^{k-1}(V).$$

The linear map sends $T \in \Lambda^k(V)$ to the antisymmetric tensor $\iota_v(T)$ that acts on vectors as follows:

$$\iota_v(T)(v_1, \ldots, v_{k-1}) = T\big(v, v_1, \ldots, v_{k-1}\big).$$

It is immediate to check that $\iota_v(T)$ is indeed antisymmetric.

Exercise 2.4.16. The following hold:

$$\iota_v \circ \iota_v = 0, \qquad \iota_v \circ \iota_w = -\iota_w \circ \iota_v.$$

2.4.7. Totally decomposable antisymmetric tensors. An antisymmetric tensor $T \in \Lambda^k(V)$ is *totally decomposable* if it may be written as

$$T = w^1 \wedge \cdots \wedge w^k$$

for some covectors $w^1, \ldots, w^k \in V^*$. This notion is similar to that of a pure tensor, only with the product $\wedge$ instead of $\otimes$.

Exercise 2.4.17. The element $T = w^1 \wedge \cdots \wedge w^k$ is non-zero $\iff$ the covectors $w^1, \ldots, w^k$ are linearly independent.

As with pure tensors, not all the antisymmetric tensors are totally decomposable:

Exercise 2.4.18. If $v_1, v_2, v_3, v_4 \in V^*$ are linearly independent then

$$v_1 \wedge v_2 + v_3 \wedge v_4$$

is not totally decomposable.

Hint. If w is totally decomposable, then $w \wedge w = 0$. □

2.4.8. Contravariant versions. We have established the theory of symmetric and antisymmetric covariant tensors, but actually everything we said also holds *verbatim* for the *contravariant* tensors: we can therefore denote by

$$S_k(V), \qquad \Lambda_k(V)$$

the subspaces of $\mathcal{T}_k(V)$ consisting of all the symmetric or antisymmetric tensors, and define

$$S_*(V) = \bigoplus_{k \geq 0} S_k(V), \qquad \Lambda_*(V) = \bigoplus_{k \geq 0} \Lambda_k(V).$$

These form two algebras, called the *contravariant symmetric algebra* and *contravariant exterior algebra*.

2.4.9. Linear maps. Every linear map $L \colon V \to W$ between vector spaces induces some algebra homomorphisms

$$\begin{aligned} L_* &\colon \mathcal{T}_*(V) \longrightarrow \mathcal{T}_*(W), & L^* &\colon \mathcal{T}^*(W) \longrightarrow \mathcal{T}^*(V), \\ L_* &\colon S_*(V) \longrightarrow S_*(W), & L^* &\colon S^*(W) \longrightarrow S^*(V), \\ L_* &\colon \Lambda_*(V) \longrightarrow \Lambda_*(W), & L^* &\colon \Lambda^*(W) \longrightarrow \Lambda^*(V). \end{aligned}$$

The passing from L to L_* or L^* is functorial, that is

$$(L' \circ L)_* = L'_* \circ L_*, \qquad \mathrm{id}_* = \mathrm{id},$$
$$(L' \circ L)^* = L^* \circ (L')^*, \qquad \mathrm{id}^* = \mathrm{id}.$$

From this we deduce that if L is an isomorphism then L_* is an isomorphism. More than that:

- if L is injective then L_* is injective and L^* is surjective,
- if L is surjective then L_* is surjective and L^* injective.

This holds because if L is injective (surjective) there is a linear map $L' \colon W \to V$ such that $L' \circ L = \mathrm{id}_V$ ($L \circ L' = \mathrm{id}_W$), as one proves with standard linear algebra.

Remark 2.4.19. The terms *covariance* and its opposite *contravariance* are used for similar objects in two quite different contexts, and this is a permanent source of confusion. In general, a mathematical entity is "covariant" if it changes "in the same way" as some other preferred entity when some modification is made. But which modifications are we considering here?

Physicists are interested in changes of frame, that is of basis, and they note that if we change a basis with a matrix A the coordinates of a vector change with $B = A^{-1}$, that is *contravariantly*. On the other hand, mathematicians are mostly interested in functoriality, and note that a map $L \colon V \to W$ induces maps $L_* \colon \mathcal{T}_*(V) \to \mathcal{T}_*(W)$ and $L^* \colon \mathcal{T}^*(W) \to \mathcal{T}^*(V)$ on tensors, and they would tend to call *contravariant* the second types of tensors because arrows are reversed. Unfortunately, the physicist and mathematicians conventions do not match.

We have chosen here the physicist convention, but we try to use the terms *covariant* and *contravariant* as little as possible. The reader can ignore all these matters – in fact, these issues start to annoy you only when you decide to write a textbook, and you are forced to choose a notation that is both reasonable and consistent.

2.4.10. Non-degenerate bilinear pairing. Let V have dimension n.

Exercise 2.4.20. Given a non-zero $\alpha \in \Lambda^k(V)$, there is a $\beta \in \Lambda^{n-k}(V)$ with $\alpha \wedge \beta \neq 0$ in $\Lambda^n(V)$.

Recall that $\Lambda^n(V)$ is isomorphic to $\mathbb{R}$. From this exercise we deduce easily that the bilinear pairing

$$\begin{aligned} \Lambda^k(V) \times \Lambda^{n-k}(V) &\longrightarrow \Lambda^n(V) \\ (\alpha, \beta) &\longmapsto \alpha \wedge \beta \end{aligned}$$

is *non-degenerate*; that is, the induced map

$$\begin{aligned} \Lambda^k(V) &\longrightarrow \mathrm{Hom}\big(\Lambda^{n-k}(V), \Lambda^n(V)\big) \\ \alpha &\longmapsto (\beta \mapsto \alpha \wedge \beta) \end{aligned}$$

is an isomorphism. Note that $\Lambda^n(V)$ is isomorphic to $\mathbb{R}$, but not canonically: to fix an isomorphism we need to equip V with some additional structure, as we will soon see.

2.4.11. The rescaled scalar product on the exterior algebra. Let V have dimension n and be equipped with a scalar product g. This induces a scalar product g on each tensor space $\mathcal{T}^k(V)$ and hence on $\Lambda^k(V)$.

Exercise 2.4.21. Let $v^1, \dots, v^k, w^1, \dots, w^k \in V^*$ be covectors. We have

$$g(v^1 \wedge \cdots \wedge v^k, w^1 \wedge \cdots \wedge w^k) = k! \det\langle v^i, w^j\rangle.$$

Hint. Use Exercise 2.3.4. □

The $k!$ factor in the formula is slightly annoying, so it is customary to replace g with the rescaled scalar product

$$\langle \alpha, \beta \rangle = \frac{1}{k!} g(\alpha, \beta).$$

Now we get the simpler formula

$$\langle v^1 \wedge \cdots \wedge v^k, w^1 \wedge \cdots \wedge w^k \rangle = \det\langle v^i, w^j\rangle.$$

In particular, if $v^1, \dots, v^n$ is an orthonormal basis of covectors the elements

$$v^{i_1} \wedge \cdots \wedge v^{i_k}$$

with $i_1 < \cdots < i_k$ form an orthonormal basis for $\Lambda^k(V)$.

2.5. Orientation

We now introduce and discuss the notion of orientation on a real vector space V and its consequences on the tensor spaces, and in particular on the exterior algebra.

2.5.1. Definition. Let us say that two basis of V are *cooriented* if the change of basis matrix relating them has positive determinant. Being cooriented is an equivalence relation on the set of all the basis in V, and one checks immediately that we get precisely two equivalence classes of bases.

Definition 2.5.1. An *orientation* on V is the choice of one of these two equivalence classes.

If V is oriented, the bases belonging to the preferred equivalence class are called *positive*, and the other *negative*. Of course V has two distinct orientations. The space $\mathbb{R}^n$ has a canonical orientation given by the canonical basis, but a space V may not have a canonical orientation in general.

Exercise 2.5.2. If $V = U \oplus W$, then an orientation on any two of the spaces U, V, W induces an orientation on the third, by requiring that, for every positive basis $u_1, \dots, u_k$ of U and $w_1, \dots, w_h$ of W, the basis $u_1, \dots, u_k, w_1, \dots, w_h$ of V is also positive.

2.5.2. Via the determinant line. An orientation on V induces one on V^* and vice-versa, as follows: we require a basis on V to be positive $\Longleftrightarrow$ its dual basis on V^* is positive.

Proposition 2.4.15 in turn shows that an orientation on V^* induces one on $\Lambda^n(V)$ and vice-versa: a basis $v^1, \ldots, v^n$ is positive in $V^* \Longleftrightarrow$ the element $v^1 \wedge \cdots \wedge v^n$ is a positive basis for the line $\Lambda^n(V)$.

Summing everything up, we could define an orientation on V to be an orientation on the determinant line $\Lambda^n(V)$.

2.5.3. Scalar product. If V is equipped with both an orientation and a scalar product g, then we get for free a canonical generator ω for the determinant line $\Lambda^n(V)$ by taking

$$\omega = v^1 \wedge \cdots \wedge v^n$$

where $v^1, \ldots, v^n$ is *any* positive orthonormal basis of V^*. The generator ω does not depend on the basis, because any two such basis are related by a matrix A with $\det A = 1$ and hence Proposition 2.4.15 applies. The element ω is also determined by requiring that

$$\omega(v_1, \ldots, v_n) = 1$$

on every positive orthonormal basis $v_1, \ldots, v_n$ of V.

2.5.4. The Hodge star operator. Let V be equipped with both an orientation and a scalar product of some signature (p, m). This induces a scalar product $\langle , \rangle$ on each $\Lambda^k(V)$, rescaled as in Section 2.4.11. Let ω be the canonical generator of $\Lambda^n(V)$. Note that $\langle \omega, \omega \rangle = (-1)^m$.

The *Hodge star operator* is the linear map

$$*\colon \Lambda^k(V) \longrightarrow \Lambda^{n-k}(V)$$

that sends $\beta \in \Lambda^k(V)$ to the unique $*\beta \in \Lambda^{n-k}(V)$ such that

$$\alpha \wedge (*\beta) = \langle \alpha, \beta \rangle \omega$$

for all $\alpha \in \Lambda^k(V)$. The map is well-defined because the bilinear pairing $\wedge$ is non-degenerate, see Section 2.4.10.

Exercise 2.5.3. The following hold:

(1) If $v^1, \ldots, v^n$ is a positive orthonormal basis for V^*, then

$$*(v^1 \wedge \cdots \wedge v^k) = (-1)^{m'} v^{k+1} \wedge \cdots \wedge v^n$$

where m' is the number of vectors in $v^1, \ldots, v^k$ with $\langle v^i, v^i \rangle = -1$.

(2) The map $*$ is an isomorphism. If m is even the map $*$ is an isometry.

(3) For every $\beta \in \Lambda^k(V)$ we have $**\beta = (-1)^{k(n-k)+m}\beta$.

If $n = 2k$ the Hodge star operator $*\colon \Lambda^k(V) \to \Lambda^k(V)$ is an automorphism. If moreover k is even (so n is divisible by four) and m is also even (for instance, if the scalar product is positive definite), the exercise says that $*$ is an isometric involution. Since $*^2 = \mathrm{id}$, the vector space $\Lambda^k(V)$ splits into its ± 1 eigenspaces

$$\Lambda^k(V) = \Lambda^k_+(V) \oplus \Lambda^k_-(V)$$

where $\alpha \in \Lambda^k_\pm(V) \iff *\alpha = \pm\alpha$. The elements in $\Lambda^k_+(V)$ and $\Lambda^l_-(V)$ are called respectively *self-dual* and *anti-self-dual*.

Exercise 2.5.4. If $\dim V = 4$, the scalar product g is positive definite, and v^1, v^2, v^3, v^4 is a positive orthonormal basis for V^*, then a basis for $\Lambda^2_+(V)$ is

$$v^1 \wedge v^2 + v^3 \wedge v^4, \quad v^1 \wedge v^3 + v^4 \wedge v^2, \quad v^1 \wedge v^4 + v^2 \wedge v^3.$$

A basis for $\Lambda^2_-(V)$ is

$$v^1 \wedge v^2 - v^3 \wedge v^4, \quad v^1 \wedge v^3 - v^4 \wedge v^2, \quad v^1 \wedge v^4 - v^2 \wedge v^3.$$

2.6. Grassmannians

After many pages of algebra, we now would like to see some geometric applications of the structures that we have just introduced. Here is one.

2.6.1. Definition. Let V be a real vector space of dimension n. Remember that the projective space $\mathbb{P}(V)$ is the set of all the vector lines in V. More generally, fix $0 < k < n = \dim V$.

Definition 2.6.1. The *Grassmannian* $\mathrm{Gr}_k(V)$ is the set consisting of all the k-dimensional vector subspaces $W \subset V$.

Recall that every $W \subset V$ determines a subspace $W^0 \subset V^*$ consisting of all the functionals that vanish on W. We have $\dim W^0 = n - \dim W$. The sets $\mathrm{Gr}_k(V)$ and $\mathrm{Gr}_{n-k}(V^*)$ may thus be identified canonically. In particular

$$\mathrm{Gr}_1(V) = \mathbb{P}(V), \qquad \mathrm{Gr}_{n-1}(V) = \mathbb{P}(V^*).$$

The simplest new interesting set to investigate is the Grassmannian $\mathrm{Gr}_2(\mathbb{R}^4)$ of vector planes in $\mathbb{R}^4$. How can we study such an object?

2.6.2. The Plücker embedding. A generic Grassmannian is not a projective space in any sense, but we now show that it can be embedded in some (bigger) projective space. We do this using the exterior algebra.

For every k-dimensional subspace $W \subset V$ of V we have the inclusion map $L\colon W \to V$ which induces an injective linear map

$$\Lambda_k(W) \longrightarrow \Lambda_k(V).$$

Since $\dim \Lambda_k(W) = 1$, the image of this map is a line in $\Lambda_k(V)$ that depends only on W. By sending W to this line we get a map

$$\mathrm{Gr}_k(V) \longrightarrow \mathbb{P}\big(\Lambda_k(V)\big)$$

called the *Plücker embedding*. Concretely, the map sends $W \subset V$ to

$$[w_1 \wedge \cdots \wedge w_k]$$

where $w_1, \ldots, w_k$ is any basis of W.

Proposition 2.6.2. *The Plücker embedding is injective.*

Proof. Consider $W \neq W'$. Let $w_1, \ldots, w_k$ and $w'_1, \ldots, w'_k$ be any basis of W and W'. Pick any vector $w \in W \setminus W'$. By Exercise 2.4.17 we have

$$w_1 \wedge \cdots \wedge w_k \wedge w = 0, \qquad w'_1 \wedge \cdots \wedge w'_k \wedge w \neq 0.$$

Therefore the tensors $w_1 \wedge \cdots \wedge w_k$ and $w'_1 \wedge \cdots \wedge w'_k$ are not proportional. $\square$

For instance, we get the Plücker embedding

$$\mathrm{Gr}_2(\mathbb{R}^4) \longleftrightarrow \mathbb{P}\big(\Lambda_2(\mathbb{R}^4)\big) \cong \mathbb{P}\Big(\mathbb{R}^{\binom{4}{2}}\Big) = \mathbb{RP}^5.$$

This map is clearly not surjective because of Exercise 2.4.18.

2.6.3. The Veronese embedding. Here is another geometric application. Fix $k > 0$ and consider the natural map $V \to S^k(V)$ defined as

$$v \longmapsto \underbrace{v \odot \cdots \odot v}_{k}.$$

This map is not linear in general, however it is injective (exercise) and it also induces an injective map between projective spaces

$$\mathbb{P}(V) \hookrightarrow \mathbb{P}\big(S^k(V)\big)$$

called the *Veronese embedding*. This map is not a projective map in general.

Exercise 2.6.3. If $V = \mathbb{R}^{n+1}$ and we use the canonical basis, we get

$$\mathbb{P}^n \longleftrightarrow \mathbb{P}^N$$

where $N = \binom{n+k}{k} - 1$. The map sends $[x_0, \ldots, x_n]$ to $[x_0^k, x_0^{k-1}x_1, \ldots]$ where the square brackets contain all the possible degree-k monomials in the variables $x_0, \ldots, x_n$. For instance for $k = n = 2$ we get

$$\mathbb{P}^2 \longleftrightarrow \mathbb{P}^5$$

given by

$$[x, y, z] \longmapsto [x^2, y^2, z^2, xy, yz, zx].$$

For $n = 1$ we get

$$\mathbb{P}^1 \longleftrightarrow \mathbb{P}^k$$

given by

$$[x, y] \longmapsto [x^k, x^{k-1}y, \ldots, xy^{k-1}, y^k].$$

2.7. Exercises

Exercise 2.7.1. Let U, V, W be finite-dimensional vector spaces. Show that there is a canonical isomorphism

$$\mathrm{Mult}(U, V; W) \longrightarrow \mathrm{Hom}(U, \mathrm{Hom}(V, W)).$$

Exercise 2.7.2. Let V be a finite-dimensional vector space. Show that every tensor T of type $(0, 2)$ may be written uniquely as a sum of a symmetric and an antisymmetric tensor. Show that this is not true for tensors of type $(0, n)$ with $n \geq 3$.

Exercise 2.7.3. Let V be a finite-dimensional vector space. Let T be a tensor of type $(0, k)$. Prove the following equivalences:

- T is antisymmetric $\Longleftrightarrow T(v_1, \ldots, v_k) = 0$ if two of the v_i's coincide.
- $S(T) = 0 \Longleftrightarrow T(v, \ldots, v) = 0$ for every $v \in V$.

Note that the two conditions are stronger than simply requiring that (in some coordinates) $T_{i_1, \ldots, i_k} = 0$ whenever two or all the indices coincide (respectively).

Exercise 2.7.4. Let V be a vector space of dimension n and $v \in V$ a fixed vector. Show that the contraction ι_v may be characterised as the unique linear map $\iota_v \colon \Lambda^k \to \Lambda^{k-1}$ that satisfies these axioms for all k:

(1) for $k = 1$ we have $\iota_v(w^*) = w^*(v)$;
(2) for every $T \in \Lambda^k(V)$ and $U \in \Lambda^k(V)$ we get

$$\iota_v(T \wedge U) = \iota_v(T) \wedge U + (-1)^k T \wedge \iota_v(U).$$

Exercise 2.7.5. Let V be equipped with both an orientation and a scalar product of some signature (p, m).

(1) For a (not necessarily orthonormal) basis $v^1, \ldots, v^n$ of V we get

$$* (v^1 \wedge \cdots \wedge v^k) = \frac{\sqrt{|\det g|}}{(n-k)!} g^{1j_1} \cdots g^{kj_k} \epsilon_{j_1 \cdots j_n} v^{j_{k+1}} \wedge \cdots \wedge v^{j_n}. \tag{5}$$

Here $\epsilon_{j_1 \cdots j_n}$ is the *Levi-Civita symbol*, which is 0 if the indices $j_1, \ldots, j_n$ are not distinct, and equals the sign of the permutation $(j_1, \ldots, j_n)$ if they are distinct. In coordinates we get

$$(*T)_{j_{k+1}, \ldots, j_n} = \frac{\sqrt{|\det g|}}{k!} g^{i_1 j_1} \cdots g^{i_k j_k} \epsilon_{j_1 \cdots j_n} T_{i_1, \ldots, i_k}.$$

If we use g to raise indices (as usual) we may write this simply as

$$(*T)_{j_{k+1}, \ldots, j_n} = \frac{\sqrt{|\det g|}}{k!} \epsilon_{j_1 \cdots j_n} T^{j_1, \ldots, j_k}.$$

Part 2

Differential topology

CHAPTER 3

Smooth manifolds

3.1. Smooth manifolds

We introduce here the notion of *smooth manifold*, the main protagonist of the book.

3.1.1. Definition. The definition of topological manifold that we have proposed in Section 1.1.8 is simple but also very poor, and it is quite hard to employ it concretely: for instance, it is already non obvious to answer such a natural question as whether $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic when $n \neq m$. To make life easier, we enrich the definition by adding a *smooth structure* that exploits the power of differential calculus.

Let M be a topological n-manifold. A *chart* is a homeomorphism $\varphi\colon U \to V$ from some open set $U \subset M$ onto an open set $V \subset \mathbb{R}^n$. The inverse map $\varphi^{-1}\colon V \to U$ is called a *parametrisation*. An *atlas* on M is a set $\{\varphi_i\}$ of charts $\varphi_i\colon U_i \to V_i$ that cover M, that is such that $\cup U_i = M$.

Let $\{\varphi_i\}$ be an atlas on M. Whenever $U_i \cap U_j \neq \varnothing$, we define a *transition map*

$$\varphi_{ij}\colon \varphi_i(U_i \cap U_j) \longrightarrow \varphi_j(U_i \cap U_j)$$

by setting $\varphi_{ij} = \varphi_j \circ \varphi_i^{-1}$. The reader should visualise this definition by looking at Figure 3.1. Note that both the domain and codomain of φ_{ij} are open sets of $\mathbb{R}^n$, and hence it makes perfectly sense to ask whether the transition functions φ_{ij} are smooth. We say that the atlas is *smooth* if all the transition functions φ_{ij} are smooth. Here is the most important definition of the book:

Definition 3.1.1. A *smooth n-manifold* is a topological n-manifold equipped with a smooth atlas.

To be more precise, we allow the same smooth manifold to be described by different atlases, as follows: we say that two smooth atlases $\{\varphi_i\}$ and $\{\varphi'_j\}$ are *compatible* if their union is again a smooth atlas; compatibility is an equivalent relation and we define a *smooth structure* on a topological manifold M to be an equivalence class of smooth atlases on M. The rigorous definition of a smooth manifold is a topological manifold M with a smooth structure on it.

Remark 3.1.2. The union of all the smooth atlases in M compatible with a given one is again a compatible smooth atlas, called a *maximal atlas*. The maximal atlas is uniquely determined by the smooth structure: hence one can

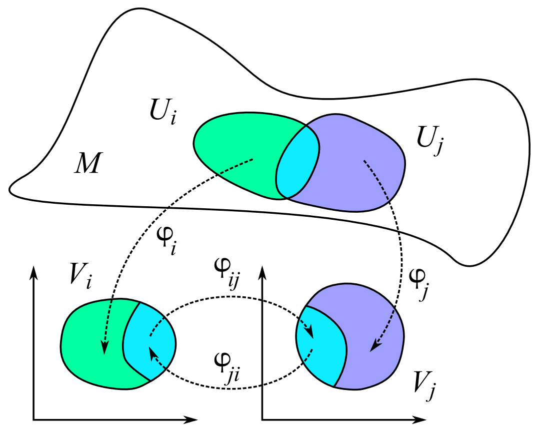

Figure 3.1. Two overlapping charts φ_i and φ_j induce a transition function $\varphi_{ij} = \varphi_j \circ \varphi_i^{-1}$.

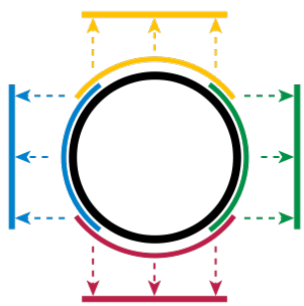

Figure 3.2. An atlas on a compact manifold..

also define a smooth manifold to be a topological manifold equipped with a maximal atlas, without using equivalence classes.

As a first example, every open subset $U \subset \mathbb{R}^n$ is naturally a smooth manifold, with an atlas that consists of a unique chart: the identity map $U \to U$.

The open subsets of $\mathbb{R}^n$ can be pretty complicated, but they are never compact. To construct some compact smooth manifolds we now build some atlases as in Figure 3.2.

3.1.2. Spheres. Recall that the *unit sphere* is

$$S^n = \{x \in \mathbb{R}^{n+1} \mid \|x\| = 1\}.$$

This is the prototypical example of a compact smooth manifold. To build a smooth atlas on S^n, we may consider the hemispheres

$$U_i^+ = \{x \in S^n \mid x^i > 0\}, \qquad U_i^- = \{x \in S^n \mid x^i < 0\}$$

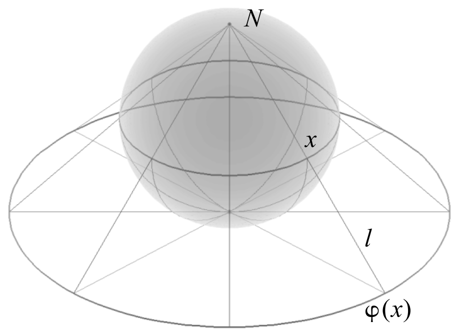

Figure 3.3. The stereographic projection sends a point $x \in S^n \setminus \{N\}$ to the point $\varphi(x)$ obtained by intersecting the line l containing N and x with the horizontal hyperplane $x^{n+1} = -1$.

for $i = 1, \ldots, n+1$ and define a chart $\varphi_i^\pm \colon U_i^\pm \to B^n$ by forgetting x^i, that is

$$\varphi_i^\pm(x^1, \ldots, x^{n+1}) = (x^1, \ldots, \check{x}^i, \ldots, x^{n+1}).$$

Proposition 3.1.3. *These charts define a smooth atlas on S^n.*

Proof. The inverse $(\varphi_i^\pm)^{-1}$ is

$$(y^1, \ldots, y^n) \longmapsto \left(y^1, \ldots, y^{i-1}, \pm\sqrt{1 - \|y\|^2}, y^i, \ldots, y^n\right).$$

The transition functions are compositions $\varphi_i^\pm \circ (\varphi_j^\pm)^{-1}$ and are smooth. □

We have equipped S^n with the structure of a smooth manifold. As we said, the same smooth structure on S^n can be built via a different atlas: for instance we describe one now that contains only two charts. Consider the north pole $N = (0, \ldots, 0, 1)$ in S^n and the *stereographic projection* $\varphi_N \colon S^n \setminus \{N\} \to \mathbb{R}^n$,

$$\varphi_N(x^1, \ldots, x^{n+1}) = \frac{2}{1 - x^{n+1}}(x^1, \ldots, x^n).$$

The geometric interpretation of the stereographic projection is illustrated in Figure 3.3. The map φ_N is a homeomorphism, so in particular $S^n \setminus \{N\}$ is homeomorphic to $\mathbb{R}^n$. We can analogously define a stereographic projection φ_S via the south pole $S = (0, \ldots, 0, -1)$, and deduce that $S^n \setminus \{S\}$ is also homeomorphic to $\mathbb{R}^n$.

Exercise 3.1.4. The two charts $\{\varphi_S, \varphi_N\}$ form a smooth atlas for S^n, compatible with the one defined above.

The atlases $\{\varphi_i^\pm\}$ and $\{\varphi_S, \varphi_N\}$ define the same smooth structure on S^n.

Remark 3.1.5. The circle S^1 is quite special: we can identify $\mathbb{C}$ with $\mathbb{R}^2$ and write $S^1 = \{e^{i\theta} \mid \theta \in \mathbb{R}\}$. The universal covering $\mathbb{R} \to S^1$, $\theta \mapsto e^{i\theta}$ is of course not injective, but it furnishes an atlas of natural charts when restricted to the open segments (a, b) with $b - a < 2\pi$. The transition maps are translations.

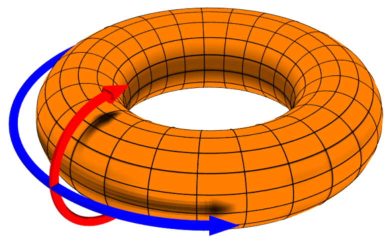

Figure 3.4. The torus $S^1 \times S^1$ embedded in $\mathbb{R}^3$. Every point $(e^{i\theta}, e^{i\varphi}) \in S^1 \times S^1$ of the torus may be interpreted on the figure as a point with (blue) longitude θ and (red) latitude φ. Note that the latitude and longitude behave very nicely on the torus, as opposite to the sphere where longitude is ambiguous at the poles. Cartographers would enjoy to live on a torus-shaped planet.

3.1.3. Projective spaces. We now consider the real projective space $\mathbb{RP}^n$. Recall the every point in $\mathbb{RP}^n$ has some homogeneous coordinates $[x_0, \ldots, x_n]$.

For $i = 0, \ldots, n$ we set $U_i \subset \mathbb{RP}^n$ to be the open subset defined by the inequality $x^i \neq 0$. We define a chart $\varphi_i \colon U_i \to \mathbb{R}^n$ by setting

$$\varphi_i\big([x_0, \ldots, x_n]\big) = \left(\frac{x_0}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_n}{x_i}\right).$$

The inverse parametrisation $\varphi_i^{-1} \colon \mathbb{R}^n \to U_i$ may be written simply as

$$\varphi_i^{-1}(y_1, \ldots, y_n) = [y_1, \ldots, y_i, 1, y_{i+1}, \ldots, y_n].$$

The open subsets $U_0, \ldots, U_n$ cover $\mathbb{RP}^n$ and the transition functions φ_{ij} are clearly smooth: hence the atlas $\{\varphi_i\}$ defines a smooth structure on $\mathbb{RP}^n$.

We have discovered that $\mathbb{RP}^n$ is naturally a smooth n-manifold. The same construction works for the *complex* projective space $\mathbb{CP}^n$ which is hence a smooth $2n$-manifold: it suffices to identify $\mathbb{C}^{n+1}$ with $\mathbb{R}^{2n+2}$ in the usual way.

Recall that $\mathbb{RP}^n$ and $\mathbb{CP}^n$ are connected and compact, see Exercise 1.4.1.

3.1.4. Products. The product $M \times N$ of two smooth manifolds M, N of dimension m, n is naturally a smooth $(m+n)$-manifold. Indeed, two smooth atlases $\{\varphi_i\}$, $\{\psi_j\}$ on M, N induce a smooth atlas $\{\varphi_i \times \psi_j\}$ on $M \times N$.

For instance the *torus* $S^1 \times S^1$ is a smooth manifold of dimension two. We take this opportunity to mention that a 2-manifold is usually called a *surface*. We will soon prove that the torus may be conveniently embedded as a submanifold of $\mathbb{R}^3$ as in Figure 3.4: to do so we will need to define the notion of *embedding* and of *submanifold*.

3.1.5. No prior topology. We now make a useful observation. We note that it is not strictly necessary to priorly have a topology to define a smooth manifold structure: we can also proceed directly with atlases as follows.

Let X be any set. We define a *smooth atlas* on X to be a collection of subsets U_i covering X and of bijections $\varphi_i \colon U_i \to V_i$ onto open subsets of $\mathbb{R}^n$, such that $\varphi_i(U_i \cap U_j)$ is open for every i, j, and the transition maps $\varphi_{ij} = \varphi_j \circ \varphi_i^{-1}$ are smooth wherever they are defined.

Exercise 3.1.6. There is a unique topology on X such that every U_i is open and every $\varphi_i \colon U_i \to V_i$ is a homeomorphism. In this topology, a subset $U \subset X$ is open $\iff$ the sets $\varphi_i(U \cap U_i)$ are open for every i.

Therefore a smooth atlas on a set X defines a compatible topology. If this topology is Hausdorff and second-countable, this gives a smooth manifold structure on X.

3.1.6. Grassmannians. We apply the "no prior topology" philosophy to define a smooth manifold structure on the Grassmannian.

Remember from Section 2.6 that the Grassmannian $\mathrm{Gr}_k(V)$ is the set of all k-dimensional vector subspaces $W \subset V$. We now define a smooth manifold structure on $\mathrm{Gr}_k(V)$ by assigning it a smooth atlas $\mathcal{A}$.

For every basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of V, we define the subspaces

$$W = \mathrm{Span}(v_1, \ldots, v_k), \quad Z = \mathrm{Span}(v_{k+1}, \ldots, v_n).$$

Of course $V = W \oplus Z$. Now we consider the set

$$U_{\mathcal{B}} = \left\{ W' \subset V \;\middle|\; V = W' \oplus Z \right\}.$$

The set $U_{\mathcal{B}}$ is a subset of $\mathrm{Gr}_k(V)$ that contains W. We now define a map

$$\begin{aligned} f_{\mathcal{B}} \colon \underbrace{Z \times \cdots \times Z}_{k} &\longrightarrow U_{\mathcal{B}} \\ (z_1, \ldots, z_k) &\longmapsto \mathrm{Span}(v_1 + z_1, \ldots, v_k + z_k). \end{aligned}$$

It is a linear algebra exercise to show that $f_{\mathcal{B}}$ is a bijection. The given basis $v_{k+1}, \ldots, v_n$ allows us to identify Z with $\mathbb{R}^{n-k}$, so we get a bijection

$$f_{\mathcal{B}} \colon \mathbb{R}^{(n-k)k} \longrightarrow U_{\mathcal{B}}.$$

The atlas $\mathcal{A}$ for $\mathrm{Gr}_k(V)$ is formed by all the maps $f_{\mathcal{B}}^{-1} \colon U_{\mathcal{B}} \to \mathbb{R}^{(n-k)k}$ as $\mathcal{B}$ varies among all the basis of V. It is now an exercise to show that the transition maps are defined on open sets and smooth. So we have given $\mathrm{Gr}_k(V)$ the structure of a smooth manifold of dimension $(n-k)k$.

3.2. Smooth maps

Every honest category of objects has its morphisms. We have defined the smooth manifolds, and we now introduce the right kind of maps between them.

We will henceforth use the following convention: if M is a given smooth manifold, we just call a *chart* on M any chart $\varphi \colon U \to V$ compatible with the smooth structure on M.

3.2.1. Definition. We say that a map $f: M \to N$ between two smooth manifolds is *smooth* if it is so when read along some charts. This means that for every $x \in M$ there are some charts $\varphi: U \to V$ and $\psi: W \to Z$ of M and N, with $x \in U$ and $f(U) \subset W$, such that the map

$$\psi \circ f \circ \varphi^{-1}: V \longrightarrow Z$$

is smooth. Note that the manifolds M and N may have different dimensions. It may be useful to visualise this definition via a commutative diagram:

$$\begin{tikzcd} U \arrow[r, "f"] \arrow[d, "\varphi"'] & W \arrow[d, "\psi"] \\ V \arrow[r, "F"'] & Z \end{tikzcd}$$

Here $F = \psi \circ f \circ \varphi^{-1}$ should be thought as "the map f read on charts".

Remark 3.2.1. If $f: M \to N$ is smooth then $\psi \circ f \circ \varphi^{-1}$ is also smooth for *any* charts φ and ψ as above. This is a typical situation: if something is smooth on *some* charts that cover M, it is so on *all* charts, because the transition functions are smooth and the composition of smooth maps is smooth.

A *curve* in M is a smooth map $\gamma: I \to M$ defined on some open interval $I \subset \mathbb{R}$, that may be bounded or unbounded. Curves play an important role in differential topology and geometry.

Exercise 3.2.2. The inclusion $S^n \hookrightarrow \mathbb{R}^{n+1}$ is a smooth map.

The space of all the smooth maps $M \to N$ is usually denoted by $C^\infty(M, N)$. We will often encounter the space $C^\infty(M, \mathbb{R})$, written as $C^\infty(M)$ for short. We note that $C^\infty(M)$ is a real commutative algebra.

3.2.2. Diffeomorphisms. A smooth map $f: M \to N$ is a *diffeomorphism* if it is a bijection and its inverse $f^{-1}: N \to M$ is also smooth. Of course a diffeomorphism is also a homeomorphism, but the converse is often not true.

Example 3.2.3. The map $f: B^n \to \mathbb{R}^n$ defined as

$$f(x) = \frac{x}{\sqrt{1 - \|x\|^2}}$$

is a diffeomorphism. Its inverse is

$$g(x) = \frac{x}{\sqrt{1 + \|x\|^2}}.$$

Two manifolds M, N are *diffeomorphic* if there is a diffeomorphism $f: M \to N$. Being diffeomorphic is clearly an equivalence relation. The open ball of radius $r > 0$ centred at $x_0 \in \mathbb{R}^n$ is by definition

$$B(x_0, r) = \{x \in \mathbb{R}^n \mid \|x - x_0\| < r\}.$$

Exercise 3.2.4. Any two open balls in $\mathbb{R}^n$ are diffeomorphic.

As a consequence, every open ball in $\mathbb{R}^n$ is diffeomorphic to $\mathbb{R}^n$ itself.

Exercise 3.2.5. The antipodal map $\iota\colon S^n \to S^n$, $\iota(x) = -x$ is a diffeomorphism.

Example 3.2.6. The following diffeomorphisms hold:

$$\mathbb{RP}^1 \cong S^1, \qquad \mathbb{CP}^1 \cong S^2.$$

These are obtained as compositions

$$\mathbb{RP}^1 \longrightarrow \mathbb{R} \cup \{\infty\} \longrightarrow S^1$$
$$\mathbb{CP}^1 \longrightarrow \mathbb{C} \cup \{\infty\} \longrightarrow S^2$$

where the first map sends $[x_0, x_1]$ to x_1/x_0, and the second is the stereographic projection. All the maps are clearly 1-1 and we only need to check that the composition is smooth, and with smooth inverse. Everything is obvious except near the point $[0, 1]$. In the complex case, if we take the parametrisation $z \mapsto [z, 1]$, by calculating we find (exercise) that the map is

$$[z, 1] \longmapsto \frac{1}{1 + 4|z|^2}\left(4\Re z, -4\Im z, 1 - 4|z|^2\right).$$

So it is smooth and has smooth inverse.

3.3. Partitions of unity

We now introduce a powerful tool that may look quite technical at a first reading, but which will have spectacular consequences in the next pages. The general idea is that smooth functions are flexible enough to be patched altogether: one can use bump functions (see Section 1.3.5) to extend smooth maps from local to global, to approximate continuous maps with smooth maps, and to do much more in the next chapters.

3.3.1. Adequate atlas. Let M be a smooth manifold. We now introduce a type of atlas that is very convenient to prove theorems.

Definition 3.3.1. An atlas $\{\varphi_i\colon U_i \to \mathbb{R}^n\}$ for M is *adequate* if

(1) the open sets $\{U_i\}$ form a locally finite covering of M,
(2) the open subsets $V_i = \varphi_i^{-1}(B^n)$ also form a covering of M.

We should visualise an adequate atlas as in Figure 3.5: a locally finite set of patches U_i diffeomorphic to $\mathbb{R}^n$, each containing a $V_i \cong B^n$, such that the V_i's cover M.

We already know that M is paracompact by Proposition 1.1.24, so every open covering has a locally finite refinement. We reprove here this fact in a stronger and more useful form.

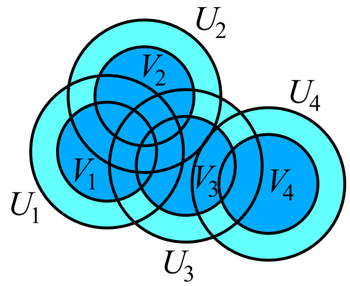

Figure 3.5. An adequate atlas.

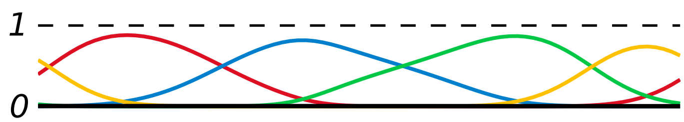

Figure 3.6. A partition of unity on S^1 (the endpoints should be identified).

Proposition 3.3.2. *Let $\{U_i\}$ be an open covering of M. There is an adequate atlas $\{\varphi_k \colon W_k \to \mathbb{R}^n\}$ such that $\{W_k\}$ refines $\{U_i\}$.*

Proof. We readapt the proof of Proposition 1.1.24. We know that M has an exhaustion by compact subsets $\{K_j\}$, and we set $K_0 = K_{-1} = \varnothing$.

We construct the atlas inductively on $j = 1, 2\ldots$ For every $p \in K_j \setminus \mathrm{int}(K_{j-1})$ there is an open set U_i containing p. We fix a chart $\varphi_p \colon W_p \to \mathbb{R}^n$ with $p \in W_p \subset \big(\mathrm{int}(K_{j+1}) \setminus K_{j-2}\big) \cap U_i$.

The open sets $\varphi_p^{-1}(B^n)$ cover the compact set $K_j \setminus \mathrm{int}(K_{j-1})$ as p varies there, and finitely many of them suffice to cover it. By taking only these finitely many φ_p for every $j = 1, 2, \ldots$ we get an adequate covering. $\square$

3.3.2. Partition of unity.

Let $\{U_i\}$ be an open covering of M.

Definition 3.3.3. A *partition of unity* subordinate to the open covering $\{U_i\}$ is a family $\{\rho_i \colon M \to \mathbb{R}\}$ of smooth functions with values in $[0, 1]$, such that the following hold:

(1) the support of ρ_i is contained in U_i for all i,
(2) every $x \in M$ has a neighbourhood where all but finitely many of the ρ_i vanish, and $\sum_i \rho_i(x) = 1$.

See an example in Figure 3.6. What is important for us, is that partitions of unity exist.

Proposition 3.3.4. *For every open covering $\{U_i\}$ of M there is a partition of unity subordinate to $\{U_i\}$.*

Proof. Fix a smooth bump function $\lambda \colon \mathbb{R}^n \to \mathbb{R}$ with values in $[0, 1]$ such that $\lambda(x) = 1$ if $\|x\| \leq 1$ and $\lambda(x) = 0$ if $\|x\| \geq 2$, see Section 1.3.5.

Pick an adequate atlas $\left\{\varphi_k : W_k \to \mathbb{R}^n\right\}$ such that $\{W_k\}$ refines $\{U_i\}$. Define the function $\bar{\rho}_k : M \to \mathbb{R}$ as $\bar{\rho}_k(p) = \lambda(\varphi_k(p))$ if $p \in W_k$ and zero otherwise. The family $\{\bar{\rho}_k\}$ is almost a partition of unity subordinate to $\{W_k\}$, except that $\sum_j \bar{\rho}_j(p)$ may be any strictly positive number (note that it is not zero because the atlas is adequate). To fix this it suffices to set

$$\rho_k(p) = \frac{\bar{\rho}_k(p)}{\sum_j \bar{\rho}_j(p)}.$$

The family $\{\rho_k\}$ is a partition of unity subordinate to $\{W_k\}$. To get one $\{\eta_i\}$ subordinate to $\{U_i\}$ we fix a function $i(k)$ such that $W_k \subset U_{i(k)}$ for every k and we define

$$\eta_i(p) = \sum_{i(k)=i} \rho_k(p).$$

The proof is complete. □

3.3.3. Extension of smooth maps. We show an application of the partitions of unity. Let M and N be two smooth manifolds. The fact that we prove here is already interesting and non-trivial when M is $\mathbb{R}^m$ or some open set in it. We first need to define a notion of smooth map for arbitrary (not necessarily open) domains.

Definition 3.3.5. Let $S \subset M$ be any subset. A map $f : S \to N$ is *smooth* if it is locally the restriction of smooth functions. That is, for every $p \in S$ there are an open neighbourhood $U \subset M$ of p and a smooth map $g : U \to N$ such that $g|_{U\cap S} = f|_{U\cap S}$.

One may wonder whether the existence of local extensions implies that of a global one. This is true if the domain is closed and the codomain is $\mathbb{R}^n$.

Proposition 3.3.6. *If $S \subset M$ is a closed subset, every smooth map $f : S \to \mathbb{R}^n$ is the restriction of a smooth map $F : M \to \mathbb{R}^n$.*

Proof. By definition for every $p \in S$ there are an open neighbourhood $U(p)$ and a local extension $g_p : U(p) \to \mathbb{R}^n$ of f. Consider the open covering

$$\left\{U(p)\right\}_{p\in S} \cup \left\{M \setminus S\right\}$$

of M, and pick a partition of unity $\{\rho_p\} \cup \{\rho\}$ subordinate to it. For every $x \in M$ we define

$$F(x) = \sum \rho_p(x) g_p(x)$$

where the sum is taken over the finitely many $p \in M$ such that $\rho_p(x) \neq 0$. The function $F : M \to \mathbb{R}^n$ is locally a finite sum of smooth functions and is hence smooth. If $x \in S$ we have

$$F(x) = \sum \rho_p(x) g_p(x) = \sum \rho_p(x) f(x) = f(x) \sum \rho_p(x) = f(x).$$

Therefore $F : M \to \mathbb{R}^n$ is a smooth global extension of f. □

Remark 3.3.7. Smooth (not even continuous) extensions cannot exist for every $S \subset M$ for obvious reasons. Take for instance $M = \mathbb{R}$ and $S = \mathbb{R}^* = \mathbb{R} \setminus \{0\}$ and $f : S \to \mathbb{R}$ with $f(x) = 1$ on $x > 0$ and $f(x) = 0$ on $x < 0$.

Remark 3.3.8. In the proof, the extension F vanishes outside $\cup_{p \in S} U(p)$. In the construction we may take the $U(p)$ to be arbitrarily small: hence we may require F to vanish outside of an arbitrary open neighbourhood of S.

3.3.4. Approximation of continuous maps. Here is another application of the partition of unity. Let M be a smooth manifold.

Proposition 3.3.9. *Let $f : M \to \mathbb{R}^n$ be a continuous map, whose restriction $f|_S$ to some (possibly empty) closed subset $S \subset M$ is smooth. For every continuous positive function $\varepsilon : M \to \mathbb{R}_{>0}$ there is a smooth map $g : M \to \mathbb{R}^n$ with $f(x) = g(x)$ for all $x \in S$ and $\|f(x) - g(x)\| < \varepsilon(x)$ for all $x \in M$.*

Proof. The map g is easily constructed locally: for every $p \in M$ there are an open neighbourhood $U(p) \subset M$ and a smooth map $g_p : U(p) \to \mathbb{R}^n$ such that $f(x) = g_p(x)$ for all $x \in U(p) \cap S$ and $\|f(x) - g_p(x)\| < \varepsilon(x)$ for all $x \in U(p)$. (This is proved as follows: if $p \in S$, let g_p be an extension of f, while if $p \notin S$ simply set $g_p(x) = f(p)$ constantly. The second condition is then achieved by restricting $U(p)$.)

We now paste the g_p to a global map by taking a partition of unity $\{\rho_p\}$ subordinated to $\big\{U(p)\big\}$ and defining

$$g(x) = \sum \rho_p(x) g_p(x).$$

The sum is taken over the finitely many $p \in M$. such that $\rho_p(x) \neq 0$. The map $g : M \to \mathbb{R}^n$ is smooth and $f(x) = g(x)$ for all $x \in S$. Moreover

$$\begin{aligned} \|f(x) - g(x)\| &= \Big\| \sum \rho_p(x) f(x) - \sum \rho_p(x) g_p(x) \Big\| \\ &\leq \sum \rho_p(x) \big\| f(x) - g_p(x) \big\| < \sum \rho_p(x) \varepsilon(x) = \varepsilon(x). \end{aligned}$$

The proof is complete. $\square$

We have proved in particular that every continuous map $f : M \to \mathbb{R}^n$ may be approximated by smooth functions.

3.3.5. Smooth exhaustions. Here is another application. A *smooth exhaustion* on a manifold M is a smooth positive function $f : M \to \mathbb{R}_{>0}$ such that $f^{-1}[0, T]$ is compact for every T.

Proposition 3.3.10. *Every manifold M has a smooth exhaustion.*

Proof. Pick a locally finite, hence countable, covering $\big\{U_i\big\}$ where $\bar{U}_i$ is compact for every i, and a subordinated partition of unity ρ_i. The function

$$f(p) = \sum_{j=1}^{\infty} j \rho_j(p)$$

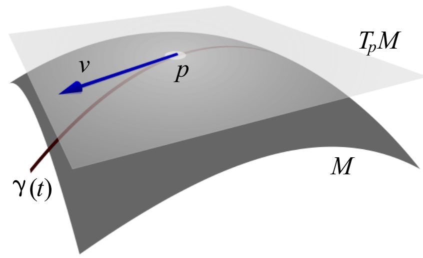

Figure 3.7. The tangent space T_pM is the set of all curves γ passing through p up to some equivalence relation.

is easily seen to be a smooth exhaustion. □

3.4. Tangent space

Let M be a smooth n-manifold. We now define for every point $p \in M$ a n-dimensional real vector space T_pM called the *tangent space* of M at p.

Heuristically, the tangent space T_pM should generalise the intuitive notions of tangent line to a curve in $\mathbb{R}^2$ or $\mathbb{R}^3$, or of a tangent plane to a surface in $\mathbb{R}^3$, as in Figure 3.7. There is however a problem here in trying to formalise this idea: our manifold M is an abstract object and is not embedded in some bigger space like the surface in $\mathbb{R}^3$ depicted in the figure! For that reason we need to define T_pM intrinsically, using only the points that are contained *inside* M and not *outside* – since there is no outside at all. We start to do this by considering all the curves passing through p: as suggested in Figure 3.7, every such curve γ should define somehow a tangent vector $v \in T_pM$. Afterwards we introduce a second more sophisticated definition where every tangent vector is introduced as a *derivation*.

3.4.1. Definition via curves. Here is a definition of the tangent space T_pM at $p \in M$. We fix a point $p \in M$ and consider all the curves $\gamma \colon I \to M$ with $0 \in I$ and $\gamma(0) = p$. (The open interval I may vary.) We want to define a notion of tangency of such curves at p. Let γ_1, γ_2 be two such curves.

If $M = \mathbb{R}^n$, the derivative $\gamma'(t)$ makes sense and we say as usual that γ_1 and γ_2 are *tangent* at p if $\gamma_1'(0) = \gamma_2'(0)$. On a more general M, we pick a chart $\varphi \colon U \to V$ and we say that γ_1 and γ_2 are *tangent* at p if the compositions $\varphi \circ \gamma_1$ and $\varphi \circ \gamma_2$ are tangent at $\varphi(p)$.[1]

This definition is chart-independent, that is it is not influenced by the choice of φ, because a transition map between two different charts transports tangent curves to tangent curves.

[1] To be precise, we may need to priorly restrict γ_1 and/or γ_2 to a smaller interval $I' \subset I$ in order for their images to lie in U.

The tangency at p is an equivalence relation on the set of all curves $\gamma\colon I \to M$ with $\gamma(0) = p$. We are ready to define T_pM.

Definition 3.4.1. The *tangent space* T_pM at $p \in M$ is the set of all curves $\gamma\colon I \to M$ with $0 \in I$ and $\gamma(0) = p$, considered up to tangency at p.

When $M = \mathbb{R}^n$, the space $T_p\mathbb{R}^n$ is naturally identified with $\mathbb{R}^n$ itself, by transforming every curve γ into its derivative $\gamma'(0)$. We will always write

$$T_p\mathbb{R}^n = \mathbb{R}^n.$$

This holds also for open subsets $M \subset \mathbb{R}^n$.

3.4.2. Definition via derivations. We now propose a more abstract and quite different definition of the tangent space at a point. It is always good to understand different equivalent definitions of the same mathematical object: the reader may choose the one she prefers, but we advise her to try to understand and remember both because, depending on the context, one definition may be more suitable than the other – for instance to prove theorems.

Let M be a smooth manifold and $p \in M$ be a point. A *derivation* v at p is an operation that assigns a number $v(f)$ to every smooth function $f\colon U \to \mathbb{R}$ defined in some open neighbourhood U of p, that fulfils the following requirements:

(1) if f and g agree on a neighbourhood of p, then $v(f) = v(g)$;
(2) v is linear, that is $v(\lambda f + \mu g) = \lambda v(f) + \mu v(g)$ for all numbers λ, μ;
(3) $v(fg) = v(f)g(p) + f(p)v(g)$.

In (2) and (3) we suppose that f and g are defined on the same open neighbourhood U. The term "derivation" is used here because the third requirement looks very much like the Leibniz rule. Here is a fresh new definition of the tangent space at a point:

Definition 3.4.2. The tangent space T_pM is the set of all derivations at p.

A linear combination $\lambda v + \lambda' v'$ of two derivations v, v' with $\lambda, \lambda' \in \mathbb{R}$ is again a derivation: therefore the tangent space T_pM has a natural structure of real vector space.

We study the model case $M = \mathbb{R}^n$. Here every vector $v \in \mathbb{R}^n$ determines the *directional derivative* ∂_v along v, defined as usual as

$$\partial_v f = \sum_{i=1}^n v^i \frac{\partial f}{\partial x^i},$$

which fulfils all the requirement (1-3) and is hence a derivation. Conversely:

Proposition 3.4.3. *If $M = \mathbb{R}^n$ every derivation is a directional derivative ∂_v along some vector $v \in \mathbb{R}^n$.*

Proof. We set $p = 0$ for simplicity. By the Taylor formula every smooth function f can be written near 0 as

$$f(x) = f(0) + \sum_i \frac{\partial f}{\partial x^i}(0)x^i + \sum_{i,j} h_{ij}(x)x^i x^j$$

for some smooth functions h_{ij}. If v is a derivation, by applying it to f we get

$$v(f) = f(0)v(1) + \sum_i \frac{\partial f}{\partial x^i}(0)v(x^i) + \sum_{i,j} v(h_{ij}x^i x^j).$$

The first and third term vanish because of the Leibniz rule (exercise: use that $v(1) = v(1 \cdot 1)$). Therefore v is the partial derivative along the vector $(v(x^1), \ldots, v(x^n))$. $\square$

We have discovered that when $M = \mathbb{R}^n$ the tangent space T_pM is naturally identified with $\mathbb{R}^n$. This works also if $M \subset \mathbb{R}^n$ is an open subset.

We have shown in particular that the two definitions – via curves and via derivations – of T_pM are equivalent at least for the open subsets $M \subset \mathbb{R}^n$. On a general M, here is a direct way to pass from one definition to the other: for every curve $\gamma\colon I \to M$ with $\gamma(0) = p$, we may define a derivation v by setting

$$v(f) = (f \circ \gamma)'(0).$$

This gives indeed a 1-1 correspondence between curves up to tangency and derivations, as one can immediately deduce by taking one chart.

Summing up, we have two equivalent definitions: the one via curves may look more concrete, but derivations have the advantage of giving T_pM a natural structure of a n-dimensional real vector space.

It is important to note that T_pM is a vector space and *nothing more than that*: for instance there is no canonical norm or scalar product on T_pM, so it does not make any sense to talk about the *lengths* of tangent vectors – tangent vectors have no lengths. We are lucky enough to have a well-defined vector space and we are content with that. To define lengths we need an additional structure called *metric tensor*, that we will introduce later on in the subsequent chapters.

3.4.3. Differential of a map. We now introduce some kind of derivative of a smooth map, called *differential*. The differential is neither a number, nor a matrix of numbers in any sense: it is “only” a linear function between tangent spaces that approximates the smooth map at every point.

Let $f\colon M \to N$ be a smooth map between smooth manifolds. The *differential* of f at a point $p \in M$ is the map

$$df_p\colon T_pM \longrightarrow T_{f(p)}N$$

that sends a curve γ with $\gamma(0) = p$ to the curve $f \circ \gamma$.

The map df_p is well-defined, because smooth maps send tangent curves to tangent curves, as one sees by taking charts. Alternatively, we may use derivations: the map df_p sends a derivation $v \in T_pM$ to the derivation $df_p(v) = v'$ that acts as $v'(g) = v(g \circ f)$.

Exercise 3.4.4. The function v' is indeed a derivation. The two definitions of df_p are equivalent; using the second one we see that df_p is linear.

The definition of df_p is clearly *functorial*, that is we have

$$d(g \circ f)_p = dg_{f(p)} \circ df_p, \qquad d(\mathrm{id}_M)_p = \mathrm{id}_{T_pM}.$$

This implies in particular that the differential df_p of a diffeomorphism $f\colon M \to N$ is invertible at every point $p \in M$.

When $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ are open subsets, the differential df_p of a smooth map $f\colon M \to N$ is a linear map

$$df_p\colon \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

because we have the natural identifications $T_pM = \mathbb{R}^m$ and $T_{f(p)}N = \mathbb{R}^n$. It is an exercise to check that df_p is just the ordinary differential of Section 1.3.1.

3.4.4. On charts. A constant refrain in differential topology and geometry is that an abstract highly non-numerical definition becomes a more concrete numerical object when read on charts. If $\varphi\colon U \to V$ and $\psi\colon W \to Z$ are charts of M and N with $f(U) \subset W$, then we may consider the commutative diagram

$$\begin{tikzcd} U \arrow[r, "f"] \arrow[d, "\varphi"'] & W \arrow[d, "\psi"] \\ V \arrow[r, "F"'] & Z \end{tikzcd}$$

where $F = \psi \circ f \circ \varphi^{-1}$ is the map f read on charts. By taking differentials we find for every $p \in U$ another commutative diagram of linear maps

$$\begin{tikzcd} T_pM \arrow[r, "df_p"] \arrow[d, "d\varphi_p"'] & T_{f(p)}N \arrow[d, "d\psi_{f(p)}"] \\ \mathbb{R}^m \arrow[r, "dF_{\varphi(p)}"'] & \mathbb{R}^n \end{tikzcd}$$

and $dF_{\varphi(p)}$ should be thought as "the differential df_p read on charts". Commutative diagrams are useful because they contain a lot of information in a single picture. The vertical arrows are isomorphisms, so one can fully recover df_p by looking at $dF_{\varphi(p)}$. In particular $dF_{\varphi(p)}$ has the same rank of df_p, and is injective/surjective $\iff$ df_p is.

It is convenient to look at $dF_{\varphi(p)}$ because it is a rather familiar object: being the differential of a smooth map $F\colon V \to Z$ between open sets $V \subset \mathbb{R}^m$

and $Z \subset \mathbb{R}^n$, the differential $dF_{\varphi(p)}$ is a quite reassuring Jacobian $n \times m$ matrix whose entries vary smoothly with respect to the point $\varphi(p) \in V$.

Example 3.4.5. The Veronese embedding $f\colon \mathbb{RP}^1 \hookrightarrow \mathbb{RP}^2$ is

$$f\big([x_0, x_1]\big) = [x_0^2, x_0x_1, x_1^2],$$

see Exercise 2.6.3. The map sends the open subset $U_0 = \{x_0 \neq 0\} \subset \mathbb{RP}^1$ into $W_0 = \{x_0 \neq 0\} \subset \mathbb{RP}^2$. We use the coordinate charts $\varphi\colon U_0 \to \mathbb{R}, [1, t] \mapsto t$ and $\psi\colon W_0 \to \mathbb{R}^2, [1, t, u] \mapsto (t, u)$. Read on these charts the map f transforms into a map $F = \psi \circ f \circ \varphi^{-1}\colon \mathbb{R} \to \mathbb{R}^2$, that is

$$F(t) = (t, t^2).$$

Its differential is $(1, 2t)$, so in particular it is injective. Analogously the chart $U_1 = \{x_1 \neq 0\} \subset \mathbb{RP}^1$ injects into $W_2 = \{x_2 \neq 0\} \subset \mathbb{RP}^2$ like $t \mapsto (t^2, t)$. We have discovered that df_p is injective for every $p \in \mathbb{RP}^1$.

Exercise 3.4.6. For every k, n and every $p \in \mathbb{RP}^n$, show that the differential df_p of the Veronese embedding $f\colon \mathbb{P}^n \hookrightarrow \mathbb{P}^N$ of Exercise 2.6.3 is injective.

3.4.5. Products. Let $M \times N$ be a product of smooth manifolds of dimensions m and n. For every $(p, q) \in M \times N$ there is a natural identification

$$T_{(p,q)}(M \times N) = T_pM \times T_qN.$$

This identification is immediate using the definition of tangent spaces via curves, since a curve in $M \times N$ is the union of two curves in M and N.

Exercise 3.4.7. The Segre embedding $f\colon \mathbb{RP}^1 \times \mathbb{RP}^1 \hookrightarrow \mathbb{RP}^3$ is

$$\big([x_0, x_1], [y_0, y_1]\big) \longmapsto [x_0y_0, x_0y_1, x_1y_0, x_1y_1].$$

See Section 2.1.5. Prove that for every $(p, q) \in \mathbb{RP}^1 \times \mathbb{RP}^1$ the differential $df_{(p,q)}$ is injective.

3.4.6. Velocity of a curve. If $\gamma\colon I \to M$ is a curve, for every $t \in I$ we get a differential $d\gamma_t\colon T_t\mathbb{R} \to T_{\gamma(t)}M$. Since $T_t\mathbb{R} = \mathbb{R}$ we may simply write $d\gamma_t\colon \mathbb{R} \to T_{\gamma(t)}M$ and it makes sense to define the *velocity* of γ at the time t as the tangent vector

$$\gamma'(t) = d\gamma_t(1) \in T_{\gamma(t)}M.$$

If we use the description of T_pM via curves, the definition of the velocity is rather tautological: the velocity of a curve at a point is the curve itself. Recall as we said above that there is no norm in $T_{\gamma(t)}M$, hence there is no way to quantify the "speed" of $\gamma'(t)$ as a number – except when it is zero.

3.4.7. Inverse Function Theorem. The Inverse Function Theorem 1.3.3 applies to this context. We say that $f\colon M \to N$ is a *local diffeomorphism* at $p \in M$ if there is an open neighbourhood $U \subset M$ of p such that $f(U) \subset N$ is open and $f|_U\colon U \to f(U)$ is a diffeomorphism.

Theorem 3.4.8. *A smooth map $f\colon M \to N$ is a local diffeomorphism at $p \in M \iff$ its differential df_p is invertible.*

Proof. Apply Theorem 1.3.3 to $\psi \circ f \circ \varphi^{-1}$ for some charts φ, ψ. $\square$

Exercise 3.4.9. Consider the map $S^n \to \mathbb{RP}^n$ that sends x to $[x]$. Prove that it is a local diffeomorphism.

3.5. Smooth coverings

In the smooth manifolds setting it is natural to consider topological coverings that are also compatible with the smooth structures, and these are called *smooth coverings*.

3.5.1. Definition. Let M and N be two smooth manifolds of the same dimension.

Definition 3.5.1. A *smooth covering* is a local diffeomorphism $f\colon M \to N$ between smooth manifolds that is also a topological covering.

For instance, the map $\mathbb{R} \to S^1, t \mapsto e^{it}$ is a smooth covering of infinite degree, and the map $S^n \to \mathbb{RP}^n$ of Exercise 3.4.9 is a smooth covering of degree two. To construct a local diffeomorphism that is not covering, pick any covering $M \to N$ (for instance, a diffeomorphism) and remove some random closed subset from the domain.

Exercise 3.5.2. A homeomorphism between smooth manifold that is also a local diffeomorphism is a diffeomorphism.

From this, deduce that the following definition is equivalent to the one given above: a *smooth covering* is a smooth map $f\colon M \to N$ such that every $p \in N$ has an open neighbourhood U where

$$f^{-1}(U) = \bigsqcup_{i \in I} U_i$$

and the restriction $f|_{U_i}\colon U_i \to U$ is a diffeomorphism for every $i \in I$.

3.5.2. Surfaces. As an example, one may use a bit of complex analysis to construct many non-trivial smooth coverings between open subsets of $\mathbb{C}$.

Exercise 3.5.3. Let $p(z) \in \mathbb{C}[z]$ be a complex polynomial of some degree $d \geq 1$. Consider the set $S = \{z \in \mathbb{C} \mid p'(z) = 0\}$, that has cardinality at most $d-1$. The restriction

$$p\colon \mathbb{C} \setminus p^{-1}\big(p(S)\big) \longrightarrow \mathbb{C} \setminus p(S)$$

is a smooth covering of degree d.

For instance, the map $f(z) = z^n$ is a degree-n smooth covering $f: \mathbb{C}^* \to \mathbb{C}^*$ where we indicate $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$.

3.5.3. From topological to smooth coverings. Let $\tilde{M} \to M$ be a covering of topological spaces. If M has a smooth manifold structure, we already know from Exercise 1.2.3 that $\tilde{M}$ is a topological manifold; more than that:

Proposition 3.5.4. *There is a unique smooth structure on $\tilde{M}$ such that $p: \tilde{M} \to M$ is a smooth covering.*

Proof. For every chart $\varphi: U \to V$ of M and every open subset $\tilde{U} \subset \tilde{M}$ such that $p|_{\tilde{U}}: \tilde{U} \to U$ is a homeomorphism, we assign the chart $\varphi \circ p|_{\tilde{U}}$ to $\tilde{M}$. These charts form a smooth atlas on $\tilde{M}$ and p is a smooth covering. Conversely, since p is a local diffeomorphism the smooth structure of $\tilde{M}$ is uniquely determined (exercise). □

As a consequence, much of the machinery on topological coverings summarised in Section 1.2.2 apply also to smooth coverings. For instance, if M is a connected smooth manifold, there is a bijective correspondence between the conjugacy classes of subgroups of $\pi_1(M)$ and the smooth coverings $\tilde{M} \to M$ considered up to isomorphism, where two smooth coverings $p: \tilde{M} \to M, p': \tilde{M}' \to M$ are isomorphic if there is a diffeomorphism $f: \tilde{M} \to \tilde{M}'$ such that $p = p' \circ f$.

3.5.4. Smooth actions. We keep adapting the topological definitions of Section 1.2.6 to this smooth setting. A *smooth action* of a group G on a smooth manifold M is a group homomorphism

$$G \longrightarrow \mathrm{Diffeo}(M)$$

where $\mathrm{Diffeo}(M)$ is the group of all the self-diffeomorphisms $M \to M$. All the results stated there apply to this smooth setting. In particular we have the following.

Proposition 3.5.5. *Let G act smoothly, freely, and properly discontinuously on a smooth manifold M. The quotient M/G has a unique smooth structure such that $p: M \to M/G$ is a smooth regular covering.*

Moreover, every smooth regular covering between smooth manifolds arises in this way.

Proof. We already know that p is a covering and M/G is a topological manifold. The smooth structure is constructed as follows: for every chart $U \to V$ on M such that $p|_U$ is injective, we add the chart $\varphi \circ p^{-1}: p(U) \to V$ to M/G. We get a smooth atlas on M/G because G acts smoothly. □

For instance, if M is a smooth manifold and $\iota: M \to M$ a fixed-point free involution (a diffeomorphism ι such that $\iota^2 = \mathrm{id}$ and $\iota(p) \neq p$ for all p), then

$M/\iota = M/G$ where $G = \langle \iota \rangle$ has order two is a smooth manifold and $M \to M/\iota$ a degree-two covering. This applies for instance to

$$\mathbb{RP}^n = S^n/\iota$$

where ι is the antipodal map. Every degree-two covering in fact arises in this way, because every degree-two covering is regular (since every index-two subgroup is normal).

3.5.5. The n-dimensional torus. Here is one example. Let $G = \mathbb{Z}^n$ act on $\mathbb{R}^n$ by translations, that is $g(v) = v + g$. The action is free and properly discontinuous, hence the quotient $T^n = \mathbb{R}^n/\mathbb{Z}^n$ is a smooth manifold called the *n-dimensional torus*. The manifold is in fact diffeomorphic to the product

$$\underbrace{S^1 \times \cdots \times S^1}_{n}$$

via the map

$$f(x^1, \ldots, x^n) = \left(e^{2\pi x^1 i}, \ldots, e^{2\pi x^n i}\right).$$

The map f is defined on $\mathbb{R}^n$ but it descends to the quotient T^n, and is invertible there. The n-torus T^n is compact and its fundamental group is $\mathbb{Z}^n$.

3.5.6. Lens spaces. Let $p \geqslant 1$ and $q \geqslant 1$ be two coprime integers and define the complex number $\omega = e^{2\pi i/p}$. We identify $\mathbb{R}^4$ with $\mathbb{C}^2$ and see the three-dimensional sphere S^3 as

$$S^3 = \left\{(z, w) \in \mathbb{C}^2 \;\middle|\; |z|^2 + |w|^2 = 1\right\}.$$

The map

$$f(z, w) = (\omega z, \omega^q w)$$

is a linear isomorphism of $\mathbb{C}^2$ that consists geometrically of two simultaneous order-p rotations on the real planes $w = 0$ and $z = 0$. The map f preserves S^3, it has order p and none of its iterates $f, f^2, \ldots, f^{p-1}$ has a fixed point in S^3. Therefore the group $G = \langle f \rangle$ generated by f acts freely on S^3, and also properly discontinuously because it is finite. The quotient

$$L(p, q) = S^3/G$$

is therefore a smooth manifold covered by S^3 called *lens space*. Its fundamental group is isomorphic to the cyclic group $G \cong \mathbb{Z}/p\mathbb{Z}$. Note that the manifold depends on both p and q.

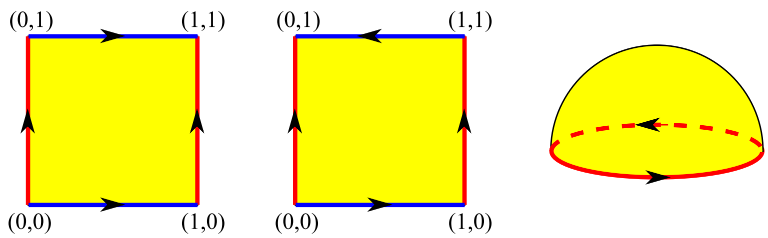

Figure 3.8. Some fundamental domains for the torus, the Klein bottle, and the projective plane. The surface is obtained from the domain by identifying the boundary curves with the same colours, respecting arrows.

3.5.7. The Klein bottle. Let G be the group of affine isometries of $\mathbb{R}^2$ generated by the maps

$$f(x, y) = (x + 1, y), \qquad g(x, y) = \big(1 - x, y + 1\big).$$

The first map is a horizontal translation, the second is a glide reflection with axis $x = 1/2$. We note that $g^{\mp 1} f g^{\pm 1} = f^{-1}$ and hence $f g^{\pm 1} = g^{\pm 1} f^{-1}$ and $g^{\mp 1} f = f^{-1} g^{\mp 1}$. This implies easily that every element of G is of the form

$$f^h g^k(x, y) = \begin{cases} (1 - x + h, y + k) & \text{if } k \text{ is odd,} \\ (x + h, y + k) & \text{if } k \text{ is even.} \end{cases}$$

The group G acts freely and properly discontinuously. The quotient surface $\mathbb{R}^2/G$ is called the *Klein bottle*.

3.5.8. Fundamental domains. Let G be a group acting smoothly, freely, and properly discontinuously on a manifold M. Sometimes we can visualise the quotient manifold M/G by drawing a *fundamental domain* for the action.

A *fundamental domain* is a closed subset $D \subset M$ such that:

- every orbit intersects D in at least one point;
- every orbit intersects $\text{int}(D)$ in at most one point.

For instance, Figure 3.8 shows some fundamental domains for:

- the action of $\mathbb{Z}^2$ to $\mathbb{R}^2$ via translations, yielding the torus $T = \mathbb{R}^2/\mathbb{Z}^2$;
- the action of G on $\mathbb{R}^2$ yielding the *Klein bottle* $K = \mathbb{R}^2/G$;
- the action of the antipodal map ι on S^2 yielding $\mathbb{RP}^2 = S^2/\iota$.

The topology of the manifold M/G can be obtained directly from D by identifying the points that lie in the same orbit. In Figure 3.8 this consists of identifying the boundary sides or curves with the same colours as suggested by the arrows.

3.6. Orientation

Some (but not all) manifolds can be equipped with an additional structure called an *orientation*. An orientation is a way of distinguishing your left hand from your right hand, through a fixed convention that holds coherently in the whole universe you are living in.

3.6.1. Oriented manifolds. Let M be a smooth manifold. We say that a compatible atlas on M is *oriented* if all the transition functions φ_{ij} have orientation-preserving differentials. That is, for every p in the domain of φ_{ij} the differential $d(\varphi_{ij})_p$ has positive determinant, for all i, j. Note that this determinant varies smoothly on p and never vanishes because φ_{ij} is a diffeomorphism: hence if the domain is connected and the determinant is positive at one point p, it is so at every point of the domain by continuity.

Definition 3.6.1. An *orientation* on M is an equivalence class of oriented atlases (compatible with the smooth structure of M), where two oriented atlases are considered as equivalent if their union is also oriented.

There are two important issues about orientations: the first is that a manifold M may have no orientation at all (see Exercise 3.6.7 below), and the second is that an orientation for M is never unique, as the following shows.

Exercise 3.6.2. If $\mathcal{A} = \{\varphi_i\}$ is an oriented atlas for M, then $\mathcal{A}' = \{r \circ \varphi_i\}$ is also an oriented atlas, where $r\colon \mathbb{R}^n \to \mathbb{R}^n$ is a fixed reflection along some hyperplane $H \subset \mathbb{R}^n$. The two oriented atlases are not orientably compatible.

We say that the orientations on M induced by $\mathcal{A}$ and $\mathcal{A}'$ are *opposite*. If M admits some orientation, we say that M is *orientable*.

Exercise 3.6.3. The sphere S^n is orientable.

Exercise 3.6.4. If M and N are orientable, then $M \times N$ also is.

3.6.2. Tangent spaces. We now exhibit an equivalent definition of orientation that involves tangent spaces. Recall the notion of orientation for vector spaces from Section 2.5.1.

Let M be a smooth manifold. Suppose that we assign an orientation to the vector space T_pM for every $p \in M$. We say that this assignment is *locally coherent* if the following holds: for every $p \in M$ there is a chart $\varphi\colon U \to V$ with $p \in U$ whose differential $d\varphi_q\colon T_qM \to T_{\varphi(q)}\mathbb{R}^n = \mathbb{R}^n$ is orientation-preserving (that is, it sends a positive basis of T_qM to a positive one of $\mathbb{R}^n$), for all $q \in U$.

Here is a new definition of orientation on M.

Definition 3.6.5. An *orientation* for M is a locally coherent assignment of orientations on all the tangent spaces T_pM.

We have two distinct notions of orientation on M, and we now show that they are equivalent. We see immediately how to pass from the first to the second: for every $p \in M$ there is some chart $\varphi\colon U \to V$ in the oriented atlas with $p \in U$ and we assign an orientation to T_pM by saying that a basis in

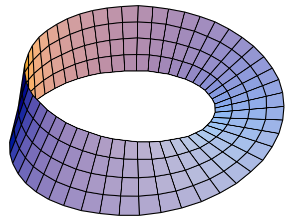

Figure 3.9. The Möbius strip is a non-orientable surface.

T_pM is positive $\Longleftrightarrow$ its image in $\mathbb{R}^n$ along $d\varphi_p$ is. The orientation of T_pM is well-defined because it is chart-independent: every other chart of the oriented atlas differs by composition with a φ_{ij} with positive differentials. We leave to the reader as an exercise to discover how to go back from the second definition to the first.

Proposition 3.6.6. *A connected smooth manifold M has either two orientations or none.*

Proof. Let $\mathcal{A}$ be an oriented atlas, and $\mathcal{A}'$ its opposite. Suppose that we have a third oriented atlas $\mathcal{A}''$. We get a partition $M = S \sqcup S'$ where S (S') is the set of points $p \in M$ where the orientation induced by $\mathcal{A}''$ on T_pM coincides with that of $\mathcal{A}$ ($\mathcal{A}'$). Both sets S, S' are open, so either $M = S$ or $M = S'$, and hence $\mathcal{A}''$ is compatible with either $\mathcal{A}$ or $\mathcal{A}'$. $\square$

Exercise 3.6.7. The *Möbius strip* shown in Figure 3.9 is non-orientable. (A rigorous definition and proof will be exhibited soon, but it is instructive to guess why that surface is not orientable only by looking at the picture.)

3.6.3. Orientation-preserving maps. Let $f: M \to N$ be a local diffeomorphism between two oriented manifolds M and N. We say that f is *orientation-preserving* if the differential $df_p: T_pM \to T_{f(p)}N$ is an orientation-preserving isomorphism for every $p \in M$. That is, we mean that it sends positive bases to positive bases. Analogously, the map f is *orientation-reversing* if df_p is so for every $p \in M$, that is it sends positive bases to negative bases.

Exercise 3.6.8. If M is connected, every local diffeomorphism $f: M \to N$ between oriented manifolds is either orientation-preserving or reversing.

As a consequence, if M is connected, to understand whether $f: M \to N$ is orientation-preserving or reversing it suffices to examine df_p at any single point $p \in M$.

Exercise 3.6.9. The orthogonal reflection π along a linear hyperplane $H \subset \mathbb{R}^{n+1}$ restricts to an orientation-reversing diffeomorphism of S^n

Hint. Suppose $H = \{x^1 = 0\}$, pick $p = (0, \dots, 0, 1)$, examine $d\pi_p$. $\square$

Corollary 3.6.10. *The antipodal map $\iota\colon S^n \to S^n$ is orientation-preserving $\iff$ n is odd.*

Proof. The map ι is a composition of $n+1$ reflections along the coordinate hyperplanes. $\square$

Let M be connected and oriented and $f\colon M \to M$ be a diffeomorphism. The condition of f being orientation-preserving or reversing is independent of the chosen orientation for M (exercise). A manifold M that admits an orientation-*reversing* diffeomorphism $M \to M$ is called *mirrorable*. For instance, the sphere S^n is mirrorable. Not all the orientable manifolds are mirrorable! This phenomenon is sometimes called *chirality*.

3.6.4. Orientability of projective spaces. We now determine whether $\mathbb{RP}^n$ is orientable or not, as a corollary of the following general fact.

Proposition 3.6.11. *Let $\pi\colon \tilde M \to M$ be a regular smooth covering of manifolds. The manifold M is orientable $\iff$ $\tilde M$ is orientable and all the deck transformations are orientation-preserving.*

Proof. If M is orientable, there is a locally coherent way to orient all the tangent spaces T_pM, which lifts to a locally coherent orientation of the tangent spaces $T_{\tilde p}\tilde M$, by requiring $d\pi_{\tilde p}$ to be orientation-preserving $\forall\, \tilde p \in \tilde M$. Every deck transformation τ turns out to be orientation preserving because $\pi\circ\tau = \pi$ implies $d\pi_{\tau(\tilde p)} \circ d\tau_{\tilde p} = d\pi_{\tilde p}$, both isomorphisms $d\pi_{\tau(\tilde p)}$ and $d\pi_{\tilde p}$ preserve orientations and hence $d\tau_{\tilde p}$ also does.

Conversely, suppose that $\tilde M$ is orientable and all the deck transformations are orientation-preserving. We can assign an orientation on T_pM by requiring that $d\pi_{\tilde p}$ be orientation-preserving for some lift $\tilde p$ of p: the definition is well-posed, because if we pick another lift $\tilde p'$ there is an orientation-preserving deck transformation τ that sends $\tilde p$ to $\tilde p'$ and we get $d\pi_{\tau(\tilde p)} \circ d\tau_{\tilde p} = d\pi_{\tilde p}$. $\square$

Corollary 3.6.12. *The real projective space $\mathbb{RP}^n$ is orientable $\iff$ n is odd.*

Proof. We have $\mathbb{RP}^n = S^n/\iota$ and the deck transformation ι is orientation-preserving $\iff$ n is odd. $\square$

Exercise 3.6.13. The projective plane $\mathbb{RP}^2$ contains an open subset diffeomorphic to the Möbius strip.

On the other hand, the n-torus and the lens spaces are orientable, because they are obtained by quotienting an orientable manifold ($\mathbb{R}^n$ or S^3) via a group of orientation-preserving diffeomorphisms acting freely and properly discontinuously.

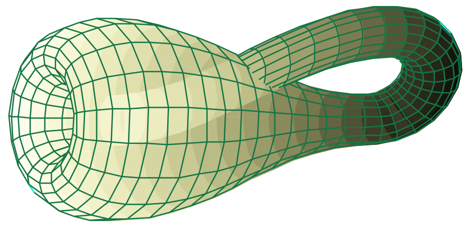

Figure 3.10. The Klein bottle immersed elegantly but non-injectively in $\mathbb{R}^3$.

3.6.5. Non-orientable surfaces. Here are two famous non-orientable surfaces. We have defined the Klein bottle in Section 3.5.7 as $\mathbb{R}^2/G$ with G generated by

$$f(x,y) = (x+1,y), \qquad g(x,y) = \big(1-x, y+1\big).$$

Since g is orientation-reversing, the Klein bottle is not orientable. The *Möbius strip* is defined analogously as $\mathbb{R}^2/\langle g\rangle$, and is also not orientable. Note that the Klein bottle is compact, while the Möbius strip is not. The Klein bottle has infinite fundamental group, so it is not homeomorphic to $\mathbb{RP}^2$. Lo dimostreremo?

As opposite to the Möbius strip, the Klein bottle cannot be embedded in $\mathbb{R}^3$, and the best that we can do is to immerse it in $\mathbb{R}^3$ non-injectively as shown in Figure 3.10. The notions of immersion and embedding will be introduced in Section 3.8.

Exercise 3.6.14. Convince yourself that by glueing the opposite sides of the central square in Figure 3.8 you get a surface homeomorphic to the one shown in Figure 3.10.

3.6.6. Orientable double cover. Non-orientable manifolds are fascinating objects, but we will see in the next chapters that it is often useful to assume that a manifold is orientable, just to make life easier. So, if you ordered an orientable manifold and you received a non-orientable one by mistake, what can you do? The best that you can do is to transform it into an orientable one by substituting it with an appropriate double cover. We now describe this operation.

We say that a manifold N is *doubly covered* by another manifold $\tilde{N}$ if there is a covering $\tilde{N} \to N$ of degree two.

Proposition 3.6.15. *Every non-orientable connected manifold M is canonically doubly covered by an orientable connected manifold $\tilde{M}$.*

Proof. We define $\tilde{M}$ as the set of all pairs (p, o) where $p \in M$ and o is an orientation for T_pM. By sending (p, o) to p we get a 2-1 map $\pi\colon \tilde{M} \to M$. We now assign to the set $\tilde{M}$ a structure of smooth connected orientable manifold and prove that π is a smooth covering.

For every chart $\varphi_i \colon U_i \to V_i$ on M we consider the set $\tilde{U}_i \subset \tilde{M}$ of all pairs (p, o) where $p \in U_i$ and o is the orientation induced by transferring back that of $\mathbb{R}^n$ via $d\varphi_p$. We also consider the map $\tilde{\varphi}_i \colon \tilde{U}_i \to V_i$, $\tilde{\varphi}_i = \varphi_i \circ \pi$. We now show that the maps

$$\tilde{\varphi}_i \colon \tilde{U}_i \longrightarrow V_i$$

constructed in this way form an oriented smooth atlas for the set $\tilde{M}$, recall the definition in Section 3.1.5.

To prove that this is an oriented smooth atlas, we first note that the sets $\tilde{U}_i$ cover $\tilde{M}$ and every $\tilde{\varphi}_i$ is a bijection. Then, we must show that for every i, j the images of $\tilde{U}_i \cap \tilde{U}_j$ along $\tilde{\varphi}_i$ and $\tilde{\varphi}_j$ are open subsets (if not empty) and the transition map $\tilde{\varphi}_{ij}$ is orientation-preservingly smooth.

We consider a point $(p, o) \in \tilde{U}_i \cap \tilde{U}_j$. The charts φ_i and φ_j both send o to the canonical orientation of $\mathbb{R}^n$, therefore the differential of the transition map φ_{ij} has positive determinant in $\varphi_i(p)$ and hence in the whole connected component W of $\varphi_i(U_i \cap U_j)$ containing $\varphi_i(p)$. This implies that $\tilde{\varphi}_i(\tilde{U}_i \cap \tilde{U}_j)$ contains the open set W. Moreover $\tilde{\varphi}_{ij}$ is orientation-preserving on W.

Now that $\tilde{M}$ is a smooth manifold, we check that π is a smooth covering: for every $p \in M$ we pick any chart $\varphi_i \colon U_i \to V_i$ with $p \in U_i$ and note that $\varphi_i' = r \circ \varphi_i$ is also a chart for any reflection r of $\mathbb{R}^n$; the two charts define two open subsets $\tilde{U}_i, \tilde{U}_i'$ of $\tilde{M}$, each projected diffeomorphically to U_i via π.

Actually, it still remains to prove that $\tilde{M}$ is connected: if it were not, it would split into two components, each diffeomorphic to M via π, but this is excluded because $\tilde{M}$ is orientable and M is not. $\square$

For instance: the Klein bottle is covered by the torus, the projective spaces are covered by spheres, and the Möbius strip is covered by $\mathbb{R} \times S^1$, with degree two in all the cases.

Corollary 3.6.16. *Every simply connected manifold is orientable.*

Proof. A simply connected manifold has no non-trivial covering! $\square$

Corollary 3.6.17. *The complex projective spaces $\mathbb{CP}^n$ are all orientable.*

Remark 3.6.18. The orientability of $\mathbb{CP}^n$ can be checked also by noting that $\mathbb{C}^n$ has a natural orientation and that the transition maps between the coordinate charts are holomorphic and hence orientation-preserving.

3.7. Submanifolds

One of the fundamental aspects of smooth manifolds is that they contain plenty of manifolds of smaller dimension, called *submanifolds*.

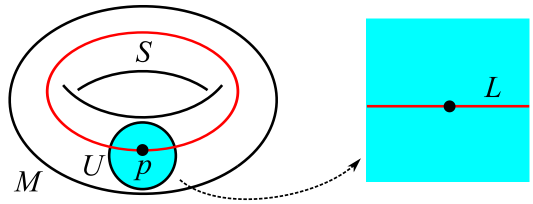

Figure 3.11. A smooth submanifold $S \subset M$ looks locally like a linear subspace $L \subset \mathbb{R}^m$.

3.7.1. Definition. Let M be a smooth m-manifold.

Definition 3.7.1. A subset $S \subset M$ is a n-dimensional *smooth submanifold* (shortly, a *n-submanifold*) if for every $p \in S$ there is a chart $\varphi\colon U \to \mathbb{R}^m$ with $p \in U$ that sends $U \cap S$ onto some linear n-subspace $L \subset \mathbb{R}^m$.

That is, the subset S looks locally like a vector n-subspace in $\mathbb{R}^m$, on some chart. Of course we must have $n \leq m$. See Figure 3.11.

A smooth n-submanifold $S \subset M$ is itself a smooth n-manifold: an atlas for S is obtained by restricting all the diffeomorphisms $U \to \mathbb{R}^m$ as above to $U \cap S$, composed with any linear isomorphism $L \to \mathbb{R}^n$. The transition maps are restrictions of smooth functions to linear subspaces and are hence smooth.

If we use the definition of tangent spaces via curves, we see immediately that for every $p \in S$ there is a canonical inclusion $i\colon T_pS \hookrightarrow T_pM$. Via derivations, the inclusion is $i(v)(f) = v(f|_S)$. We will see T_pS as a linear n-subspace of T_pM.

When $m = n$, a submanifold $N \subset M$ is just an open subset of M.

Example 3.7.2. Every linear subspace $L \subset \mathbb{R}^n$ is a submanifold.

Example 3.7.3. The graph S of a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a n-submanifold of $\mathbb{R}^n \times \mathbb{R}^m$ diffeomorphic to $\mathbb{R}^n$. The map $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ that sends (x, y) to $\big(x, y + f(x)\big)$ is a diffeomorphism that sends the linear space $L = \{y = 0\}$ to S.

As a consequence, a subset $S \subset \mathbb{R}^n$ that is locally the graph of some smooth function is a submanifold. For instance, the sphere $S^n \subset \mathbb{R}^{n+1}$ can be seen locally at every point (up to permuting the coordinates) as the graph of the smooth function $x \mapsto \sqrt{1 - \|x\|^2}$ and is hence a n-submanifold in $\mathbb{R}^{n+1}$.

If $S \subset \mathbb{R}^n$ is a k-submanifold, the tangent space T_pS at a point $p \in S$ may be represented very concretely as a k-dimensional vector subspace of $T_p\mathbb{R}^n = \mathbb{R}^n$.

Exercise 3.7.4. For every $p \in S^n$ we have

$$T_pS^n = p^\perp$$

where $p^\perp$ indicates the vector space orthogonal to p. (We will soon deduce this exercise from a general theorem.)

Example 3.7.5. A projective k-dimensional subspace S of $\mathbb{RP}^n$ or $\mathbb{CP}^n$ is the zero set of $n-k$ independent homogeneous linear equations. It is a smooth submanifold, because read on each coordinate chart it becomes a linear k-subspace in $\mathbb{R}^n$ or $\mathbb{C}^n$. It is diffeomorphic to $\mathbb{RP}^k$ or $\mathbb{CP}^k$.

Exercise 3.7.6. Let M, N be smooth manifolds. For every $p \in M$ the subset $\{p\} \times N$ is a submanifold of $M \times N$ diffeomorphic to N.

3.8. Immersions, embeddings, and submersions

We now study some particular kinds of nice maps called *immersions*, *embeddings*, and *submersions*.

3.8.1. Immersions. A smooth map $f: M \to N$ between smooth manifolds of dimension m and n is an *immersion* at a point $p \in M$ if the differential

$$df_p\colon T_pM \longrightarrow T_{f(p)}N$$

is injective. This implies in particular that $m \leq n$.

It is a remarkable fact that every immersion may be described locally in a very simple form, on appropriate charts. This is the content of the following proposition.

Proposition 3.8.1. *Let $f: M \to N$ be an immersion at $p \in M$. There are charts $\varphi\colon U \to \mathbb{R}^m$ and $\psi\colon W \to \mathbb{R}^n$ with $p \in U \subset M$ and $f(U) \subset W \subset N$ such that $\psi \circ f \circ \varphi^{-1}(x^1, \ldots, x^m) = (x^1, \ldots, x^m, 0, \ldots, 0)$.*

The proposition can be memorised via the following commutative diagram:

$$\begin{tikzcd} U \arrow[r, "f"] \arrow[d, "\varphi"'] & W \arrow[d, "\psi"] \\ \mathbb{R}^m \arrow[r, "F"'] & \mathbb{R}^n \end{tikzcd} \tag{6}$$

where $F(x^1, \ldots, x^m) = (x^1, \ldots, x^m, 0, \ldots, 0)$. Read on some charts, every immersion looks like F.

Proof. We can replace M and N with any open neighbourhoods of p and $f(p)$, in particular by taking charts we may suppose that $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ are some open subsets.

We know that $df_p\colon \mathbb{R}^m \to \mathbb{R}^n$ is injective. Therefore its image L has dimension m. Choose an injective linear map $g\colon \mathbb{R}^{n-m} \to \mathbb{R}^n$ whose image is in direct sum with L and define

$$G\colon M \times \mathbb{R}^{n-m} \longrightarrow \mathbb{R}^n$$

Figure 3.12. A non-injective immersion $S^1 \to \mathbb{R}^2$ (left) and an injective immersion $\mathbb{R} \to \mathbb{R}^2$ that is not an embedding (right).

by setting $G(x, y) = f(x) + g(y)$. Its differential at $(p, 0)$ is $dG_{(p,0)} = (df_p, g)$ and it is an isomorphism. By the Implicit Function Theorem the map G is a local diffeomorphism at $(p, 0)$. Therefore there are open neighbourhoods U_1, U_2, W of $p, 0, f(p)$ such that

$$G|_{U_1 \times U_2} \colon U_1 \times U_2 \to W$$

is a diffeomorphism, and we call ψ its inverse. Now for every $x \in U_1$ we get

$$\psi\big(f(x)\big) = \psi\big(G(x, 0)\big) = (x, 0).$$

Therefore we get the commutative diagram

$$\begin{tikzcd}
U_1 \arrow[r, "f"] \arrow[d, equal] & W \arrow[d, "\psi"] \\
U_1 \arrow[r, "F"'] & U_1 \times U_2
\end{tikzcd}$$

with $F(x) = (x, 0)$ as required. To conclude, we may take neighbourhoods U_1, U_2 diffeomorphic to $\mathbb{R}^m$, $\mathbb{R}^{n-m}$ and the diagram transforms into (6). $\square$

A map $f\colon M \to N$ is an *immersion* if it is so at every $p \in M$. An immersion is locally injective because of Proposition 3.8.1, but it may not be so globally: see for instance Figure 3.12-(left).

3.8.2. Embeddings. We have discovered that an immersion has a particularly nice local behaviour. We now introduce some special type of immersions that also behave nicely globally.

Definition 3.8.2. A smooth map $f\colon M \to N$ is an *embedding* if it is an immersion and a homeomorphism onto its image.

The latter condition means that $f\colon M \to f(M)$ is a homeomorphism, so in particular f is injective. We note that f may be an injective immersion while not being a homeomorphism onto its image! An example is shown in Figure 3.12-(right). We really need the “homeomorphism onto its image” condition here, injectivity is not enough for our purposes.

The importance of embeddings relies in the following.

Proposition 3.8.3. *If $f\colon M \to N$ is an embedding, then $f(M) \subset N$ is a smooth submanifold and $f : M \to f(M)$ a diffeomorphism.*

Proof. For every $p \in M$ there are open neighbourhoods $U \subset M$, $V \subset N$ of p, $f(p)$ such that $f|_U \colon U \to V \cap f(M)$ is a homeomorphism.

By Proposition 3.8.1, after taking a smaller V there is a chart that sends $\big(V, V \cap f(M)\big)$ to $(\mathbb{R}^n, L)$ for some linear subspace L. Therefore $f(M)$ is a smooth submanifold, and f is a diffeomorphism onto $f(M)$. $\square$

Figure 3.12-(right) shows that the image of an injective immersion needs not to be a submanifold. Conversely:

Exercise 3.8.4. If $S \subset N$ is a smooth submanifold, then the inclusion map $i \colon S \hookrightarrow N$ is an embedding.

We now look for a simple embedding criterion. Recall that a map $f \colon X \to Y$ is *proper* if $C \subset Y$ compact implies $f^{-1}(C) \subset X$ compact.

Exercise 3.8.5. A proper injective immersion $f \colon M \to N$ is an embedding.

In particular, if M is compact then f is certainly proper, and we can conclude that every injective immersion of M is an embedding. This is certainly a fairly simple embedding criterion.

Example 3.8.6. Fix two positive numbers $0 < a < b$ and consider the map $f \colon S^1 \times S^1 \to \mathbb{R}^3$ given by

$$f(e^{i\theta}, e^{i\varphi}) = \big((a\cos\theta + b)\cos\varphi, (a\cos\theta + b)\sin\varphi, a\sin\theta\big).$$

Using the coordinates θ and φ, the differential is

$$\begin{pmatrix} -a\sin\theta\cos\varphi & -(a\cos\theta + b)\sin\varphi \\ -a\sin\theta\sin\varphi & (a\cos\theta + b)\cos\varphi \\ a\cos\theta & 0 \end{pmatrix}$$

and it has rank two for all θ, φ. Therefore f is an injective immersion and hence an embedding since $S^1 \times S^1$ is compact. The image of f is the standard torus in space already shown in Figure 3.4.

Example 3.8.7. Let p, q be two coprime integers. The map $g \colon S^1 \to S^1 \times S^1$ given by

$$g(e^{i\theta}) = \big(e^{ip\theta}, e^{iq\theta}\big)$$

is injective (exercise) and its differential in the angle coordinates is $(p, q) \neq (0, 0)$. Therefore g is an embedding.

The composition $f \circ g \colon S^1 \to \mathbb{R}^3$ with the map f of Example 3.8.6 is also an embedding, and its image is called a *torus knot*: see Figure 3.13. More generally, a *knot* is an embedding $S^1 \hookrightarrow \mathbb{R}^3$.

Exercise 3.8.8. Let p, q be two real numbers with irrational ratio p/q. The map $h \colon \mathbb{R} \to S^1 \times S^1$ defined by

$$h(t) = \big(e^{ipt}, e^{iqt}\big)$$

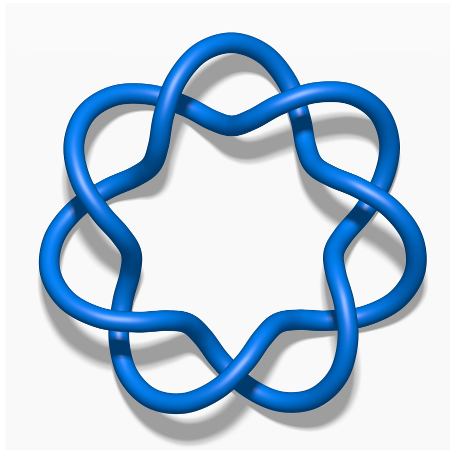

Figure 3.13. A knot is an embedding $S^1 \hookrightarrow \mathbb{R}^3$. This is a torus knot: what are the parameters p and q here?

is an injective immersion but is *not* an embedding. Its image is in fact a dense subset of the torus.

Exercise 3.8.9. If M is compact and N is connected, and $\dim M = \dim N$, every embedding $M \to N$ is a diffeomorphism.

3.8.3. Submersions. We now describe some maps that are somehow dual to immersions. A smooth map $f: M \to N$ is a *submersion* at a point $p \in M$ if the differential df_p is surjective. This implies that $m \geq n$. Again, every such map has a simple local form.

Proposition 3.8.10. *Let $f: M \to N$ be a submersion at $p \in M$. There are charts $\varphi: U \to \mathbb{R}^m$ and $\psi: W \to \mathbb{R}^n$ with $p \in U \subset M$ and $f(U) \subset W \subset N$ such that $\psi \circ f \circ \varphi^{-1}(x^1, \ldots, x^m) = (x^1, \ldots, x^n)$.*

The proposition can be memorised via the following commutative diagram:

$$\begin{tikzcd}
U \arrow[r, "f"] \arrow[d, "\varphi"'] & W \arrow[d, "\psi"] \\
\mathbb{R}^m \arrow[r, "F"'] & \mathbb{R}^n
\end{tikzcd}$$

where $F(x^1, \ldots, x^m) = (x^1, \ldots, x^n)$. Read on some charts, every submersion looks like F.

Proof. The proof is very similar to that of Proposition 3.8.1. We can replace M and N with any open neighbourhoods of p and $f(p)$, in particular by taking charts we suppose that $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ are open subsets.

We know that $df_p: T_pM \to T_{f(p)}N$ is surjective, hence its kernel K has dimension $m - n$. Choose a linear map $g: \mathbb{R}^m \to \mathbb{R}^{m-n}$ that is injective on K

and define

$$G\colon M \longrightarrow N \times \mathbb{R}^{m-n}$$

by setting $G(x) = \big(f(x), g(x)\big)$. Its differential at p is $dG_p = (df_p, g)$ and is an isomorphism. By the Implicit Function Theorem the map G is a local diffeomorphism at p.

Therefore there are open neighbourhoods U, W_1, W_2 of $p, f(p), 0$ such that $G(U) = W_1 \times W_2$ and $G|_U$ is a diffeomorphism. Now $f\big(G^{-1}(x,y)\big) = x$ and we conclude similarly as in the proof of Proposition 3.8.1. □

A smooth map $f: M \to N$ is a *submersion* if it is so at every $p \in M$.

3.8.4. Regular values. We have proved that the image of an embedding is a submanifold, and now we show that (somehow dually) the preimage of a submersion is also a submanifold. In fact, one does not really need the map to be a submersion: some weaker hypothesis suffices, that we now introduce.

Let $f: M \to N$ be a smooth map between manifolds of dimension $m \geq n$ respectively. A point $p \in M$ is *regular* if the differential df_p is surjective (that is if f is a submersion at p), and *critical* otherwise.

Proposition 3.8.11. *The regular points form an open subset of M.*

Proof. Read on charts, the differential df_p becomes a $n \times m$ matrix that depends smoothly on the point p. The matrices with maximum rank m form an open subset in the set of all $n \times m$ matrices. □

A point $q \in N$ is a *regular value* if the counterimage $f^{-1}(q)$ consists entirely of regular points, and it is *singular* otherwise. The map f is a submersion $\Longleftrightarrow$ all the points in the codomain are regular values.

Proposition 3.8.12. *If $q \in N$ is a regular value, then $S = f^{-1}(q)$ is either empty or a smooth $(m-n)$-submanifold. Moreover for every $p \in S$ we have*

$$T_pS = \ker df_p.$$

Proof. Thanks to Proposition 3.8.10 there are charts at p and $f(p)$ that transform f locally into a projection $\pi\colon \mathbb{R}^m \to \mathbb{R}^n$. On these charts $f^{-1}(q)$ is the linear subspace $\ker \pi$, hence a $(m-n)$-submanifold. The tangent space at p is $\ker \pi = \ker df_p$. □

Using this proposition we can re-prove that the sphere S^n is a submanifold of $\mathbb{R}^{n+1}$: pick the smooth map $f(x) = \|x\|^2$ and note that $S^n = f^{-1}(1)$. The gradient df_x is $(2x^1, \ldots, 2x^n)$, hence every non-zero point $x \in \mathbb{R}^{n+1}$ is regular for f, and therefore every non-zero point $y \in \mathbb{R}$ is a regular value: in particular 1 is regular and the proposition applies.

We can also deduce Exercise 3.7.4 quite easily: for every $x \in S^n$ we get

$$T_xS^n = \ker df_x = \ker(2x^1, \ldots, 2x^n) = x^\perp.$$

3.9. Examples

Some familiar spaces are actually smooth manifolds in a natural way. We list some of them and state a few results that will be useful in the sequel.

3.9.1. Matrix spaces. Every finite-dimensional real vector space V is naturally a smooth manifold diffeomorphic to $\mathbb{R}^n$: as a smooth atlas, pick all the isomorphisms $V \to \mathbb{R}^n$. Since V is also a vector space, the tangent space at every point $p \in V$ is naturally identified with V itself.

The vector space $M(m,n)$ of all $m \times n$ matrices is hence diffeomorphic to $\mathbb{R}^{mn}$. The subset consisting of all the matrices with maximal rank is open, and is hence also a smooth manifold.

In particular, the set $M(n)$ of all the square $n \times n$ matrices is a smooth manifold, and the open subset $\mathrm{GL}(n,\mathbb{R})$ of all the invertible $n \times n$ matrices is a smooth manifold, both of dimension n^2. For every $A \in M(n)$ we identify $T_AM(n) = M(n)$, and also $T_A\mathrm{GL}(n,\mathbb{R}) = M(n)$ for every $A \in \mathrm{GL}(n,\mathbb{R})$.

The subspaces $S(n)$ and $A(n)$ of all the symmetric and antisymmetric matrices are submanifolds of dimension $(n+1)n/2$ and $(n-1)n/2$ respectively.

A less trivial example is the set of $n \times n$ matrices with unit determinant:

$$\mathrm{SL}(n,\mathbb{R}) = \big\{A \in M(n) \bigm| \det A = 1\big\}.$$

Proposition 3.9.1. *The set* $\mathrm{SL}(n,\mathbb{R})$ *is a submanifold of* $M(n)$ *of codimension 1. We have*

$$T_I\mathrm{SL}(n,\mathbb{R}) = \big\{A \in M(n) \bigm| \mathrm{tr}A = 0\big\}.$$

Proof. The determinant is a smooth map $\det\colon M(n) \to \mathbb{R}$. We show that $1 \in \mathbb{R}$ is a regular value. For every $A \in \mathrm{SL}(n,\mathbb{R})$ and $B \in M(n)$ we easily get

$$\det(A + tB) = \det(I + tBA^{-1}) = 1 + t\mathrm{tr}(BA^{-1}) + o(t).$$

Therefore $d\det_A(B) = \mathrm{tr}(BA^{-1})$ and by taking $B = A$ we deduce that $d\det_A$ is surjective. Hence 1 is a regular value, so by Proposition 3.8.12 the preimage $\mathrm{SL}(n,\mathbb{R})$ is a smooth submanifold and $T_I\mathrm{SL}(n,\mathbb{R}) = \ker d\det_I$ is as stated. $\square$

3.9.2. Orthogonal matrices. Another important example is the set of all the orthogonal matrices

$$\mathrm{O}(n) = \big\{A \in M(n) \bigm| {}^{\mathrm{t}}AA = I\big\}.$$

Proposition 3.9.2. *The set* $\mathrm{O}(n)$ *is a submanifold of* $M(n)$ *of dimension* $(n-1)n/2$. *We have*

$$T_I\mathrm{O}(n) = A(n).$$

Proof. Consider the smooth map

$$\begin{aligned} f\colon\quad M(n) &\longrightarrow S(n), \\ A &\longmapsto {}^{\mathrm{t}}AA. \end{aligned}$$

Note that $O(n) = f^{-1}(I)$. We now show that $I \in S(n)$ is a regular value. For every $A \in O(n)$ we have

$$\begin{aligned} f(A+tB) &= {}^t(A+tB)(A+tB) = {}^tAA + t({}^tBA + {}^tAB) + t^2\,{}^tBB \\ &= I + t({}^tBA + {}^tAB) + o(t). \end{aligned}$$

and hence

$$df_A(B) = {}^tBA + {}^tAB.$$

For every symmetric matrix $S \in S(n)$ there is a B such that ${}^tBA + {}^tAB = S$ (exercise). Then df_A is surjective for all $A \in O(n)$ and I is a regular value.

We deduce from Proposition 3.8.12 that $O(n) = f^{-1}(I)$ is a smooth manifold of dimension $\dim M(n) - \dim S(n) = (n-1)n/2$. Moreover

$$T_I O(n) = \ker df_I = \{B \mid {}^tB + B = 0\} = A(n).$$

The proof is complete. $\square$

3.9.3. Fixed rank. We now exhibit some natural submanifolds in the space $M(m,n)$ of all $m \times n$ matrices. For every $0 \leq k \leq \min\{m,n\}$, we define $M_k(m,n) \subset M(m,n)$ to be the subset consisting of all the matrices having rank k.

Proposition 3.9.3. *The subspace $M_k(m,n)$ is a submanifold in $M(m,n)$ of codimension $(m-k)(n-k)$.*

Proof. Consider a matrix $P_0 \in M_k(m,n)$. Up to permuting rows and columns, we may suppose that $P_0 = \left(\begin{smallmatrix} A_0 & B_0 \\ C_0 & D_0 \end{smallmatrix}\right)$ where $A_0 \in \mathrm{GL}(k,\mathbb{R})$.

On an open neighbourhood of P_0 every matrix P is also of this type $P = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ with $A \in \mathrm{GL}(k,\mathbb{R})$ and if we set $Q = \left(\begin{smallmatrix} A^{-1} & -A^{-1}B \\ 0 & I_{n-k} \end{smallmatrix}\right) \in \mathrm{GL}(n,\mathbb{R})$ we find

$$PQ = \begin{pmatrix} I_k & 0 \\ CA^{-1} & D - CA^{-1}B \end{pmatrix}.$$

Since $\mathrm{rk}P = \mathrm{rk}PQ$, we deduce that

$$\mathrm{rk}P = k \iff D = CA^{-1}B.$$

Therefore $M_k(m,n)$ is a manifold parametrised locally by (A,B,C), of codimension $(m-k)(n-k)$. $\square$

3.9.4. Square roots. Let $S^+(n) \subset S(n)$ be the open subset of all positive-definite symmetric matrices. We will neeed the following.

Proposition 3.9.4. *Every $S \in S^+(n)$ has a unique square root $\sqrt{S} \in S^+(n)$, that depends smoothly on S.*

Proof. The existence and uniqueness of $\sqrt{S}$ are consequences of the spectral theorem. Smoothness may be proved by showing that the map $f\colon S^+(n) \to S^+(n)$, $A \mapsto A^2$ is a submersion: being a 1-1 correspondence, it is then a diffeomorphism.

To show that f is a submersion, up to conjugacy we may suppose that D is diagonal, and write

$$f(D+tM) = (D+tM)^2 = D^2 + t(DM+MD) + o(t).$$

We have

$$(DM+MD)_{ij} = D_{ii}M_{ij} + M_{ij}D_{jj} = (D_{ii}+D_{jj})M_{ij}.$$

Since $D_{ii} > 0$ for all i, if $M \neq 0$ then $DM + MD \neq 0$, so df_D is injective and hence invertible. $\square$

3.9.5. Some matrix decompositions. It is often useful to decompose a matrix into a product of matrices of some special types. Let $T(n)$ be the set of all upper triangular matrices with positive entries on the diagonal.

Proposition 3.9.5. *For every $A \in \mathrm{GL}(n,\mathbb{R})$ there are unique $O \in \mathrm{O}(n)$ and $T \in T(n)$ such that $A = OT$. Both O and T depend smoothly on A.*

Proof. Write $A = (v^1 \ldots v^n)$ and orthonormalise its columns via the Gram–Schmidt algorithm to get $O = (w^1 \ldots w^n)$. The algorithm may in fact be interpreted as a right multiplication by some triangular T. Conversely, if $A = OT$ then O is uniquely determined: the vector w^{i+1} must be the unit vector orthogonal to $\mathrm{Span}(v^1, \ldots, v^i)$ lying on the same side as v^{i+1}. $\square$

Corollary 3.9.6. *We have the diffeomorphisms*

$$\mathrm{GL}(n,\mathbb{R}) \cong \mathrm{O}(n) \times T(n) \cong \mathrm{O}(n) \times \mathbb{R}^{n(n+1)/2}.$$

In particular there is a smooth strong deformation retraction of $\mathrm{GL}(n,\mathbb{R})$ onto the compact subset $\mathrm{O}(n)$. We also deduce a similar result for $\mathrm{SL}(n,\mathbb{R})$. Let $ST(n) \subset T(n)$ be the submanifold of all upper triangular matrices with positive entries on the diagonal and unit determinant.

Corollary 3.9.7. *We have the diffeomorphisms*

$$\mathrm{SL}(n,\mathbb{R}) \cong \mathrm{SO}(n) \times ST(n) \cong \mathrm{SO}(n) \times \mathbb{R}^{n(n+1)/2-1}.$$

The decomposition $M = OT$ is nice, but we will later need one that is "more invariant".

Proposition 3.9.8. *For every $A \in \mathrm{GL}(n,\mathbb{R})$ there are unique $O \in \mathrm{O}(n)$ and $S \in S^+(n)$ such that $A = OS$. Both O and S depend smoothly on A.*

Proof. Pick $S = \sqrt{{}^{\mathrm{t}}\!AA}$. Write $O = AS^{-1}$ and note that O is orthogonal:

$${}^{\mathrm{t}}OO = {}^{\mathrm{t}}S^{-1}\,{}^{\mathrm{t}}\!AAS^{-1} = S^{-1}S^2S^{-1} = I.$$

Conversely, if $A = OS$ then ${}^{\mathrm{t}}\!AA = {}^{\mathrm{t}}S\,{}^{\mathrm{t}}OOS = S^2$. $\square$

The decomposition $A = OS$ is also known as the *polar decomposition* and is "more invariant" than $A = OT$ because it satisfies the following property:

Proposition 3.9.9. *If $A' = PAQ$ for some orthogonal matrices $P, Q \in \mathrm{O}(n)$, then $A' = O'S'$ with $O' = POQ$ and $S' = Q^{-1}SQ$.*

Proof. By multiplying we indeed get $A' = S'O'$. □

3.9.6. Connected components. Recall that every $A \in \mathrm{O}(n)$ has $\det A = \pm 1$. We define

$$\mathrm{SO}(n) = \big\{A \in \mathrm{O}(n) \mid \det A = 1\big\}$$

Proposition 3.9.10. *The manifold* $\mathrm{O}(n)$ *has two connected components, one of which is* $\mathrm{SO}(n)$*.*

Proof. We first prove that $\mathrm{SO}(n)$ is path-connected. Let R_θ be the θ-rotation 2×2 matrix. The real Jordan theorem implies that every real square matrix has either an invariant line or plane. By applying this fact iteratively we deduced that every $A \in \mathrm{SO}(n)$ is similar $A = M^{-1}BM$ via a matrix $M \in \mathrm{SO}(n)$ to a $B \in \mathrm{SO}(n)$ of type

$$B = \begin{pmatrix} R_{\theta_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & R_{\theta_m} \end{pmatrix} \quad \text{or} \quad B = \begin{pmatrix} R_{\theta_1} & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & R_{\theta_m} & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

depending on whether $n = 2m$ or $n = 2m + 1$, for some angles $\theta_1, \dots, \theta_m$. By sending continuously the angles to zero we get a path connecting B to I_n and by conjugating everything with M we get one connecting A to I_n.

Finally, two matrices in $\mathrm{O}(n)$ with determinant 1 and -1 cannot be path-connected because the determinant is a continuous function. □

Corollary 3.9.11. *The manifold* $\mathrm{GL}(n, \mathbb{R})$ *has two connected components, consisting of matrices with positive and negative determinant, respectively.*

Corollary 3.9.12. *The manifold* $\mathrm{SL}(n, \mathbb{R})$ *is connected.*

3.10. Homotopy and isotopy

There are plenty of smooth maps $M \to N$ between two given smooth manifolds, and in some cases it is natural to consider them up to some equivalence relation. We introduce here a quite mild relation called *smooth homotopy* and a stronger one, that works only for embeddings, called *isotopy*.

3.10.1. Smooth homotopy. We introduce the following notion.

Definition 3.10.1. A *smooth homotopy* between two given smooth maps $f, g \colon M \to N$ is a smooth map $F \colon M \times \mathbb{R} \to N$ such that $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$ for all $x \in M$.

In general topology, a homotopy is just a continuous map $F: X \times [0,1] \to Y$ where X, Y are topological spaces. In this smooth setting we must (a bit reluctantly) substitute $[0,1]$ with $\mathbb{R}$ because we need the domain to be a smooth manifold. Anyway, the behaviour of $F(x,t)$ when $t \notin [0,1]$ is of no interest for us, and we may require $F(x,\cdot)$ to be constant outside that interval:

Proposition 3.10.2. *If F is a smooth homotopy between f and g, then there is another smooth homotopy F' such that $F'(x,t)$ equals $f(x)$ for all $t \leq 0$ and $g(x)$ for all $t \geq 1$.*

Proof. Take a smooth transition function $\Psi: \mathbb{R} \to \mathbb{R}$ as in Section 1.3.6, such that $\Psi(t) = 0$ for all $t \leq 0$ and $\Psi(t) = 1$ for all $t \geq 1$. Define $F'(x,t) = F\big(x, \Psi(t)\big)$. □

Two smooth maps $f, g: M \to N$ are *smoothly homotopic* if there is a smooth homotopy between them.

Proposition 3.10.3. *Being smoothly homotopic is an equivalence relation.*

Proof. The only non-trivial part is the transitive property. Let F be a smooth homotopy between f and g, and G be a smooth homotopy between g and h. We must glue them to an isotopy H between f and g.

To do this smoothly, we first modify F and G as in the proof of Proposition 3.10.2, taking a transition function Ψ such that $\Psi(x) = 0$ for all $x \leq \frac{1}{3}$ and $\Psi(x) = 1$ for all $x \geq \frac{2}{3}$. Now $F(x,\cdot)$ and $G(x,\cdot)$ are constant outside $\left[\frac{1}{3}, \frac{2}{3}\right]$ and can be glued by writing

$$H(x,t) = \begin{cases} F(x,2t) & \text{for } t \leq \frac{1}{2}, \\ G(x,2t-1) & \text{for } t \geq \frac{1}{2}. \end{cases}$$

The map H is smooth and the proof is complete. □

Example 3.10.4. Let M be a smooth manifold. Any two maps $f, g: M \to \mathbb{R}^n$ are smoothly homotopic: indeed, every $f: M \to \mathbb{R}^n$ is smoothly homotopic to the constant map $c(x) = 0$, simply by taking

$$F(x,t) = tf(x).$$

3.10.2. Isotopy. We now introduce an enhanced version of smooth homotopy, called *isotopy*, that is nicely tailored to work with embeddings.

Definition 3.10.5. An *isotopy* between two embeddings $f, g: M \to N$ is a smooth homotopy $F: M \times \mathbb{R} \to N$ between them, such that $F_t(x) = F(x,t)$ is an embedding $F_t: M \to N$ for all $t \in [0,1]$.

We can prove as above that the isotopy between embeddings is an equivalence relation. Being isotopic is much stronger than being homotopic: for instance two embeddings $f, g: M \to \mathbb{R}^n$ are always smoothly homotopic, but they may not be isotopic in many interesting cases.

As an example, two knots $f, g\colon S^1 \hookrightarrow \mathbb{R}^3$ may not be isotopic. The *knot theory* is an area of topology that studies precisely this phenomenon: its main (and still unachieved) goal would be to classify all knots up to isotopy in a satisfactory way.

Another interesting challenge is to study the set of all self-diffeomorphisms $M \to M$ of one fixed manifold M up to isotopy. Note that if M is compact and connected, every level F_t in one such isotopy is a diffeomorphism by Exercise 3.8.9. This is already a fundamental and non-trivial problem when $M = S^n$ is a sphere; the one-dimensional case is the only one that can be solved easily:

Proposition 3.10.6. *Every self-diffeomorphism $\varphi\colon S^1 \to S^1$ is isotopic either to the identity or to a reflection $z \mapsto \bar{z}$, depending on whether φ is orientation-preserving or not.*

Proof. Suppose that $\varphi\colon S^1 \to S^1$ is orientation-preserving. We lift φ to a map $\tilde{\varphi}\colon \mathbb{R} \to \mathbb{R}$ between universal covers, and note that $\tilde{\varphi}'(x) > 0$ for all $x \in \mathbb{R}$. Consider the map

$$\tilde{F}_t(x) = t\tilde{\varphi}(x) + (1-t)x.$$

Since $\tilde{F}_t(x + 2k\pi) = \tilde{F}_t(x) + 2k\pi$ the map descends to a map $F_t\colon S^1 \to S^1$. When $t \in [0,1]$ we get $\tilde{F}_t'(x) = t\tilde{\varphi}'(x) + (1-t) > 0$, hence each F_t is an embedding. Therefore F_t is an isotopy between id and φ. $\square$

Here is another interesting question, that we will be able to solve in the positive in the next chapters.

Question 3.10.7. Let M be a connected n-manifold. Are two orientation-preserving embeddings $f, g\colon \mathbb{R}^n \hookrightarrow M$ always isotopic?

3.11. The Whitney embedding

We now show that every manifold may be embedded in some Euclidean space. This result was proved by Whitney in the 1930s.

3.11.1. Borel and zero-measure subsets. We start with some preliminaries that are of independent interest.

Let M be a smooth n-manifold. As in every topological space, a *Borel subset* of M is any subspace $S \subset M$ that can be constructed from the open sets through the operations of relative complement, countable unions and intersections.

Exercise 3.11.1. A subset $S \subset M$ is Borel $\Longleftrightarrow$ its image along any chart is a Borel subset of $\mathbb{R}^n$.

Let $S \subset M$ be a Borel set. Although there is no notion of measure for S, we may still say that S has *measure zero* if the image $\varphi(U \cap S)$ along any chart $\varphi\colon U \to V$ has measure zero, with respect to the Lebesgue measure in

$\mathbb{R}^n$. Note that any diffeomorphism sends zero-measure sets to zero-measure sets (Remark 1.3.6), so it suffices to check this for a set of charts covering S.

Proposition 3.11.2. *Let $f\colon M \to N$ be a smooth map between manifolds of dimensions m, n. If $m < n$, the image of f is a zero-measure set.*

Proof. This holds on charts by Corollary 1.3.8. □

In particular, the image of f has empty interior.

3.11.2. The compact case. We now prove that every compact manifold embeds in some Euclidean space. Not only the statement seems very strong, but its proof is actually relatively easy.

Theorem 3.11.3. *Every compact smooth manifold M embeds in some $\mathbb{R}^n$.*

Proof. Since M is compact, it has a finite adequate atlas $\{\varphi_i\colon U_i \to \mathbb{R}^m\}$ that consists of some k charts. The open subsets $V_i = \varphi_i^{-1}(B^m)$ cover M. Let $\lambda\colon \mathbb{R}^m \to \mathbb{R}$ be a bump function with $\lambda(x) = 1$ if $\|x\| \leq 1$, $0 < \lambda(x) < 1$ if $1 < \|x\| < 2$, and $\lambda(x) = 0$ if $\|x\| \geq 2$, see Section 1.3.5.

For every $i = 1, \ldots, k$ we define the smooth map $\lambda_i\colon M \to \mathbb{R}$ by setting $\lambda_i(p) = \lambda\big(\varphi_i(p)\big)$ if $p \in U_i$ and zero otherwise. Note that $\lambda_i(p) = 1$ if and only if $p \in V_i$, so the function λ_i detects whether a point belongs to V_i. Analogously we define the smooth map $\psi_i\colon M \to \mathbb{R}^m$ by setting $\psi_i(p) = \lambda_i(p)\varphi_i(p)$ when $p \in U_i$ and zero otherwise. Note that the functions ψ_i and φ_i coincide on V_i.

Let $n = k(m+1)$. We define $F\colon M \to \mathbb{R}^n$ by setting

$$F(p) = \big(\psi_1(p), \ldots, \psi_k(p), \lambda_1(p), \ldots, \lambda_k(p)\big).$$

The codomain is indeed $\mathbb{R}^m \times \ldots \times \mathbb{R}^m \times \mathbb{R} \times \ldots \times \mathbb{R} = \mathbb{R}^n$. We now show that F is an injective immersion, and hence an embedding since M is compact.

Every point $p \in M$ belongs to some V_i, and on this open set $\psi_i = \varphi_i$ is a local diffeomorphism; therefore the differential $d(\psi_i)_p$ has rank m, and hence also dF_p has the maximum rank m. We deduce that F is an immersion.

We prove that F is injective. Suppose that $F(p) = F(q)$. The point p belongs to some V_i, so $\lambda_i(p) = \lambda_i(q) = 1$, which implies that also q belongs to V_i. Now $\psi_i = \varphi_i$ is injective on V_i, and therefore $p = q$. □

We now want to improve the theorem in two directions: we remove the compactness hypothesis, and we prove that the dimension $n = 2m+1$ suffices.

3.11.3. Immersions. Let M be a manifold of dimension m, not necessarily compact. We know from Proposition 3.3.9 that every continuous map $f\colon M \to \mathbb{R}^n$ into a Euclidean space can be perturbed to a smooth map. We now show that if $n \geq 2m$ the map can be perturbed to an immersion.

Theorem 3.11.4. *Let $f\colon M \to \mathbb{R}^n$ be a continuous map, and $n \geq 2m$. For every $\varepsilon > 0$ there is an immersion $F\colon M \to \mathbb{R}^n$ with $\|F(p) - f(p)\| < \varepsilon\ \forall p \in M$.*

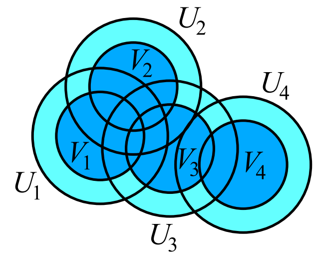

Figure 3.14. We pass from F^{i-1} to F^i by modifying the function only in U_i, with the purpose to get an immersion on $\bar{V}_i$.

Proof. By Proposition 3.3.9, we may suppose that f is smooth.

Let $\{\varphi_i : U_i \to \mathbb{R}^m\}$ be an adequate atlas, with countably many indices $i = 1, 2, \ldots$ The open subsets $V_i = \varphi_i^{-1}(B^m)$ also form a covering of M. Let $\psi_i : M \to \mathbb{R}^m$ be defined as in the proof of Theorem 3.11.3, so that $\psi_i \equiv \varphi_i$ on V_i and $\psi_i \equiv 0$ outside U_i. We set

$$M_i = \bigcup_{j=1}^{i} V_j$$

and note that $\{\bar{M}_i\}$ is a covering of M with compact subsets.

We define a sequence $F^0, F^1, \ldots$ of maps $F^i : M \to \mathbb{R}^n$ such that:

(1) $\|F^i(p) - f(p)\| < \varepsilon$ for all $p \in M$,
(2) $F^i \equiv F^{i-1}$ outside of U_i,
(3) dF^i_p is injective for all $p \in \bar{M}_i$.

See Figure 3.14. Since $\{U_i\}$ is locally finite, the maps F^i stabilise on every compact set and converge to an immersion $F : M \to \mathbb{R}^n$ as required.

We define F^i inductively on i as follows. We set $F^0 = f$ and

$$F^i = F^{i-1} + A_i \psi_i$$

for some appropriate matrix $A = A_i \in M(n, m)$ that we now choose accurately so that the conditions (1-3) will be satisfied.

We note that F^i satisfies (2). Condition (1) is also fine as long as $\|A\|$ is sufficiently small. To get (3) we need a bit of work. By the inductive hypothesis dF^{i-1}_p is injective for all $p \in \bar{M}_{i-1}$, and it will keep being so if $\|A\|$ is sufficiently small. It remains to consider the points $p \in \bar{M}_i \setminus \bar{M}_{i-1}$.

At every $p \in \bar{V}_i$ we have $\psi_i = \varphi_i$ and

$$dF^i_p = dF^{i-1}_p + A d(\varphi_i)_p.$$

Therefore dF^i_p is not injective if and only if

$$A = B - d(F^{i-1} \circ \varphi_i^{-1})_{\varphi_i(p)}$$

for some matrix $B \in M(n, m)$ of rank $k < m$.

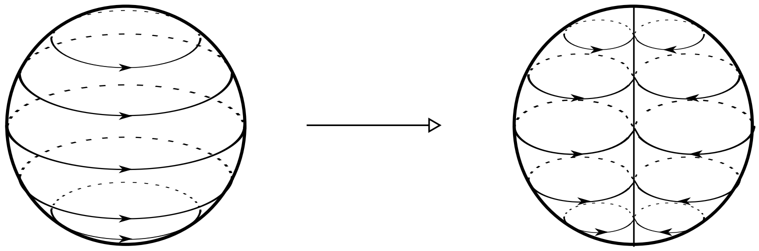

Figure 3.15. Can you perturb this continuous map $f\colon S^2 \to \mathbb{R}^3$ to an immersion? Probably not... At every horizontal level except the poles, the map is as in Figure 3.16 below. The map f is an immersion everywhere except at the poles, but it seems hard to eliminate the singular points at the poles just by perturbing f. If we are allowed to raise the dimension of the target, then f can certainly be perturbed to an immersion $S^2 \to \mathbb{R}^4$ and to an embedding $S^2 \to \mathbb{R}^5$ by Whitney's Theorems 3.11.4 and 3.11.7, although both perturbations may be hard to see...

By Proposition 3.9.3, the space $M_k(m,n)$ of all rank-k matrices is a manifold of dimension $mn-(m-k)(n-k)$. For every $k<m$ consider the map

$$\begin{aligned}\Psi\colon B^m \times M_k(n,m) &\longrightarrow M(n,m)\\ (x,B) &\longmapsto B - d(F^{i-1}\circ \varphi_i^{-1})_x.\end{aligned}$$

The dimensions of the domain and codomain are

$$m+mn-(m-k)(n-k), \qquad mn.$$

Since $n \geq 2m$ and $k \leq m-1$ we have

$$m-(m-k)(n-k) \leq m-1\cdot(n-m+1) = 2m-n-1 < 0.$$

By Proposition 3.11.2 the image of Ψ has zero measure for all k. Therefore it suffices to pick A with small $\|A\|$ and away from these zero-measure sets. $\square$

In particular, every continuous map $\mathbb{R}\to\mathbb{R}^2$ or $S^1\to\mathbb{R}^2$ can be perturbed to an immersion. If S is a surface, every continuous map $S\to\mathbb{R}^4$ can be perturbed to an immersion.

We cannot remove the condition $n\geq 2m$ in general. For instance, no map $S^1\to\mathbb{R}$ can be perturbed to an immersion, because there are no immersions $S^1\to\mathbb{R}$ at all. The dimensions $m=2$ and $n=3$ seem also problematic: as a challenging example, consider the continuous map $f\colon S^2\to\mathbb{R}^3$ drawn in Figure 3.15. Can you perturb f to an immersion?

Remark 3.11.5. The proof of Theorem 3.11.4, especially in the choice of the matrix A, suggests that any "generic" smooth perturbation of f should be an immersion. This suggestion can be made precise by endowing the space of all maps $M\to\mathbb{R}^n$ with the appropriate topology: we do not pursue this here.

Corollary 3.11.6. *Every m-manifold M immerses in $\mathbb{R}^{2m}$.*

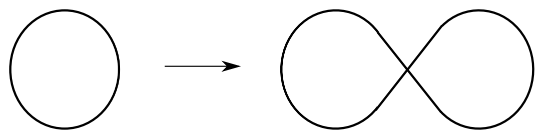

Figure 3.16. This immersion $S^1 \to \mathbb{R}^2$ cannot be perturbed to an embedding.

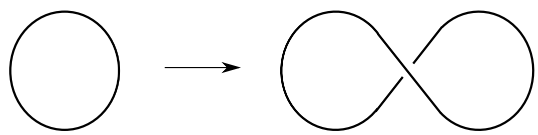

Figure 3.17. It suffices to raise the dimension of the target by one, and the immersion can now be perturbed to an injective immersion.

Proof. Pick a constant map $f : M \to \mathbb{R}^{2m}$ and apply Theorem 3.11.4. □

3.11.4. Injective immersions. Can we perturb an immersion $M^m \to \mathbb{R}^n$ to an *injective* immersion? This may not be possible in some cases, see Figure 3.16. In fact, Figure 3.17 suggests that we could achieve injectivity just by adding one dimension to the codomain: the immersion can be perturbed to be injective in $\mathbb{R}^3$, not in $\mathbb{R}^2$. We now show that this is a general principle.

Theorem 3.11.7. *Let $f : M \to \mathbb{R}^n$ be an immersion, and $n \geq 2m+1$. For every $\varepsilon > 0$ there is an injective immersion $F : M \to \mathbb{R}^n$ with $\|F(p) - f(p)\| < \varepsilon\ \forall p \in M$.*

Proof. We adapt the proof of Theorem 3.11.4 to this context. By Proposition 3.8.1 the map f is locally injective, so by Proposition 3.3.2 we can find an adequate atlas $\{\varphi_i : U_i \to \mathbb{R}^m\}$ such that $f|_{U_i}$ is injective for all i.

We define again $V_i = \varphi_i^{-1}(B^m)$ and $M_i = \cup_{j \leq i} V_j$. Let $\lambda_i : M \to \mathbb{R}$ be a bump function with $\lambda_i \equiv 1$ on V_i and $\lambda_i \equiv 0$ outside U_i.

We now construct a sequence $F^0, F^1, \ldots$ of immersions $F^i : M \to \mathbb{R}^n$, that satisfy the following conditions:

(1) $\|F^i(p) - f(p)\| < \varepsilon$ for all $p \in M$,
(2) $F^i \equiv F^{i-1}$ outside of U_i,
(3) $F^i|_{U_j}$ is injective for all j,
(4) F^i is injective on $\bar{M}_i$.

Again, we conclude that F^i converge to some F that is an injective immersion.

We set $F^0 = f$. Given F^{i-1}, we define

$$F^i = F^{i-1} + \lambda_i v_i$$

where $v = v_i \in \mathbb{R}^n$ is some vector that we now determine. If $\|v\|$ is sufficiently small, then F^i is an immersion and (1) is satisfied. Moreover (2) is automatic.

Now let $U \subset M \times M$ be the open subset

$$U = \big\{(p, q) \in M \times M \bigm| \lambda_i(p) \neq \lambda_i(q)\big\}.$$

We define $\Psi\colon U \to \mathbb{R}^n$ by setting

$$\Psi(p, q) = -\frac{F^{i-1}(p) - F^{i-1}(q)}{\lambda_i(p) - \lambda_i(q)}.$$

We deduce that $F^i(p) = F^i(q)$ if and only if one of the following holds:

(a) $(p, q) \in U$ and $v = \Psi(p, q)$, or
(b) $(p, q) \notin U$ and $F^{i-1}(p) = F^{i-1}(q)$.

Since $\dim U = 2m$, the image $\Psi(U)$ form a zero-measure subset and we may require that v be disjoint from it. This excludes (a) and therefore F^i is injective where F^{i-1} is injective: we get (3).

To show (4), suppose that $F^i(p) = F^i(q)$ for some $p, q \in \bar{M}_i$. We must have $\lambda_i(p) = \lambda_i(q)$ and $F^{i-1}(p) = F^{i-1}(q)$. If $\lambda_i(p) = 0$, then $p, q \in \bar{M}_{i-1}$ and we get $p = q$ by the induction hypothesis. If $\lambda_i(p) > 0$, then $p, q \in U_i$ and we get $p = q$ by the induction hypothesis again. □

3.11.5. Embeddings. We now want to make one step further, and promote injective immersions to embeddings. The following result is the main achievement of this section.

Theorem 3.11.8 (Whitney embedding Theorem). *For every smooth m-manifold M there is a proper embedding $M \hookrightarrow \mathbb{R}^{2m+1}$.*

Proof. Pick a smooth exhaustion $g\colon M \to \mathbb{R}_{>0}$ from Proposition 3.3.10 and consider the proper map $f\colon M \to \mathbb{R}^{2m+1}, f(p) = \big(g(p), 0, \ldots, 0\big)$. By applying Theorems 3.11.4 and 3.11.7 with any fixed $\varepsilon > 0$ we can perturb f to an injective immersion, that is easily seen to be still proper. Being proper, it is an embedding by Exercise 3.8.5. □

Concerning properness, we note the following.

Exercise 3.11.9. An embedding $i\colon M \hookrightarrow \mathbb{R}^n$ is proper $\Longleftrightarrow$ $i(M)$ is a closed subset of $\mathbb{R}^n$.

Corollary 3.11.10. *Every m-manifold M is diffeomorphic to a closed submanifold of $\mathbb{R}^{2m+1}$.*

For instance, every surface embeds properly in $\mathbb{R}^5$.

3.12. Exercises

Exercise 3.12.1. Construct two smooth atlases in $\mathbb{R}$ that are not compatible. Show that the two resulting smooth manifolds are diffeomorphic.

Remark 3.12.2. Every topological manifold of dimension $n \leq 3$ has in fact a unique (up to diffeomorphisms) smooth structure. Things become more complicated in dimension $n \geq 4$ where a given topological manifold can have no smooth structure at all, or can have many pairwise non-diffeomorphic smooth structures.

Exercise 3.12.3. Let M, N be two topological manifolds and $f: M \to N$ a local homeomorphism. Given a smooth structure on M, show that there is precisely one smooth structure on N such that f becomes a local diffeomorphism.

G

Exercise 3.12.4. Consider the group G of affine isometries of $\mathbb{R}^3$ generated by:

$$f(x,y,z) = (x+1,y,z), \qquad g(x,y,z) = (x,y+1,z),$$

$$h(x,y,z) = (-x,-y,z+1).$$

Show that G acts freely and properly discontinuously and that the 3-manifold $\mathbb{R}^3/G$ is compact and orientable, but not homeomorphic to the 3-torus $S^1 \times S^1 \times S^1$. Show that this 3-manifold is doubly covered by the 3-torus.

Exercise 3.12.5. Let G be the group of affine transformations of $\mathbb{R}^2$ generated by

$$f(x,y) = (2x, \tfrac{1}{2}y).$$

Show that G acts freely but not properly discontinuously on the manifold $M = \mathbb{R}^2 \setminus \{0\}$. Show that the resulting map $M \to M/G$ is a covering map, but the quotient M/G is not Hausdorff.

Exercise 3.12.6. Let M and N be manifolds. Show that $M \times N$ is orientable if and only if both M and N are.

Exercise 3.12.7. Let N be a manifold, $M \subset N$ a smooth submanifold, and $S \subset M$ a smooth submanifold. Show that $S \subset N$ is a smooth submanifold.

Exercise 3.12.8. Let $f: M \to N$ be a smooth map between smooth manifolds. Show that the following map is an embedding:

$$i: M \hookrightarrow M \times N, \quad p \longmapsto \big(p, f(p)\big)$$

Exercise 3.12.9. Every immersion $f: M \to N$ between manifolds of the same dimension is an open map. If M is compact and N is connected, it is a smooth covering of finite degree.

Exercise 3.12.10. Every injective immersion $f: M \to N$ between manifolds of the same dimension is an embedding. If M is compact and N is connected, it is a diffeomorphism.

Exercise 3.12.11. Prove that a submersion is an open map. Deduce that if M is compact there is no submersion $M \to \mathbb{R}^n$.

Exercise 3.12.12. Prove that the following map $f: \mathbb{RP}^2 \to \mathbb{R}^4$ is an embedding:

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

Exercise 3.12.13. Construct for all n an embedding

$$\underbrace{S^1 \times \cdots \times S^1}_{n} \hookrightarrow \mathbb{R}^{n+1}.$$

Exercise 3.12.14. Prove that the Plücker embedding defined in Section 2.6.2 is indeed an embedding.

CHAPTER 4

Bundles

We introduce here a notion that is ubiquitous in modern geometry, that of a *bundle*. We start with the more general concept of *fibre bundle*, and then we turn to *vector bundles*.

4.1. Fibre bundles

In the previous chapter we have introduced the *immersions* $M \to N$, and we have proved that they behave nicely near each point $p \in M$ of the domain. After that, we have discussed the enhanced notion of *embedding* that is also nice at every point $q \in N$ of the codomain.

Here we do a similar thing with *submersions*. These are maps that behave nicely at every point $p \in M$ of the domain, and we would like to enhance the definition of submersion by requiring it to be nice also at every point $q \in N$ of the codomain. This leads to the notion of *fibre bundle*.

4.1.1. Definition. We work as usual in the smooth manifolds context.

Definition 4.1.1. Let F be a smooth manifold. A *smooth fibre bundle* with fibre F is a smooth map

$$\pi \colon E \longrightarrow B$$

between two smooth manifolds E, B called the *total space* and the *base space*, that satisfies the following *local triviality* condition. Every $p \in B$ has an open *trivialising* neighbourhood $U \subset B$ whose counterimage $\pi^{-1}(U)$ is diffeomorphic to a product $U \times F$, via a map $\varphi \colon \pi^{-1}(U) \to U \times F$ such that the following diagram commute:

$$\begin{tikzcd} \pi^{-1}(U) \arrow[r, "\varphi"] \arrow[d, "\pi"'] & U \times F \arrow[dl, "\pi_1"] \\ U & \end{tikzcd}$$

where $\pi_1 \colon U \times F \to U$ is the projection onto the first factor.

The definition might look slightly technical, but on the contrary is indeed very natural: in a fibre bundle $E \to B$, every fibre is diffeomorphic to F, and locally the fibration looks like a product $U \times F$ projecting onto the first factor.

Example 4.1.2. The *trivial bundle* is the product $E = B \times F$, with the projection $\pi \colon E \to B$ onto the first factor.

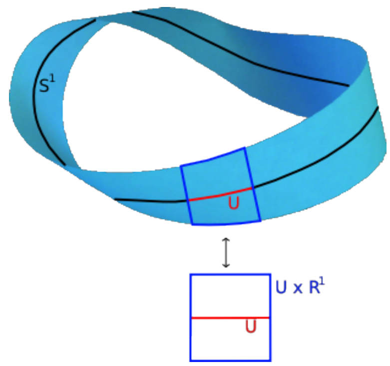

Figure 4.1. The Möbius strip is the total space of a fibre bundle with base a circle and fibre $\mathbb{R}$. Although it is locally trivial (as every fibre bundle), it is globally non-trivial: the fibre $\mathbb{R}$ makes a "twist" when transported all through the base circle.

immersion	submersion	local diffeomorphism	smooth homotopy
embedding	fibre bundle	smooth covering	isotopy

Table 4.1. We summarise here some of the most important definitions in differential topology. Every notion in the second row is an improvement of the one above.

The prototype of a non-trivial fibre bundle is the *Möbius strip* shown in Figure 4.1, which is the total space of a fibre bundle with $F = \mathbb{R}$ and $B = S^1$.

If the fibre F is diffeomorphic to the line $\mathbb{R}$, the circle S^1, the sphere S^n, the torus T, etc. we say correspondingly that E is a *line, circle, sphere*, or *torus bundle* over B. For instance, the Möbius strip is a line bundle over S^1.

Two fibre bundles $\pi\colon E \to B$ and $\pi'\colon E' \to B$ are *isomorphic* if there is a diffeomorphism $\psi\colon E \to E'$ such that $\pi = \pi' \circ \psi$. We say that a fibre bundle is *trivial* if it is isomorphic to the trivial bundle.

Remark 4.1.3. Every fibre bundle is a submersion, but not every submersion is a fibre bundle. Table 4.1 summarises some important definitions that we have introduced up to now. Recall that immersions and submersions are somehow dual notions, and every concept in the second row is an improvement of the one lying above.

Example 4.1.4. Both the torus T and the Klein bottle K are total spaces of fibre bundles over S^1 with fibre S^1. A fibration on the torus is $(e^{i\theta}, e^{i\varphi}) \mapsto e^{i\theta}$ and is clearly trivial. A fibration on the Klein bottle is suggested in Figure 4.2: this fibration is certainly not trivial, because K is not diffeomorphic to $S^1 \times S^1$.

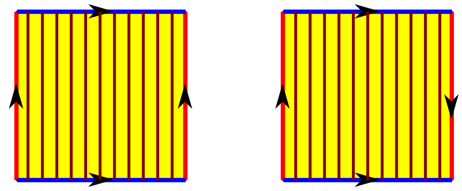

Figure 4.2. The torus and the Klein bottles are both total spaces of circle fibrations over the circle. The first is trivial, the second is not.

Note that in general two fibre bundles over the same B that are isomorphic must have diffeomorphic total spaces, but the converse is not necessarily true.

Example 4.1.5. A smooth covering between manifolds is precisely the same thing as a fibre bundle with zero-dimensional fibre F.

4.1.2. Sections. A *section* of a fibre bundle $E \to B$ is a smooth map $s\colon B \to E$ such that $\pi \circ s = \mathrm{id}_B$.

Example 4.1.6. On a trivial fibre bundle $B \times F \to B$ every map $f\colon B \to F$ determines a section $s(p) = \big(p, f(p)\big)$, and every section is obtained in this way, so sections and maps $B \to F$ are roughly the same thing.

On non-trivial bundles sections are more subtle: there are fibre bundles that have no sections at all, for instance non-trivial smooth coverings. We will often confuse a section s with its image $s(B)$; we can do this unambiguously since $s(B)$ determines s.

Exercise 4.1.7. Show that any two sections on the Möbius strip bundle intersect. This also implies that the bundle is non-trivial.

4.2. Vector bundles

A *vector bundle* is a particular fibre bundle where every fibre has a structure of finite-dimensional real vector space. This is an extremely useful concept in differential topology and geometry.

4.2.1. Definition. A *smooth vector bundle of rank k* is a smooth fibre bundle $E \to M$ with fibre $F = \mathbb{R}^k$, where the fibre $E_p = \pi^{-1}(p)$ of every point $p \in M$ has an additional structure of a real vector space of dimension k, compatible with the smooth structure in the following way: every $p \in M$ must have a trivialising open neighbourhood U such that the following diagram commutes

$$\begin{tikzcd} \pi^{-1}(U) \arrow[r, "\varphi"] \arrow[d, "\pi"'] & U \times \mathbb{R}^k \arrow[dl, "\pi_1"] \\ U & \end{tikzcd}$$

via a diffeomorphism φ that sends every fibre E_p to $\{p\} \times \mathbb{R}^k$ isomorphically as vector spaces. Note that the dimensions k and n of the fibre and of M may be arbitrary.

The simplest example of a vector bundle over M is the trivial one $M \times \mathbb{R}^k$. In general, the natural number $k > 0$ is the *rank* of the vector bundle. A vector bundle with rank $k = 1$ is called a *line bundle*. Vector bundles arise quite naturally in various contexts, as we will soon see.

Exercise 4.2.1. Recall that $\mathbb{RP}^n$ may be interpreted as the space of all the vector lines $l \subset \mathbb{R}^{n+1}$. Consider the space

$$E = \left\{ (l, v) \in \mathbb{RP}^n \times \mathbb{R}^{n+1} \;\middle|\; v \in l \right\}.$$

This is a smooth $(n+1)$-submanifold of $\mathbb{RP}^n \times \mathbb{R}^{n+1}$ and the map $\pi \colon E \to \mathbb{RP}^n$ that sends (l, v) to l is a smooth line bundle with fibre $F = \mathbb{R}$, called the *tautological line bundle*. Here $\pi^{-1}(l)$ is naturally identified to l itself and is hence a vector line.

4.2.2. Morphisms. A *morphism* between two vector bundles $E \to M$ and $E' \to M'$ is a commutative diagram

$$\begin{tikzcd} E \arrow[r, "F"] \arrow[d, "\pi"'] & E' \arrow[d, "\pi'"] \\ M \arrow[r, "f"'] & M' \end{tikzcd}$$

where F and f are smooth maps, and F is a linear map on each fibre (that is $F|_{E_p} \colon E_p \to E'_{f(p)}$ is linear for each $p \in M$).

Note that the dimensions of the manifolds M, M' and of their fibres are arbitrary, so this is a quite general notion. As usual, we say that a morphism is an isomorphism if it is invertible on both sides: this is in fact equivalent to requiring that both maps f and F be diffeomorphisms.

In some cases we might prefer to consider vector bundles on a fixed base manifold M, and in that setting it is natural to consider only morphisms where f is the identity map on M.

4.2.3. The zero-section. As opposite to more general fibre bundles, every vector bundle $E \to M$ has a canonical section $s \colon M \to E$, called the *zero-section*, defined as $s(p) = 0$ where 0 is the zero in the vector space E_p, for all $p \in M$. It is convenient to identify the image $s(M)$ of the zero-section with M itself. We will always consider the base space M embedded canonically in E through its zero-section.

4.2.4. Manipulations of vector bundles. Roughly speaking, every operation on vector spaces translates into one on vector bundles over a fixed base manifold M. For instance, given two vector bundles $E \to M$ and $E' \to M$ we may define:

- their sum $E \oplus E' \to M$,
- the dual $E^* \to M$,
- their tensor product $E \otimes E' \to M$.

To do so we simply need to perform these operations fibrewise. If E_p, E'_p are the fibres over p in E, E', then the fibre of $E \oplus E'$ is by definition $E_p \oplus E'_p$, so

$$E \oplus E' = \bigsqcup_{p \in M} E_p \oplus E'_p.$$

Of course, to complete the construction we need to build a natural smooth structure on $E \oplus E'$, and this is done as follows: if $U \times \mathbb{R}^k$ and $U \times \mathbb{R}^h$ are local trivialisations of E and E', then $U \times (\mathbb{R}^k \oplus \mathbb{R}^h)$ is a local trivialisation for $E \oplus E'$ and we equip it with the obvious product smooth structure.

The dual and tensor product bundles are defined analogously. More vector bundles may be constructed by combining these operations.

Example 4.2.2. The vector bundle $\mathrm{Hom}(E, E') \to M$ is by definition the vector bundle $E^* \otimes E' \to M$. The fiber over $p \in M$ is $\mathrm{Hom}(E_p, E'_p) = E^*_p \otimes E'_p$, see Corollary 2.1.14.

4.2.5. Subbundle and quotient bundle. The notion of vector subspace translates into that of *subbundle*. Given a vector bundle $\pi\colon E \to M$, a subset $E' \subset E$ is a *h-subbundle* if it fulfills the following requirement: every $p \in M$ has a trivialising neighbourhood $U \subset M$ with a diffeomorphism

$$\varphi\colon \pi^{-1}(U) \longrightarrow U \times \mathbb{R}^k = U \times \mathbb{R}^h \times \mathbb{R}^{k-h}$$

with $\varphi\big(E' \cap \pi^{-1}(U)\big) = U \times \mathbb{R}^h \times \{0\}$. Shortly: a subbundle $E' \subset E$ looks locally like $U \times \mathbb{R}^h \times \{0\} \subset U \times \mathbb{R}^h \times \mathbb{R}^{k-h}$ above $U \subset M$.

If follows readily from the definition that $E' \subset E$ is a submanifold and the restriction $\pi|_{E'}\colon E' \to M$ is a rank-h bundle, where the fiber E'_p at every point $p \in M$ is a h-subspace of E_p.

Example 4.2.3. The line bundle of Exercise 4.2.1 is a subbundle of the trivial bundle $\mathbb{RP}^n \times \mathbb{R}^{n+1}$ over $\mathbb{RP}^n$.

If E' is a subbundle of E, we can define the *quotient bundle* $E/E' \to M$, whose fibre over $p \in M$ is the quotient vector space E_p/E'_p. The smooth structure is obtained from the diffeomorphisms φ considered above by identifying $\mathbb{R}^k/\mathbb{R}^h$ with $\mathbb{R}^{k-h}$ in the obvious way. The resulting maps

$$\begin{tikzcd} E' \arrow[r] \arrow[d, "\pi"'] & E \arrow[r] \arrow[d, "\pi"] & E/E' \arrow[d, "\pi"] \\ M \arrow[r, "\mathrm{id}"'] & M \arrow[r, "\mathrm{id}"'] & M \end{tikzcd}$$

are bundle morphisms.

4.2.6. Restriction and pull-back. So far we have only described some manipulations of vector bundles on a fixed base manifold M. Some interesting operations arise also by varying the base manifold.

For instance we can change the base while keeping the fibres fixed: if $N \subset M$ is a submanifold, then every vector bundle $E \to M$ restricts to a vector bundle $E|_N \to N$ with the same fibres E_p in the obvious way. We call this operation the *restriction* to a submanifold. We get a bundle morphism

$$\begin{tikzcd} E|_N \arrow[r] \arrow[d, "\pi"'] & E \arrow[d, "\pi"] \\ N \arrow[r, hook] & M \end{tikzcd}$$

More generally, let $f: N \to M$ be any smooth map and $E \to M$ be a vector bundle. The *pull-back* of f is a new vector bundle $f^*E \to N$ constructed as follows: the total space is

$$f^*E = \big\{(p, v) \in N \times E \ \big|\ f(p) = \pi(v)\big\} \subset N \times E.$$

The map $\pi: f^*E \to N$ is $\pi(p, v) = p$. The fibre $(f^*E)_p$ over p is naturally identified with $E_{f(p)}$ and is hence a vector space.

Proposition 4.2.4. *The total space f^*E is a smooth submanifold of $N \times E$ and $f^*E \to N$ is a vector bundle.*

Proof. By restricting to a trivialising neighbourhood for E it suffices to consider the case where $N = \mathbb{R}^n$, $M = \mathbb{R}^m$, and $E = \mathbb{R}^m \times \mathbb{R}^k$. We get

$$f^*E = \big\{(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \ \big|\ f(x) = y\big\}.$$

Everything now follows from Example 3.7.3. $\square$

We draw the commutative diagram

$$\begin{tikzcd} f^*E \arrow[r, dotted] \arrow[d, dotted] & E \arrow[d, "\pi"] \\ N \arrow[r, "f"'] & M \end{tikzcd}$$

The dotted arrows indicate the maps that are induced by pulling-back π along f. The restriction is a particular kind of pull-back where $N \subset M$ is a submanifold and f is the inclusion map.

Exercise 4.2.5. If f is constant, then f^*E is trivial.

4.2.7. Homotopy invariance of pull-backs. The pull-back of a bundle along a map is in fact invariant up to homotopy.

Theorem 4.2.6. *The pull-backs f^*E, g^*E of a vector bundle $E \to M$ along two homotopic maps $f, g: N \to M$ are always isomorphic.*

The proof of this theorem will be quite straightforward after that we introduce connections, so we defer it to the next chapters. TBD

Corollary 4.2.7. *Every vector bundle on a contractible manifold is trivial.*

Proof. Let $E \to M$ be a vector bundle on a contractible M. The identity $\mathrm{id}\colon M \to M$ is homotopic to a constant map $c\colon M \to M$, so $E = \mathrm{id}^*E$ is isomorphic to c^*E, which is trivial by Exercise 4.2.5. □

In particular every vector bundle over $\mathbb{R}^n$ is trivial.

4.3. Tangent bundle

We now introduce the most important vector bundle on a smooth n-manifold M, the *tangent bundle*. We will also define some of its relatives, like the *cotangent*, the *normal*, and the more general *tensor bundle*.

4.3.1. Definition. Let M be a smooth manifold. As a set, the *tangent bundle* of M is the disjoint union

$$TM = \bigsqcup_{p \in M} T_pM$$

of all its tangent spaces. There is an obvious projection $\pi\colon TM \to M$ that sends T_pM to p.

The set TM has a natural structure of smooth manifold induced from that of M as follows. Every chart $\varphi\colon U \to V$ of M induces an isomorphism $d\varphi_p\colon T_pM \to \mathbb{R}^n$ for every $p \in U$. Therefore it induces an overall identification $\varphi_*\colon \pi^{-1}(U) \to V \times \mathbb{R}^n$ via

$$\varphi_*(v) = \big(\varphi(p), d\varphi_p(v)\big)$$

where $p = \pi(v)$, for every $v \in \pi^{-1}(U)$. We define an atlas on TM by taking all the charts φ_* of this type. The same charts φ_* furnish the local trivializations needed to prove that $TM \longrightarrow M$ is indeed a vector bundle.

If $\dim M = n$, then $\dim TM = 2n$. We think of M embedded in TM as the zero-section, as usual with vector bundles.

Example 4.3.1. The tangent bundle of an open subset $U \subset \mathbb{R}^n$ is canonically identified with the trivial bundle

$$TU = U \times \mathbb{R}^n$$

because every tangent space in U is canonically identified with $\mathbb{R}^n$.

More generally, we can write the tangent bundle TM of a submanifold $M \subset \mathbb{R}^n$ of any dimension $m < n$ quite explicitly:

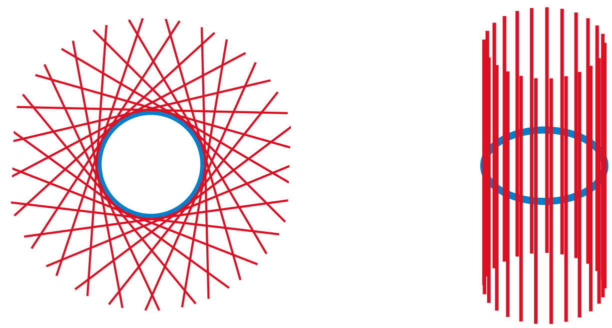

Figure 4.3. The tangent bundle of S^1 is trivial.

Example 4.3.2. The tangent bundle of a submanifold $M \subset \mathbb{R}^n$ is naturally a submanifold $TM \subset \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$, defined by

$$TM = \big\{(p, v) \bigm| p \in M, v \in T_pM\big\}.$$

For instance, we have

$$TS^n = \big\{(x, v) \bigm| \|x\| = 1, v \in x^\perp\big\}.$$

Example 4.3.3. As suggested by Figure 4.3, the tangent bundle of S^1 is trivial. A bundle isomorphism $f \colon S^1 \times \mathbb{R} \to TS^1$ is the following:

$$f(e^{i\theta}, t) = \big(e^{i\theta}, te^{i(\theta+\pi/2)}\big).$$

Is the tangent bundle of S^2 also trivial? And that of S^3?

Exercise 4.3.4. The tangent bundle TM is always an orientable manifold (even when M is not!).

Every smooth map $f \colon M \to N$ induces a morphism of tangent bundles

$$\begin{tikzcd} TM \arrow[r, "f_*"] \arrow[d, "\pi"'] & TN \arrow[d, "\pi"] \\ M \arrow[r, "f"'] & N \end{tikzcd}$$

by setting $f_*(v) = df_p(v)$ where $p = \pi(v)$ for all $v \in TM$. The restriction of f_* to each fibre T_pM is the differential $df_p \colon T_pM \to T_{f(p)}N$.

If f is a diffeomorphism, then f_* is an isomorphism.

4.3.2. Cotangent bundle. The *cotangent bundle* T^*M of a smooth manifold M is by definition the vector bundle dual to the tangent bundle TM. The fibre T_p^*M at $p \in M$ is the vector space dual to the tangent space T_pM and is called the *cotangent space* at p.

The cotangent bundle has some curious features that are lacking in the tangent bundle. One is the following: every smooth function $f\colon M \to \mathbb{R}$ induces a differential $df_p\colon T_pM \to \mathbb{R}$ at every $p \in M$, which is an element

$$df_p \in T_p^*M$$

of the *cotangent* space. We can therefore interpret the family of differentials $\{df_p\}_{p\in M}$ as a section of the cotangent bundle, and call it simply df.

We have discovered that every smooth function $f\colon M \to \mathbb{R}$ induces a section df of the cotangent bundle called its *differential*.

Remark 4.3.5. When $M = \mathbb{R}^n$, both the tangent and the cotangent space at every $p \in M$ are identified to $\mathbb{R}^n$ and the differential df is simply the gradient ∇f, that assigns a vector $(\nabla f)_p \in \mathbb{R}^n$ to every point $p \in \mathbb{R}^n$. Note however that the tangent and cotangent spaces at a point $p \in M$ are *not* canonically identified on a general smooth manifold M. A map $f\colon M \to \mathbb{R}$ induces a section of the cotangent bundle, not of the tangent bundle!

4.3.3. Normal bundle. Let M be a smooth manifold and $N \subset M$ a submanifold. We can find two natural vector bundles based on N: the tangent bundle TN and the restriction $TM|_N$ of the tangent bundle of M to N. The first is naturally a subbundle of the second, since at every $p \in N$ we have a natural inclusion $T_pN \subset T_pM$.

The *normal bundle* at N is the quotient

$$\nu N = TM|_N / TN.$$

An interesting feature of the normal bundle is that the total space νN is a manifold of the same dimension as the ambient space M. Indeed if $\dim M = m$ and $\dim N = n$ we get

$$\dim \nu N = (m - n) + n = m.$$

This preludes to an important topological application of νN called *tubular neirghbourhood* that will be revealed in the next chapters.

Example 4.3.6. On a submanifold $M \subset \mathbb{R}^n$ we may use the Euclidean scalar product to identify ν_pM with $T_pM^\perp$ for every $p \in M$. We get an orthogonal decomposition

$$T_pM \oplus \nu_pM = \mathbb{R}^n$$

for every p. Therefore we can interpret νM as a submanifold

$$\nu M = \big\{(p, v) \bigm| p \in M, v \in \nu_pM\big\} \subset \mathbb{R}^n \times \mathbb{R}^n.$$

For instance we have

$$\nu S^n = \big\{(x, v) \bigm| \|x\| = 1, v \in \mathrm{Span}(x)\big\}.$$

It is easy to deduce that the normal bundle of S^n inside $\mathbb{R}^{n+1}$ is trivial, because we may identify $S^n \times \mathbb{R}$ and νS^n by sending (x, λ) to $(x, \lambda x)$. Therefore we get a connected sum of bundles

$$TS^n \oplus \nu S^n = S^n \times \mathbb{R}^{n+1}$$

where two of them νS^n and $S^n \times \mathbb{R}^{n+1}$ are trivial, but the third one TS^n is often not trivial. It is possible to add a trivial bundle to a non-trivial one, and get a trivial bundle as a result.

4.3.4. Tensor bundle. For every $h, k \geq 0$ we may construct the *tensor bundle* $\mathcal{T}_h^k(M)$ via tensor products of the tangent and cotangent bundles:

$$\mathcal{T}_h^k(M) = \underbrace{T(M) \otimes \cdots \otimes T(M)}_{h} \otimes \underbrace{T^*(M) \otimes \cdots \otimes T^*(M)}_{k}.$$

The fiber over p is the tensor space $\mathcal{T}_h^k(T_pM)$. We define analogously the symmetric and antisymmetric tensor bundles

$$S^k(M), \qquad \Lambda^k(M)$$

as the subbundles of $\mathcal{T}^k(M)$ whose fibres over p are $S^k(T_pM)$ and $\Lambda^k(T_pM)$. In particular $\mathcal{T}_1(M)$ is the tangent bundle and $\mathcal{T}^1(M) = S^1(M) = \Lambda^1(M)$ is the cotangent bundle. We also define the trivial tensor bundle $\mathcal{T}_0^0(M) = M \times \mathbb{R}$, coherently with the fact that a tensor of type $(0, 0)$ is just a scalar in $\mathbb{R}$.

4.4. Sections

The most important feature of vector bundles is that they contain plenty of sections. Sections are not as exoteric as they might look like: in fact, many mathematical entities that will be introduced in this book – like *vector fields*, *differential forms*, and *metric tensors* – are sections in some appropriate vector bundles, so it makes perfectly sense to study them in more detail. The effort we are making now in treating these abstract objects in full generality will be soon rewarded.

4.4.1. Vector space. Let $\pi: E \to M$ be a vector bundle. The space of all sections $s: M \to E$ is usually denoted by

$$\Gamma(E).$$

This space is naturally a vector space: the sum $s + s'$ of two sections s and s' is defined by setting $(s + s')(p) = s(p) + s'(p)$ for every $p \in M$, using the vector space structure of E_p, and the product with scalars is analogous. The origin of the vector space $\Gamma(E)$ is of course the zero-section.

Moreover, for every smooth function $f: M \to \mathbb{R}$ and every section s we can define a new section fs by setting $(fs)(p) = f(p)s(p)$. Therefore $\Gamma(E)$ is also a module over the ring $C^\infty(M)$.

If E and E' are two bundles over M, with sections s and s', then one can define the sections $s \oplus s'$ and $s \otimes s'$ of $E \oplus E'$ and $E \otimes E'$ in the obvious way, by setting $(s \oplus s')(p) = \big(s(p), s'(p)\big)$ and $(s \otimes s')(p) = s(p) \otimes s'(p)$.

4.4.2. Extensions of sections. We now show that vector bundles have plenty of sections, and we do this by proving that every "locally defined" section may be extended to a global one.

Let $\pi \colon E \to M$ be a vector bundle and s be a section. On a trivialising neighbourhood U, we get a diffeomorphism $\varphi \colon \pi^{-1}(U) \to U \times \mathbb{R}^k$ and hence

$$\varphi\big(s(p)\big) = \big(p, s'(p)\big)$$

for some smooth map $s' \colon U \to \mathbb{R}^k$. In other words, every smooth section s can be read as a function $s' \colon U \to \mathbb{R}^k$ on every trivalising neighbourhood U.

The fact that sections look locally like functions has some interesting consequences: for instance, we now show that sections defined only partially may be extended globally.

Let $S \subset M$ be any subset. We say that a smooth map $s \colon S \to E$ is a *partial section* if $\pi \circ s = \mathrm{id}_S$. Recall from Definition 3.3.5 the correct meaning of "smooth" here.

Proposition 4.4.1. *If $S \subset M$ is a closed subset, every partial section $s \colon S \to E$ may be extended to a global one $M \to E$.*

Proof. We adapt the proof Proposition 3.3.6 to this context. Locally, sections are like maps $U \to \mathbb{R}^k$ and can hence be extended. Therefore for every $p \in S$ there are an open trivialising neighbourhood U_p and a local extension $g_p \colon U_p \to E$ of s. We then proceed with a partition of unity following the same proof of Proposition 3.3.6. □

Remark 4.4.2. By construction, we may suppose (if needed) that s vanishes outside of any given neighbourhood of S.

Exercise 4.4.3. Let $E \to M$ be a vector bundle of rank $k \geq 1$. If M is not a finite collection of points, the vector space $\Gamma(E)$ has infinite dimension.

4.4.3. Zeroes. Let $\pi \colon E \to M$ be a vector bundle over some smooth manifold M. We say that a section $s \colon M \to E$ *vanishes* at a point $p \in M$ if $s(p) = 0$. In that case p is called a *zero* of s. The section is *nowhere vanishing* if $s(p) \neq 0$ for all $p \in M$.

Here is one important thing to keep in mind about sections of vector bundles: although there are plenty of them, it may be hard – and sometimes impossible – to construct one that is nowhere vanishing. As an example:

Exercise 4.4.4. The Möbius strip line bundle $E \to S^1$ has no nowhere-vanishing section.

4.4.4. Frames. Let $\pi\colon E \to M$ be a rank-k vector bundle. A *frame* for π consists of k sections $s_1, \ldots, s_k$ such that the vectors $s_1(p), \ldots, s_k(p)$ are independent, and hence form a basis for E_p, for every $p \in M$.

Every s_i is in particular a nowhere-vanishing section: finding a frame is even harder than constructing a nowhere-vanishing section. In fact, the following shows that frames exist only on trivial bundles.

Proposition 4.4.5. *A bundle has a frame $\Longleftrightarrow$ the bundle is trivial.*

Proof. On a trivial bundle $E = M \times \mathbb{R}^k$, the sections $s_i(p) = (p, e_i)$ with $i = 1, \ldots, k$ form a frame. Conversely, a frame $s_1, \ldots, s_k$ on $\pi\colon E \to M$ provides a bundle isomorphism $F\colon M \times \mathbb{R}^k \to E$ by writing

$$F\big(p, (\lambda_1, \ldots, \lambda_k)\big) = \lambda_1 s_1(p) + \ldots + \lambda_k s_k(p).$$

The proof is complete. $\square$

In light of this result, a frame is also called a *trivialisation* of the bundle E because it specifies a precise isomorphism of E with the trivial bundle $M \times \mathbb{R}^k$. A nontrivial bundle $E \to M$ has no global frame, but it has many local frames: we define a *local frame* to be a frame on a trivialising open set $U \subset M$. Every trivialising open set has a local frame, induced by the trivialising chart.

4.4.5. Tensor fields. We now introduce the most important types of sections in differential topology and geometry, called *tensor fields*. These are ubiquitous in this book.

Let M be a smooth manifold. A *tensor field* of type (h, k) is a section s of the tensor bundle $\mathcal{T}_h^k(M)$ of M, that is

$$s \in \Gamma\big(\mathcal{T}_h^k(M)\big).$$

In other words, we have a tensor $s(p) \in \mathcal{T}_h^k(T_pM)$ that varies smoothly with the point $p \in M$.

Since $\mathcal{T}_0^0(M) = M \times \mathbb{R}$ is the trivial line bundle, a tensor field of type $(0, 0)$ is just a smooth function $s\colon M \to \mathbb{R}$.

A tensor field of type $(1, 0)$ assigns a tangent vector at every point and is called a *vector field*: vector fields are extremely important in differential topology and we will study them in the next chapter with some detail.

A tensor field of type $(0, 1)$ may be called a *covector field*, but the term *1-form* is more often employed. More generally, a *k-form* is a section of the antisymmetric tensor bundle $\Lambda^k(M)$. These are also important objects and we will dedicate the Chapter 7 to them.

A symmetric tensor field of type $(0, 2)$ assigns a bilinear symmetric form to every tangent space: this notion will open the doors to *differential geometry*.

Most of the operations that we defined on tensors apply naturally to tensor fields. For instance, the tensor product $s \otimes s'$ of two tensor fields s and s' of type (h, k) and (h', k') is a tensor field of type $(h + h', k + k')$, and the contraction of a tensor field of type (h, k) is a tensor field of type $(h-1, k-1)$. It suffices to apply these constructions pointwise at every $p \in M$.

4.4.6. Coordinates. Let s be a tensor field of type (h, k) on M and let $\varphi\colon U \to V$ be a chart. We now want to express s in coordinates with respect to the chart φ.

As we already noticed, for every $p \in U$ the differential $d\varphi_p$ identifies the tangent space T_pM with $\mathbb{R}^n$, and we deduce from that an identification of the tensor space $\mathcal{T}_h^k(T_pM)$ with $\mathcal{T}_h^k(\mathbb{R}^n)$. The tensor field s, restricted to U, may therefore be represented as a smooth map

$$s'\colon V \longrightarrow \mathcal{T}_h^k(\mathbb{R}^n).$$

How can we write such a map? The vector space $\mathcal{T}_h^k(\mathbb{R}^n)$ has a canonical basis that consists of the elements

$$e_{i_1} \otimes \cdots \otimes e_{i_h} \otimes e^{j_1} \otimes \cdots \otimes e^{j_k}$$

where $1 \leq i_1, \ldots, i_h, j_1, \ldots, j_k \leq n$ and $e_1, \ldots, e_n$ is the canonical basis of $\mathbb{R}^n$, see Section 2.2.2. Therefore s' may be written uniquely as

$$s'(x) = s_{j_1,\ldots,j_k}^{i_1,\ldots,i_h}(x) e_{i_1} \otimes \cdots \otimes e_{i_h} \otimes e^{j_1} \otimes \cdots \otimes e^{j_k}$$

where we employ the Einstein convention and the coefficients vary smoothly with respect to $x \in V$. Shortly, the coordinates of s with respect to φ are the coefficients

$$s_{j_1,\ldots,j_k}^{i_1,\ldots,i_h}$$

that are real numbers that depend smoothly on x.

4.4.7. Changes of coordinates. If we pick another chart around a point $p \in M$, the same tensor field s is represented via some different coordinates

$$\hat{s}_{j_1,\ldots,j_k}^{i_1,\ldots,i_h}$$

and the transformation law relating the two different coordinates is prescribed by Proposition 2.2.12. It is convenient here to denote the coordinates of the two charts by $x^1, \ldots, x^n$ and $\hat{x}^1, \ldots, \hat{x}^n$ respectively, so that the differential of the transition map may be written simply as

$$\frac{\partial \hat{x}^i}{\partial x^j}.$$

The transformation law says that

$$\hat{s}_{j_1\ldots j_k}^{i_1\ldots i_h} = \frac{\partial \hat{x}^{i_1}}{\partial x^{l_1}} \cdots \frac{\partial \hat{x}^{i_h}}{\partial x^{l_h}} \frac{\partial x^{m_1}}{\partial \hat{x}^{j_1}} \cdots \frac{\partial x^{m_k}}{\partial \hat{x}^{j_k}} s_{m_1\ldots m_k}^{l_1\ldots l_h}.$$

For instance, for a vector field we have

$$\hat{s}^i = \frac{\partial \hat{x}^i}{\partial x^j} s^j$$

while for a covector field we get

$$\hat{s}_j = \frac{\partial x^i}{\partial \hat{x}^j} s_i.$$

Note that everything is designed so that every two repeated indices stay one on the top and the other on the bottom, in every formula. This is a convention that helps us to prevent mistakes; another trick consists of replacing the notations e_i and e^j with the symbols $\frac{\partial}{\partial x^i}$ and dx^j. We will explain this in the subsequent chapters.

4.5. Riemannian metric

It is sometimes useful to equip a vector bundle with some additional structure, called *Riemannian metric*. Not only this structure is interesting in its own right, but it is also useful as an auxiliary tool.

4.5.1. Definition. Let $\pi\colon E \to M$ be a vector bundle. Consider the bundle $E^* \otimes E^* \to M$. Remember that the fibre above $p \in M$ is the space $E_p^* \otimes E_p^*$ of all tensors on E_p of type $(0,2)$. Remember also that scalar products are particular kinds of symmetric tensors of type $(0,2)$.

Definition 4.5.1. A *Riemannian metric* in π is a section g of $E^* \otimes E^*$ such that $g(p)$ is a positive-definite scalar product on E_p for every $p \in M$.

In other words, a Riemannian metric is a positive-definite scalar product $g(p)$ on each fibre E_p, that varies smoothly with p. On a trivialising chart U the bundle E looks like $U \times \mathbb{R}^k$ and g can be represented concretely as a positive-definite symmetric matrix g_{ij} smoothly varying with $p \in U$.

Proposition 4.5.2. *Every vector bundle has a Riemannian metric.*

Proof. We fix an open covering $\{U_i\}$ of trivialising sets for the bundle. Above every U_i the bundle is like $U_i \times \mathbb{R}^k$, so we can identify $E_p = \mathbb{R}^k$ for every $p \in U_i$ and assign it the Euclidean scalar product, that we name $g(p)_i$.

To patch the $g(p)_i$ altogether, we pick a partition of unity $\{\rho_i\}$ subordinate to the covering. For every $p \in M$ we define

$$g(p) = \sum_i \rho_i(p) g(p)_i.$$

This is a positive-definite scalar product, because a linear combination of positive definite scalar products with positive coefficients is always a positive-definite scalar product. $\square$

Example 4.5.3. The *Euclidean metric* on the trivial bundle $M \times \mathbb{R}^k$ is the assignment of the Euclidean scalar product on every fibre $\mathbb{R}^k$.

If $E \to M$ has a Riemannian metric, then every subbundle and every restriction to a submanifold also inherits a Riemannian metric.

4.5.2. Orthonormal frames. Let $E \to M$ be a vector bundle equipped with a Riemannian metric. An *orthonormal frame* is a frame $s_1, \ldots, s_k$ where $s_1(p), \ldots, s_k(p)$ form an orthonormal basis for every $p \in M$.

Proposition 4.5.4. *Every frame transforms canonically into an orthonormal frame via the Gram – Schmidt algorithm.*

Proof. This sentence already says everything. The Gram – Schmidt algorithm transforms $s_1(p), \ldots, s_k(p)$ into k orthonormal vectors in a way that depends smoothly on p, as one can see on a chart. $\square$

Corollary 4.5.5. *A bundle has an orthonormal frame $\Longleftrightarrow$ it is trivial.*

Proof. We already know that a bundle has a frame $\Longleftrightarrow$ it is trivial. $\square$

4.5.3. Isotopies. We will soon need an appropriate notion of isotopy between bundle isomorphisms.

Let $E \to M$ and $E' \to M$ be two vector bundles, and $f, g : E \to E'$ be two isomorphisms. An *isotopy* between f and g is a smooth map

$$F : E \times \mathbb{R} \longrightarrow E'$$

such that each $F_t = F(\cdot, t)$ is an isomorphism, and $F_0 = f$, $F_1 = g$.

4.5.4. Isometries. An *isometry* between vector bundles E, E' with Riemannian metrics g, g' is an isomorphism $F : E \to E'$ that preserves the metric, that is with $g'\big(F(v), F(w)\big) = g(v, w)$ for all $v, w \in E_p$ and all $p \in M$.

The following proposition says that, maybe a bit surprisingly, isometry between vector bundles is not a stronger relation than isomorphism. This fact extends the well-known linear algebra theorem that says that two real vector spaces equipped with positive definite scalar products are isometric if and only if they are isomorphic.

Proposition 4.5.6. *Two isomorphic vector bundles equipped with arbitrary Riemannian metrics are always isometric, via an isometry that is isotopic to the initial isomorphism.*

Proof. If the two bundles are trivial, this follows from Proposition 4.5.4, since we would find orthonormal frames on both, and an isometry would be constructed by sending the first to the second.

In general, we can use this argument only locally, and more work is needed to pass from local to global: the Gram – Schmidt process is not "invariant enough" for this purpose and we will need the "more invariant" OS decompositon of Proposition 3.9.8.

We may reduce to the case where $\pi : E \to M$ is a vector bundle and g, g' are two arbitrary Riemannian metrics on it; we must construct an isometry $E \to E$ with respect to the metrics g and g', isotopic to the identity.

Let U be a trivialising neighbourhood. Pick two orthonormal frames s_i and s_i' for g and g' on U. We may represent every isomorphism $E|_U \to E|_U$ with respect to these frames as a matrix $A(p) \in \mathrm{GL}(n, \mathbb{R})$ that depends smoothly on $p \in U$. The isomorphism is an isometry $\Longleftrightarrow A(p) \in \mathrm{O}(n)$ for every $p \in U$.

Let $A = A(p)$ represent the identity isomorphism in these basis. Use Proposition 3.9.8 to decompose A as $A = OS$ with $O \in \mathrm{O}(n)$ and $S \in S^+(n)$. The matrix $O(p)$ defines an isometry for every $p \in U$.

As a consequence of Proposition 3.9.9, we see easily that the isometry defined by $O(p)$ actually does not depend on the orthogonal frames s_i and s_i' chosen above! Therefore by covering M with charts we get a well-defined global isometry $E \to E$ with respect to the metrics g and g'.

An isotopy between O and A is $B(p) = O(p)\big(tI + (1-t)S(p)\big)$, using that $S^+(n)$ is convex. This is well defined again by Proposition 3.9.9. □

4.5.5. Unitary sphere bundle. Let $\pi\colon E \to M$ be a vector bundle. Let us equip it with a Riemannian metric g. Every fibre E_p has a positive-definite scalar product $g(p)$ and hence every vector $v \in E_p$ has a *norm*

$$\|v\| = \sqrt{g(v, v)}.$$

The associated *unitary sphere bundle* is the submanifold

$$S(E) = \big\{ v \in E \bigm| \|v\| = 1 \big\}.$$

The projection π restricts to a projection $\pi\colon S(E) \to M$ whose fibre $S(E)_p$ is the unitary sphere in E_p.

Proposition 4.5.7. *The projection $\pi\colon S(E) \to M$ is indeed a sphere bundle. It does not depend, up to isotopy, on the chosen metric g.*

By "isotopy" we mean that the sphere bundles constructed from two metrics g and g' are related by a self-isomorphism of $E \to M$ isotopic to the identity.

Proof. We prove the local triviality. On a trivialising open set U the bundle E is isometric to the Euclidean $U \times \mathbb{R}^k$, so $S(E)|_U$ is like $U \times S^{k-1}$. If we pick another metric g', we get an E' isometric to E by Proposition 4.5.6, via an isometry that is isotopic to the identity. Hence $S(E')$ is isotopic to $S(E)$. □

4.5.6. Orthogonal bundle. Let $E \to M$ be a vector bundle equipped with a Riemannian metric. For every subbundle $E' \to M$ we have an *orthogonal bundle* $(E')^\perp \to M$, whose fiber $(E')^\perp_p$ is the orthogonal subspace to $E'_p \subset E_p$ with respect to the metric.

The orthogonal bundle is canonically isomorphic to the normal bundle E/E' and may be seen as a realisation of it as a subbundle of E.

Example 4.5.8. If the tangent bundle TM of a manifold M is equipped with a Riemannian metric, the normal bundle νN of any submanifold $N \subset M$

may be seen (using the Riemannian metric) as a subbundle of $TM|_N$, so that we get an orthogonal sum

$$TM|_N = TN \oplus \nu N.$$

4.5.7. Dual vector bundle. Here is another instance where a Riemannian metric may be used as an auxiliary tool, to prove theorems.

Proposition 4.5.9. *Every vector bundle $E \to M$ is isomorphic to its dual $E^* \to M$.*

Proof. Pick a Riemannian metric on M. The scalar product on E_p may be used to identify E_p with its dual E_p^* as described in Section 2.3.3. This furnishes the bundle isomorphism $E \to E'$. □

Example 4.5.10. A Riemannian metric on the tangent bundle TM determines an identification of the tangent and the cotangent bundles over M. More generally, it furnishes some bundle isomorphisms

$$\mathcal{T}_h^k(M) \cong \mathcal{T}_{h+k}(M) \cong \mathcal{T}^{h+k}(M).$$

4.5.8. Shrinking vector bundles. A Riemannian metric may be used to *shrink* a vector bundle. We will need this technical operation at some point.

Lemma 4.5.11. *Let $E \to M$ be a vector bundle. For every neighbourhood $W \subset E$ of the zero-section M there is an embedding $g\colon E \to W$ with*

- $g|_M = \mathrm{id}_M$,
- $g(E_p) \subset E_p$ *for every* $p \in M$.

Moreover there is an isotopy g_t between $g_0 = \mathrm{id}_E$ and $g_1 = g$ through embeddings $g_t\colon E \to E$ that also fulfill these two requirements.

Proof. Fix a Riemannian metric on E. Using a partition of unity, we can prove (exercise) that there is a smooth positive function $\varepsilon\colon M \to \mathbb{R}$ such that W contains all the vectors $v \in E_p$ with $\|v\| < \varepsilon(p)$, for all $p \in M$. Define

$$g(v) = \varepsilon\big(\pi(v)\big)\frac{v}{\sqrt{1+\|v\|^2}}.$$

This map fulfills the requirements. An isotopy is obtained by convex combination $g_t(v) = (1-t)v + tg(v)$. □

4.5.9. Trivialising sums. The tangent bundle TS^n of a sphere is often non-trivial, but it suffices to add the normal bundle of S^n in $\mathbb{R}^{n+1}$ to get a trivial bundle, that is:

$$TS^n \oplus \nu S^n = S^n \times \mathbb{R}^{n+1}.$$

This is in fact an instance of a more general phenomenon:

Exercise 4.5.12. For any vector bundle $E \to M$ there is another vector bundle $E' \oplus M$ such that $E \oplus E' \to M$ is trivial.

TBD

4.6. Exercises

Exercise 4.6.1. Let S be an orientable surface. Show that the tangent bundle TS is trivial $\Longleftrightarrow$ there is a nowhere-vanishing vector field on S.

Show that this is false for the Klein bottle K: the tangent bundle TK is not trivial but K has a nowhere-vanishing section.

Exercise 4.6.2. Prove that there are precisely two vector bundles with rank 1 over S^1 up to isomorphism.

Exercise 4.6.3. Construct a fibre bundle $E \to K$ with fibre $F = S^1$ over the Klein bottle K, such that E is an orientable 3-manifold.

Hint. Use Exercise 3.12.4. □

Exercise 4.6.4. Show that every non-orientable manifold M of dimension n is contained in an orientable manifold of dimension $n + 1$.

Exercise 4.6.5. Let $\pi\colon E \to M$ be a bundle with connected fibre F. Fix any base-point $x_0 \in E$. Show that $\pi_*\colon \pi_1(E, x_0) \to \pi_1(M, \pi(x_0))$ is a surjective homomorphism. If it is a vector bundle, show that it is an isomorphism (construct a deformation retract of E onto the zero-section).

Exercise 4.6.6. The *Grassmann bundle* $\mathrm{Gr}^k(E) \to M$ of a bundle $E \to M$ is the fibre bundle whose fiber $\mathrm{Gr}^k(E)_p$ over p consists of all k-planes in E_p. Prove that $\mathrm{Gr}^k(E)$ has a natural smooth structure, and that there is a natural 1-1 correspondence between sections of $\mathrm{Gr}^k(E)$ and k-plane subbundles of E.

CHAPTER 5

The basic toolkit

We now introduce some fundamental notions that apply to every context in differential topology: we start with *vector fields*, their flows and Lie brackets; then we turn to *distributions*, *foliations*, and the Fobenius Theorem; finally, we introduce the two most important tools to understand embedded submanifolds, namely *tubular neighbourhoods* and *transversality*.

5.1. Vector fields

5.1.1. Definition. Let M be a smooth manifold. A section $X: M \to TM$ of the tangent bundle is called a *vector field*: it assigns a tangent vector $X(p) \in T_p(M)$ to every point $p \in M$ that varies smoothly with p. Remember that sections are smooth by definition, and hence vector fields also are.

Some vector fields on the torus are drawn in Figure 5.1. Recall that a zero of X is a point p such that $X(p) = 0$. Note that the vector fields sketched in the figure have no zeroes.

Example 5.1.1. When $n = 2m - 1$ is odd, the following is a nowhere-vanishing vector field on $S^n \subset \mathbb{R}^{2m}$:

$$(x^1, \ldots, x^{2m}) \longmapsto (-x^2, x^1, \ldots, -x^{2m}, x^{2m-1}).$$

Exercise 5.1.2. Write a smooth vector field on S^n that vanishes only at the poles $(\pm 1, 0, \ldots, 0)$.

We denote by $\mathfrak{X}(M)$ the set of all the vector fields on M. Recall from Section 4.4 that $\mathfrak{X}(M) = \Gamma(TM)$ is a vector space and also a $C^\infty(M)$-module.

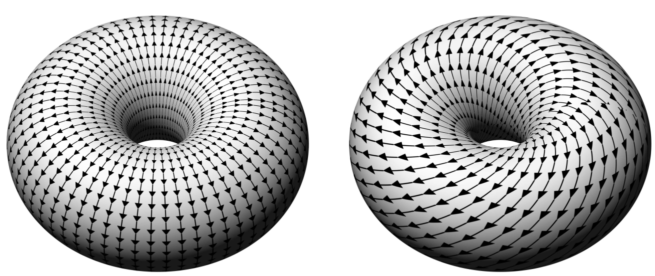

Figure 5.1. Nowhere-vanishing vector fields on the torus.

5.1.2. Diffeomorphisms. Many of the mathematical objects that we define are naturally transported along smooth maps $f: M \to N$, either from M to N or vice-versa from N to M, but this is *not* the case with vector fields: there is no meaningful way to transport a vector field along a generic map f, neither forward from M to N nor backwards from N to M.

On the other hand, every intrinsic (that is, coordinates-independent) notion can be transported in both directions if $f: M \to N$ is a diffeomorphism. If f is a diffeomorphism, every vector field X in M induces a vector field Y on N via differentials, that is by imposing:

$$Y\big(f(p)\big) = df_p\big(X(p)\big) \quad \text{for every } p \in M.$$

This gives an isomorphism between $\mathfrak{X}(M)$ and $\mathfrak{X}(N)$ induced by f.

5.1.3. On charts. If X is a vector field on M and $\varphi: U \to V \subset \mathbb{R}^n$ is a chart, we can restrict X to a vector field on U and then transport it into a vector field in V via the diffeomorphism φ. As we noticed in Section 4.4.6, the transported vector field assumes the familiar form of a smooth map $V \to \mathbb{R}^n$ because $TV = V \times \mathbb{R}^n$, and we may write it as a vector

$$\big(X^1(x), \ldots, X^n(x)\big)$$

in $\mathbb{R}^n$ that varies smoothly on $x \in V$. Here X^i is the i-coordinate of X in the chosen chart, a real number that depend smoothly on $x \in V$. We can use the Einstein notation and write the transported vector field in V more concisely as

$$X^i e_i$$

where $e_1, \ldots, e_n$ is the canonical basis of $\mathbb{R}^n$. It turns out that it is more comfortable to use the symbol $\frac{\partial}{\partial x^i}$ instead of e_i, and we write instead

$$X^i \frac{\partial}{\partial x^i}.$$

Why do we prefer the awkward notation $\frac{\partial}{\partial x^i}$ to the more familiar e_i? The partial derivative symbol is appropriate here for three reasons: (i) it is coherent with the interpretation of tangent vectors as derivations, (ii) there is no risk of confusing it with anything else, and more importantly (iii) it helps us write the coordinate changes correctly via the chain rule. Indeed, if we pick another chart we get different coordinates

$$\bar{X}^i \frac{\partial}{\partial \bar{x}^i}$$

and we know from Section 4.4.7 that the coordinates of a vector change contravariantly, hence

$$\bar{X}^j = X^i \frac{\partial \bar{x}^j}{\partial x^i}. \tag{7}$$

Thanks to the partial derivative notation, there is no need to remember the formula by heart: it suffices to apply formally the chain rule and we get

$$X^i \frac{\partial}{\partial x^i} = X^i \frac{\partial \bar{x}^j}{\partial x^i} \frac{\partial}{\partial \bar{x}^j}.$$

This gives (7). Beware that one possible source of confusion is that the coordinates of a vector change contravariantly, while the vectors themselves of the basis change *covariantly*: indeed we have

$$\frac{\partial}{\partial \bar{x}^j} = \frac{\partial x^i}{\partial \bar{x}^j} \frac{\partial}{\partial x^i}$$

and the change of basis matrix here is the *inverse* of the one that we find in (7). Luckily, we can relax: the partial derivative notation helps us write the correct form in any context.

5.1.4. Vector fields on subsets. Let M be a smooth manifold. It is sometimes useful to have vector fields defined not on the whole of M, but only on some subset $S \subset M$. By definition, a vector field in S is a smooth partial section $S \to TM$ of the tangent bundle, see Section 4.4.2. The following example may be quite common.

Example 5.1.3. If $f \colon N \hookrightarrow M$ is an embedding, every vector field X in N induces a vector field Y on the image $S = f(N)$ by setting

$$Y\big(f(p)\big) = df_p(X(p)).$$

We now rephrase Proposition 4.4.1 in this context:

Proposition 5.1.4. *If $S \subset M$ is a closed subset, every vector field on S may be extended to a global one on M.*

We may also require that the extended vector field vanishes outside of an arbitrary neighbourhood of S.

Corollary 5.1.5. *Let $N \subset M$ be a closed submanifold. Every vector field in N extends to a vector field in M that vanishes outside of any given neighbourhood of N.*

5.2. Flows

It is hard to overestimate the importance of vector fields in differential topology: they appear naturally everywhere, not only as intrinsically interesting objects, but also as very powerful tools to prove deep theorems.

In this section, we show that a vector field X on a smooth manifold M defines an infinitesimal way to deform M through a *flow* which moves every point of p along an *integral curve*, a curve that is tangent to X at every point.

Flows are powerful tools, and we will use them here to promote isotopies to *ambient isotopies* on every compact manifold.

5.2.1. Integral curves. Let M be a smooth manifold and X a given vector field on M. An *integral curve* of X is a smooth curve $\gamma\colon I \to M$ such that

$$\gamma'(t) = X\big(\gamma(t)\big)$$

for all $t \in I$.

Example 5.2.1. The curve $\gamma(t) = \frac{1}{\sqrt{m}}(\cos t, \sin t, \ldots, \cos t, \sin t)$ is an integral curve of the vector field in S^n described in Example 5.1.1.

An integral curve $\gamma\colon I \to M$ is *maximal* if there is no other integral curve $\eta\colon J \to M$ with $I \subsetneq J$ and $\gamma(t) = \eta(t)$ for all $t \in I$. Every integral curve can be extended to a maximal one by enlarging the domain as much as possible. A straightforward application of the Cauchy – Lipschitz Theorem 1.3.5 proves the existence and uniqueness of maximal integral curves:

Proposition 5.2.2. *Let X be a vector field in M. For every $p \in M$ there is a unique maximal integral curve $\gamma_p\colon I_p \to M$ with $\gamma(0) = p$.*

Proof. Pick a chart $\varphi\colon U \to \mathbb{R}^n$ and translate locally everything into $\mathbb{R}^n$. The vector field X transforms into a smooth map $\mathbb{R}^n \longrightarrow \mathbb{R}^n$, that we still denote by X for simplicity. An integral curve γ satisfies $\gamma'(t) = X\big(\gamma(t)\big)$. The local existence and uniqueness of γ follows from the Cauchy – Lipschitz Theorem 1.3.5. The maximal integral curve is also clearly unique. $\square$

5.2.2. Flows. One very nice feature of the Cauchy – Lipschitz Theorem is that the unique solution depends smoothly on the initial data. This allows us to gather all the integral curves into a single smooth global dynamical object called *flow*.

For every $p \in M$ we have an interval $I_p \subset \mathbb{R}$ and a maximal integral curve $\gamma_p\colon I_p \to M$. We first gather all the intervals into a set

$$U = \bigcup_{p \in M} \big(\{p\} \times I_p\big) \subset M \times \mathbb{R}.$$

Then we define the map $\Phi\colon U \to M$ by gathering all the integral curves:

$$\Phi(p, t) = \gamma_p(t).$$

The Cauchy –Lipschitz Theorem 1.3.5, applied locally at every point $(p, t) \in U$, implies that U is open and Φ is smooth.

The map Φ is the *flow* of the vector field X. If $U = M \times \mathbb{R}$ we say that the vector field X is *complete*. A vector field is complete if all its maximal integral curves are defined over $\mathbb{R}$.

Example 5.2.3. Pick $M = \mathbb{R}^n$ and $X = \frac{\partial}{\partial x^1}$ constantly. In this case we have $U = M \times \mathbb{R}$ and $\Phi(x, t) = x + te_1$, so X is complete. If we remove from M a random closed subset the resulting vector field X is probably not complete anymore.

Here is a simple completeness criterion.

Lemma 5.2.4. *If $M \times (-\varepsilon, \varepsilon) \subset U$ for some $\varepsilon > 0$, then X is complete.*

Proof. If at every moment of your life you are guaranteed to live at least ε more seconds, you never die.

More details follow. We fix an arbitrary point $p \in M$ and we must prove that $I_p = \mathbb{R}$. Pick any $t \in I_p$. The integral curves emanating from p and $\Phi(p,t)$ differ only by a translation of the domain: hence $I_p = I_{\Phi(p,t)} + t$ and

$$\Phi\big(\Phi(p,t), u\big) = \Phi(p, t+u) \tag{8}$$

for every $u \in I_{\Phi(p,t)}$. By hypothesis $(-\varepsilon, \varepsilon) \subset I_{\Phi(p,t)}$ and hence $(t-\varepsilon, t+\varepsilon) \subset I_p$. Since this holds for every $t \in I_p$ we get $I_p = \mathbb{R}$. $\square$

Corollary 5.2.5. *Every vector field on a compact M is complete.*

Proof. By compactness any neighbourhood U of $M \times \{0\}$ in $M \times \mathbb{R}$ must contain $M \times (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. $\square$

Let now X be a complete vector field on a smooth manifold M and Φ be its flow. We denote by $\Phi_t \colon M \to M$ the level map $\Phi_t(p) = \Phi(p,t)$.

Proposition 5.2.6. *The map Φ_t is a diffeomorphism for all $t \in \mathbb{R}$. Moreover*

$$\Phi_{-t} = \Phi_t^{-1}, \qquad \Phi_{t+s} = \Phi_t \circ \Phi_s$$

for all $t, s \in \mathbb{R}$.

Proof. The equality (8) implies that $\Phi_{t+s} = \Phi_t \circ \Phi_s$ for all $t, s \in \mathbb{R}$. This in turn gives $\Phi_{-t} = \Phi_t^{-1}$ and hence Φ_t is a diffeomorphism. $\square$

A smooth map $\Phi \colon M \times \mathbb{R} \to M$ with these properties is also called a *one-parameter group of diffeomorphisms*. Indeed we may consider this family as a group homomorphism $\mathbb{R} \to \mathrm{Diffeo}(M)$, $t \mapsto \Phi_t$ where $\mathrm{Diffeo}(M)$ is the group of all diffeomorphisms $M \to M$.

It is indeed a remarkable fact that by constructing different vector fields on a compact manifold M we get plenty of one-parameter families of diffeomorphisms for M.

Example 5.2.7. The vector field on S^n constructed in Example 5.1.1 generates the flow

$$\Phi(x^1, \ldots, x^{2m}, t) = \big(x^1 \cos t - x^2 \sin t, x^2 \cos t + x^1 \sin t, \ldots\big).$$

5.2.3. Straightening a vector field. Let X be a vector field on a smooth manifold M, and $p \in M$ a point. Among the infinitely many possible charts near p, is there one that transports X into a reasonably nice vector field in $\mathbb{R}^n$? The answer is positive if X does not vanish at p.

Proposition 5.2.8 (Straightening vector fields). *If $X(p) \neq 0$, there is a chart $U \to V$ with $p \in U$ that transports X into $\frac{\partial}{\partial x^1}$.*

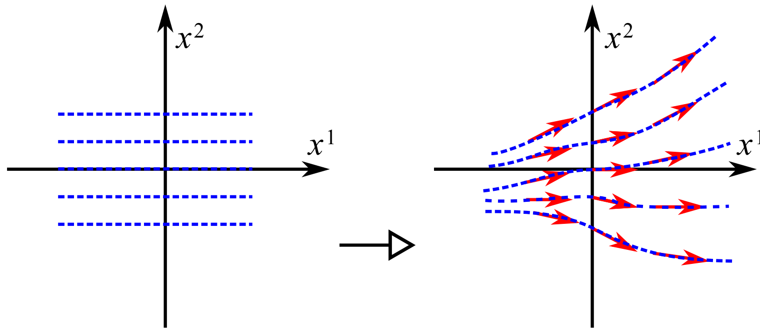

Figure 5.2. Using the flow we may construct a map ψ that sends the horizontal lines to the integral lines of X. This map straightens X.

Proof. After taking a chart we may suppose that $M = \mathbb{R}^n$, $p = 0$, and $X(p) = \frac{\partial}{\partial x^1}$. We now use the flow Φ of X to construct a chart that straightens the field X as sketched in Figure 5.2. We set

$$\psi(x^1, \ldots, x^n) = \Phi\big((0, x^2, \ldots, x^n), x^1\big).$$

The vector field X may not be complete, so $\psi(x)$ is well-defined only for sufficiently small $\|x\|$. We get a map $\psi\colon B(0,\varepsilon) \to \mathbb{R}^n$ for some $\varepsilon > 0$.

We prove that the differential $d\psi_0$ is the identity. We first note that $\psi(0, x^2, \ldots, x^n) = (0, x^2, \ldots, x^n)$, so ψ is the identity on the hyperplane $x^1 = 0$, and hence $d\psi_0(e_i) = e_i$ for $i = 2, \ldots, n$. Moreover $\gamma(t) = \psi(t, 0, \ldots, 0) = \Phi(0, t)$ is an integral curve of X, hence $d\psi_0(e_1) = \gamma'(0) = \frac{\partial}{\partial x^1} = e_1$.

Since $d\psi_0$ is invertible, the map ψ is a local diffeomorphism at 0. By construction ψ sends the lines $x + te_1$ to some integral curves of X as sketched in Figure 5.2, so it sends the vector field $\frac{\partial}{\partial x^1}$ to X. $\square$

5.3. Ambient isotopy

The previous discussion on flows and diffeomorphisms leads us naturally to a stronger form of isotopy, called *ambient isotopy*, that involves a smooth distortion of the ambient space.

5.3.1. Definition. Let M be a smooth manifold.

Definition 5.3.1. An *ambient isotopy* in M is an isotopy F between the identity $\mathrm{id}\colon M \to M$ and some diffeomorphism $\varphi\colon M \to M$, such that every level $F_t\colon M \to M$ is a diffeomorphism.

For instance, every flow Φ generated by some complete vector field X on M is an ambient isotopy between the identity Φ_0 and the diffeomorphism Φ_1.

Let now M, N be two manifolds. We say that two embeddings $f, g\colon M \to N$ are *ambiently isotopic* if there is an ambient isotopy F on N with $F_0 = \mathrm{id}$

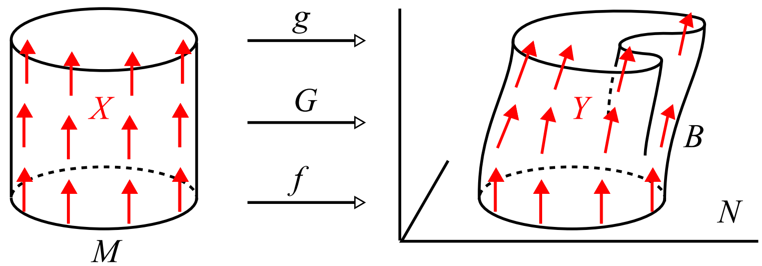

Figure 5.3. The vertical vector field X on $M \times [0,1]$ is transported via G into a vector field Y defined only on the compact set B.

and $F_1 = \varphi$ such that $g = \varphi \circ f$. We check that this notion is indeed stronger than that of an isotopy.

Proposition 5.3.2. *If f, g are ambiently isotopic, they are isotopic.*

Proof. An isotopy G_t between f and g is $G_t(x) = F_t(f(x))$. $\square$

We now use the flows to show that, if M is compact, the two notions actually coincide.

Theorem 5.3.3. *If M is compact, any two embeddings $f, g\colon M \to N$ are isotopic $\Longleftrightarrow$ they are ambiently isotopic.*

Proof. Let $F\colon M \times \mathbb{R} \to N$ be an isotopy relating f and g. We define

$$G\colon M \times \mathbb{R} \longrightarrow N \times \mathbb{R}$$

by setting $G(p,t) = \big(F(p,t), t\big)$. We note that G is time-preserving and proper (because M is compact, exercise). Moreover

$$dG_{(p,t)} = \begin{pmatrix} d(F_t)_p & * \\ 0 & 1 \end{pmatrix}$$

and hence G is an injective immersion. Being proper, the map G is an embedding (see Exercise 3.8.5) and therefore its image $G(M \times \mathbb{R})$ is a submanifold of $N \times \mathbb{R}$.

The vertical vector field $X = \frac{\partial}{\partial t}$ on $M \times [0,1]$ is transported via G to a vector field Y defined only on the compact set $B = G\big(M \times [0,1]\big)$, by setting $Y\big(G(p,t)\big) = dG_{(p,t)}\big(\frac{\partial}{\partial t}\big)$ as in Example 5.1.3. See Figure 5.3.

The vector field Y is defined only on the compact subset $B \subset N \times \mathbb{R}$, but we extend it to a vector field on the whole of $N \times \mathbb{R}$ that vanishes outside of some compact neighbourhood V of B. After that, we abruptly modify it by setting everywhere its t-component to be constantly 1. The resulting vector field (that we still name Y for simplicity) has two keys properties:

(1) it coincides with the original Y on B, since its t-component was already 1 from the beginning by construction;

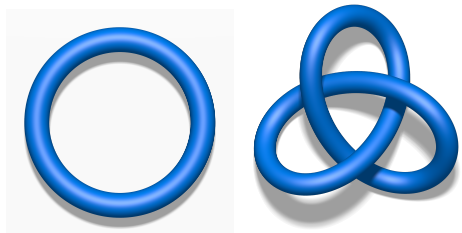

Figure 5.4. The trivial and the trefoil knot are not isotopic. This is certainly true... but how can we prove it?

(2) it coincides with $\frac{\partial}{\partial t}$ outside of V.

We now consider the flow Φ of Y in $N\times\mathbb{R}$. The vector field Y is complete: to show this, we note that V is compact and $\Phi_t(p,u) = (p, u+t)$ outside V, and these two facts easily imply that there is an $\varepsilon > 0$ such that Φ is defined at every time $|t| < \varepsilon$, so Lemma 5.2.4 applies.

Since the t-component of Y is constantly 1 we get

$$\Phi_t(p,0) = \big(H(p,t), t\big)$$

for some smooth map $H\colon N\times\mathbb{R}\to N$. We write $H_t(p) = H(p,t)$ and note that $H_t\colon N\to N$ is diffeomorphism for every t, since Φ_t is. Moreover $H_0 = \mathrm{id}$ and hence H furnishes an ambient isotopy. Finally, we have $H\big(f(p), t\big) = F(p,t)$ for every $(p,t)\in M\times[0,1]$ because $Y = dG\big(\frac{\partial}{\partial t}\big)$ on B. Therefore H is an ambient isotopy relating f and g. $\square$

Corollary 5.3.4. *Every connected smooth manifold M is* homogeneous, *that is for every two points $p,q\in M$ there is a diffeomorphism $f\colon M\to M$ isotopic to the identity such that $f(p)=q$.*

Proof. There is a smooth arc $\gamma\colon\mathbb{R}\to M$ with $\gamma(0)=p$ and $\gamma(1)=q$ (exercise). This arc may be interpreted as an isotopy between two embeddings $\{\mathrm{pt}\}\to M$ that send a single point to p and to q, respectively. This isotopy may be promoted to an ambient isotopy, that sends p to q. $\square$

How can we prove that two given homotopic embeddings are actually not isotopic? For instance, how can we prove the intuitive fact that the two knots in Figure 5.4 are not isotopic? Recall that a knot is an embedding $S^1\hookrightarrow\mathbb{R}^3$. Here is one answer: if they were isotopic, they would also be ambiently isotopic (because S^1 is compact), and hence there would be a diffeomorphism of the whole $\mathbb{R}^3$ sending the first to the second. This implies in particular that they would have homeomorphic complements. One can then try to calculate the fundamental groups of the complements and prove that they are not isomorphic: this strategy works for the two knots depicted in the figure.

5.4. Lie brackets

We now introduce an operation on vector fields called *Lie bracket*. The Lie bracket $[X, Y]$ of two vector fields X and Y in M is a third vector field that measures the "lack of commutativity" of X and Y.

5.4.1. Vector fields as derivations. Let X be a vector field on a smooth manifold M. For every open subset $U \subset M$ and every smooth function $f \in C^\infty(U)$ we may define a new function $Xf \in C^\infty(U)$ by setting

$$(Xf)(p) = X(p)(f)$$

for every $p \in U$. Recall that $X(p) \in T_pM$ is a derivation and hence transforms any locally defined function f into a real number $X(p)(f)$, so the definition of the function Xf makes sense.

In coordinates, the vector field X is written as

$$X^i \frac{\partial}{\partial x^i}$$

and the new function Xf is simply

$$X^i \frac{\partial f}{\partial x^i}.$$

This shows in particular that Xf is smooth.

We have just discovered that we can employ vector fields to "derive" functions. We use the term "derivation" here, because the Leibniz rule

$$X(fg) = (Xf)g + f(Xg)$$

is satisfied by construction for every functions f and g defined on some common open set $U \subset M$. Of course the derived function Xf depends heavily on the vector field X.

Another way of seeing Xf is as the result of a contraction of the differential df, a tensor field of type $(0, 1)$, with X, a tensor field of type $(1, 0)$. The result is a tensor field Xf of type $(0, 0)$, that is a smooth function.

5.4.2. Lie brackets. Let X and Y be two vector fields on a smooth manifold M. The *Lie bracket* $[X, Y]$ of X and Y is a new vector field, uniquely determined by requiring that

$$[X, Y]f = XYf - YXf$$

for every function f defined on any open subset $U \subset M$.

Proposition 5.4.1. *The vector field $[X, Y]$ is well-defined.*

Proof. For the moment, the bracket $[X, Y] = XY - YX$ is just an operator on smooth functions defined on any open subset $U \subset M$. For every $f, g \in$

$C^\infty(U)$ we get

$$\begin{aligned} XY(fg) &= X\big((Yf)g\big) + X\big(f(Yg)\big) \\ &= (XYf)g + (Yf)(Xg) + (Xf)(Yg) + f(XYg), \\ YX(fg) &= (YXf)g + (Xf)(Yg) + (Yf)(Xg) + f(YXg) \end{aligned}$$

from which we deduce that

$$[X,Y](fg) = \big([X,Y]f\big)g + f\big([X,Y]g\big).$$

We have proved that $[X,Y]$ is also a derivation. This allows us to define $[X,Y]$ as a vector field, by setting

$$[X,Y](p)(f) = [X,Y](f)(p)$$

for every $p \in M$ and every f defined near p. The proof is complete. □

5.4.3. Lie algebra. We introduce an important concept.

Definition 5.4.2. A *Lie algebra* is a real vector space A equipped with an antisymmetric bilinear operation $[,]$ called *Lie bracket* that satisfies the *Jacobi identity*

$$\big[[x,y],z\big] + \big[[y,z],x\big] + \big[[z,x],y\big] = 0$$

for every $x,y,z \in A$.

Let M be a smooth manifold. Recall that $\mathfrak{X}(M)$ is the vector space consisting of all the vector fields in M.

Exercise 5.4.3. The space $\mathfrak{X}(M)$ with the Lie bracket $[\,,]$ is a Lie algebra.

5.4.4. In coordinates. The definition of the Lie bracket is quite abstract and it is now due time to write an explicit formula that is valid in coordinates with respect to any chart.

Exercise 5.4.4. In coordinates we get

$$[X,Y]^i = X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}.$$

The reader may also wish to define $[X,Y]$ directly via this formula, but in that case she needs to verify that this definition is chart-independent, a fact that is not immediately obvious: if we modify the formula randomly, for instance by inserting a factor 2 after the minus sign, the definition is *not* chart-independent anymore.

In the definition of the Lie bracket of two vector fields we have seen the appearance of a recurrent theme in differential topology and geometry: the eternal quest for intrinsic (that is, chart-independent) definitions. One may fulfil this task either working entirely in coordinates, or using some more abstract arguments as we just did. As usual, both viewpoints are important.

The following exercises may be solved working in coordinates.

Exercise 5.4.5. For every $X, Y \in \mathfrak{X}(M)$ and $f, g \in C^\infty(M)$ we have

$$[fX, gY] = fg[X, Y] + f(Xg)Y - g(Yf)X.$$

Exercise 5.4.6. On an open set of $\mathbb{R}^n$, for every i, j we have

$$\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0.$$

More generally, we have

$$\left[\frac{\partial}{\partial x^i}, Y^j\frac{\partial}{\partial x^j}\right] = \frac{\partial Y^j}{\partial x^i}\frac{\partial}{\partial x^j} = \frac{\partial Y}{\partial x^i}.$$

Exercise 5.4.7. Let A, B be two $n \times n$ matrices. Consider the vector fields in $\mathbb{R}^n$ defined as

$$X(x) = Ax, \qquad Y(x) = Bx.$$

Their Lie bracket is

$$[X, Y](x) = (BA - AB)x.$$

If you get $(AB - BA)x$, you made a (quite common) mistake. This is a very instructing exercise.

5.4.5. Diffeomorphism invariance. The Lie bracket $[X, Y]$ is an important object because it is intrinsically defined given X and Y only, and this is enough to dignify it: in differential topology and geometry we long for intrinsically defined objects, because they usually have nice functorial properties. Indeed it follows readily from the definition that the bracket commutes with diffeomorphisms: a diffeomorphism $f: M \to N$ between manifolds that sends the fields X_1, X_2 to Y_1, Y_2 respectively, necessarily sends $[X_1, X_2]$ to $[Y_1, Y_2]$.

More than that, one can show the following. If $f: M \to N$ is any smooth map between manifolds, we say that two vector fields $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$ are *f-related* if $df_p(X(p)) = Y(f(p))$ for all $p \in M$.

Exercise 5.4.8. If X_1, X_2 are f-related to Y_1, Y_2 respectively, then $[X_1, X_2]$ is f-related to $[Y_1, Y_2]$.

Corollary 5.4.9. *Let $N \subset M$ be a submanifold. If X, Y are vector fields on N, and $\bar{X}, \bar{Y}$ are any extensions of X, Y to some open subset $U \subset M$ containing N, then at every point $p \in N$ we get*

$$[X, Y](p) = [\bar{X}, \bar{Y}](p).$$

We now introduce a more geometric interpretation of the Lie bracket.

5.4.6. Non-commuting flows. Let X and Y be two vector fields on a smooth manifold M, and let Φ, Ψ be their corresponding flows. Consider a point $p \in M$. In general, the two flows do not commute, that is $\Phi_s(\Psi_t(p))$ may be different from $\Psi_t(\Phi_s(p))$ whenever they are defined. We now show that the Lie bracket $[X, Y]$ at p measures this possible lack of commutation.

Proposition 5.4.10. *On any chart, we have*

$$\Psi_t(\Phi_s(p)) - \Phi_s(\Psi_t(p)) = st[X,Y](p) + o(s^2+t^2).$$

Note that the whole expression makes sense only on a chart, that is on some open subset $V \subset \mathbb{R}^n$ with $p \in V$ and with s, t sufficiently small. On a general smooth manifold M the points $\Psi_t(\Phi_s(p))$ and $\Phi_s(\Psi_t(p))$ are probably distinct points in M and there is no way of estimating their "distance". The expression is however very useful because it holds on *every* chart.

Proof. We fix p and consider the smooth function

$$F(s,t) = \Psi_t(\Phi_s(p)) - \Phi_s(\Psi_t(p)).$$

We have $F(s,0) = F(0,t) = 0$ for all s, t. Since $F \equiv 0$ on the axis $s = 0$ and $t = 0$, the second-order Taylor expansion of F reduces to

$$F(s,t) = st\frac{\partial^2 F}{\partial s \partial t}\Big|_{s=t=0} + o(s^2+t^2).$$

There is only one second-order term that we now calculate. We have

$$\frac{\partial}{\partial t}\Psi_t(\Phi_s(p))\Big|_{t=0} = Y\big(\Phi_s(p)\big)$$

and then

$$\frac{\partial^2}{\partial s \partial t}\Psi_t(\Phi_s(p))\Big|_{s=t=0} = \frac{\partial}{\partial s}Y\big(\Phi_s(p)\big)\Big|_{s=0} = X^j(p)\frac{\partial Y}{\partial x^j}(p).$$

Therefore

$$\frac{\partial^2 F}{\partial s \partial t}(0,0) = X^j(p)\frac{\partial Y}{\partial x^j}(p) - Y^j(p)\frac{\partial X}{\partial x^j}(p) = [X,Y](p)$$

by Exercise 5.4.4. The proof is complete. □

We say that two vector fields X and Y *commute* if $[X,Y] = 0$ everywhere. The corresponding flows Φ and Ψ *commute locally* if

$$\Phi_s(\Psi_t(p)) = \Psi_t(\Phi_s(p))$$

for every p and sufficiently small s, t. These two notions coincide:

Proposition 5.4.11. *Two vector fields commute $\Longleftrightarrow$ their flows do locally.*

Proof. If the flows commute, then $[X,Y] = 0$ because of Proposition 5.4.10. Conversely, suppose that $[X,Y] = 0$.

Consider a point $p \in M$. If $X(p) = Y(p) = 0$, we get $\Phi_s(p) = \Psi_t(p) = p$ and we are done. Otherwise, suppose that $X(p) \neq 0$. On a chart we can straighten X and get $X = \frac{\partial}{\partial x^1}$ and $\Phi_s(p) = p + se_1$.

Now $[X,Y] = 0$ and Exercise 5.4.6 imply that

$$\frac{\partial Y}{\partial x^1} = 0.$$

The field Y is hence invariant by translations along e_1. Therefore $\Psi_t(p+se_1) = \Psi_t(p) + se_1$, that is Ψ_t commutes with Φ_s. □

Exercise 5.4.12. If both X and Y are complete, their flows Φ and Ψ commute locally if and only if they commute *globally*, that is $\Phi_s(\Psi_t(p)) = \Psi_t(\Phi_s(p))$ for every p, s, t.

5.4.7. Multiple straightenings. Can we straighten two or more vector fields simultaneously? It should not be a surprise now that the answer depends on their Lie brackets. Let $X_1, \ldots, X_k$ be vector fields on a smooth manifold M, and $p \in M$ be a point.

Proposition 5.4.13. *Suppose that $X_1(p), \ldots, X_k(p)$ are independent vectors. There is a chart $U \to V$ that transports $X_1, \ldots, X_k$ into $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^k}$ $\Longleftrightarrow$ $[X_i, X_j] = 0$ for all i, j on some neighbourhood of p.*

Proof. If there is a chart of this type, then clearly $[X_i, X_j] = 0$. We now prove the converse and suppose $[X_i, X_j] = 0$ for all i, j. The proof is similar to that of Proposition 5.2.8.

By taking a chart we may suppose that M is an open set in $\mathbb{R}^n$, $p = 0$, and $X_i(0) = \frac{\partial}{\partial x^i}$ for all $i = 1, \ldots, k$. Let Φ^i_t be the flow of X_i. Define

$$\psi(x^1, \ldots, x^n) = \Phi^k_{x^k}(\cdots(\Phi^1_{x^1}(0, \ldots, 0, x^{k+1}, \ldots, x^n))\cdots).$$

The differential $d\psi_0$ is the identity, because

$$\psi(0, \ldots, 0, x^{k+1}, \ldots, x^n) = (0, \ldots, 0, x^{k+1}, \ldots, x^n)$$

and $\gamma_i(t) = \psi(te_i)$ with $i = 1, \ldots, k$ is an integral curve for X_i, so $\gamma_i'(0) = \frac{\partial}{\partial x^i}$.

We deduce that ψ is a local diffeomorphism. It is clear that ψ sends the lines $x + te_k$ to integral curves for X_k, so it sends $\frac{\partial}{\partial x^k}$ to X_k. Since $[X_i, X_j] = 0$, the flows Φ^i_t commute and we can permute them in the definition of ψ at our pleasure: so we can put Φ^i_t at the end of the composition and the same argument shows that ψ sends $\frac{\partial}{\partial x^i}$ to X_i for all i. $\square$

5.4.8. Lie derivative. We have just noted that a vector field X may be used to derive functions. Can we also use X to derive other objects, for instance another vector field Y or more generally any tensor field s? The answer is positive, and this operation is called the *Lie derivative*.

We first note that every diffeomorphism $f: M \to N$ induces an isomorphism between the corresponding tensor bundles

$$f_*: \mathcal{T}^k_h M \longrightarrow \mathcal{T}^k_h N$$

induced from that of the tangent bundles $f_*: TM \to TN$, and we may use f_* to transfer tensor fields from M to N and viceversa.

Let now X be a vector field on a smooth manifold M, and let s be any tensor field on M, of some type (h, k). The *Lie derivative* $\mathcal{L}_X s$ is a new tensor field of the same type (h, k), morally obtained by deriving s along X, and defined as follows.

Let Φ_t be the flow generated by X. For every point $p \in M$, there is a sufficiently small $\varepsilon > 0$ such that Φ_t is defined on a neighbourhood of p and is a local diffeomorphism at p for all $|t| < \varepsilon$. Therefore $(\Phi_t)_*(s)$ is another tensor field defined on a neighbourhood of $\Phi_t(p)$, that varies smoothly in t, and we now want to compare s and $(\Phi_t)_*(s)$.

We note that the tensor

$$(\Phi_{-t})_*\big(s(\Phi_t(p))\big)$$

is well-defined and lies in $\mathcal{T}_h^k(T_pM)$ for every sufficiently small t and varies smoothly in t, so it makes sense to define its derivative

$$\big(\mathcal{L}_X s\big)(p) = \frac{d}{dt}\Big|_{t=0}(\Phi_{-t})_*\big(s(\Phi_t(p))\big).$$

We have defined a linear map

$$\mathcal{L}_X \colon \Gamma\big(\mathcal{T}_h^k(M)\big) \longrightarrow \Gamma\big(\mathcal{T}_h^k(M)\big)$$

that "derives" any tensor field along X.

Exercise 5.4.14. The following hold:

- if $f \in C^\infty(M)$, then $\mathcal{L}_X f = Xf$;
- if Y is a vector field, then $\mathcal{L}_X Y = [X, Y]$;
- for every tensor fields S and T of any types we have

$$\mathcal{L}_X(S \otimes T) = (\mathcal{L}_X S) \otimes T + S \otimes (\mathcal{L}_X T);$$

- the Lie derivative commutes with contractions.

The Lie derivative $\mathcal{L}_X s$ measures how s changes along X, in fact it follows readily from the definition that $\mathcal{L}_X s \equiv 0$ on $M \iff$ the tensor field s is invariant under the flow Φ_t.

It is important to note here that, as opposite to the directional derivative in $\mathbb{R}^n$, the value of $\mathcal{L}_X s$ at a point p depends on the local behaviour of X near p, not on the directional vector $X(p)$ alone! To get a derivation that, like the directional derivative in $\mathbb{R}^n$, depends in p only on the directional vector based at p, we need to introduce an additional structure called *connection*. We will do this later on in this book. The Lie derivative is the maximum we can get on a smooth manifold without equipping it with some additional structure.

5.5. Foliations

We now introduce some higher-dimensional analogues of vector fields and integral curves, where we replace vectors with k-dimensional subspaces, and integral curves with k-dimensional submanifolds.

5.5.1. Foliations. Let M be a smooth n-manifold. An *immersed submanifold* in M is the image of an immersion $S \to M$. A *horizontal* affine k-plane in $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$ is an affine k-plane of type $\mathbb{R}^k \times \{c\}$ for some fixed $c \in \mathbb{R}^{n-k}$.

Definition 5.5.1. A k-dimensional *foliation* in M is a partition $\mathscr{F} = \{\lambda_i\}$ of M into injectively immersed k-dimensional connected submanifolds $\lambda_i \subset M$ called *leaves*, such that the following holds: for every $p \in M$ there is a chart $\varphi\colon U \to \mathbb{R}^n$ with $p \in U$ such that $\varphi(\lambda_i \cap U)$ is the union of some parallel horizontal affine k-planes, for every i.

In other words, at every point p there is a chart φ that transforms the partition $\mathscr{F}$ near p into the partition of $\mathbb{R}^n$ into parallel horizontal k-planes. We say that such a chart φ is *compatible* with the foliation.

Remark 5.5.2. For a fixed leaf λ_i, the image $\varphi(\lambda_i \cap U)$ along a compatible chart φ may consist of infinitely many k-planes. These are countable, because λ_i is the image of an immersed submanifold $S \to M$ and S has countable base.

We also note that a foliation contains uncountably many leaves: this is a consequence of the previous remark, or of the more general fact that the union of countably many immersed manifolds of smaller dimension than M has measure zero and hence cannot cover M.

Example 5.5.3. The following are foliations:

1. the partition of $\mathbb{R}^n$ into all the affine spaces parallel to a fixed vector subspace $L \subset \mathbb{R}^n$;
2. if $E \to B$ is a fibre bundle, the partition of E into the fibres E_p;
3. for a fixed *slope* $\nu \in \mathbb{R}$, the family of all curves $\alpha\colon \mathbb{R} \to S^1 \times S^1$ of type $\alpha(t) = \left(e^{it}, e^{i(\nu t + \mu)}\right)$ as μ varies.

Exercise 5.5.4. In the last example, the leaves are compact $\Longleftrightarrow \lambda \in \mathbb{Q}$. If $\lambda \in \mathbb{R} \setminus \mathbb{Q}$ every leaf is dense.

We now furnish an equivalent definition of foliation.

Definition 5.5.5. A k-dimensional *foliation* in M is an atlas $\{\varphi_i\colon U_i \to \mathbb{R}^n\}$ compatible with the smooth structure of M whose transition maps φ_{ij} are all locally of the following form:

$$\varphi_{ij}(x, y) = \left(\varphi^1_{ij}(x, y), \varphi^2_{ij}(y)\right).$$

Here we represent $\mathbb{R}^n$ as $\mathbb{R}^k \times \mathbb{R}^{n-k}$, both as a domain and as a codomain.

In other words, we require that the last $n-k$ coordinates of φ_{ij} should depend locally only on the last $n-k$ coordinates of the point. By "locally" we mean as usual that every point p in the domain of φ_{ij} has a neighbourhood such that φ_{ij} is of that form.

The two definitions look very different but are indeed equivalent! If $\mathscr{F}$ is a foliation in the partition sense, by considering only charts that are compatible with $\mathscr{F}$ we get an atlas as in Definition 5.5.5 (exercise). Conversely, given an atlas $\mathcal{A} = \{\varphi_i\}$ of this kind, the transition maps preserve locally the k-dimensional affine horizontal subspaces $\{y = c\}$ which hence glue to form immersed submanifolds in M.

To construct the immersed manifolds rigorously, we proceed as follows. We assign to $\mathbb{R}^k$ and $\mathbb{R}^{n-k}$ respectively the Euclidean and the discrete topology, and we give the product topology to $\mathbb{R}^k \times \mathbb{R}^{n-k}$. Note that this topology is finer than the Euclidean one. We now use this model to define a finer topology on M, by declaring a set in M to be open if it intersects the domain U_i of every chart $\varphi_i \in \mathcal{A}$ into a subset whose image in $\varphi_i(U_i) \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$ is open in the new finer topology.

The manifold M with the finer topology decomposes into (uncountably many) connected components $\{M_j\}$. The atlas $\{\varphi_i : U_i \to \mathbb{R}^n\}$ furnishes to every M_j a structure of smooth manifold: the only tricky part here is to prove that it has a countable base, and is left as an exercise. *Hint:* Select a countable sub-atlas $\mathcal{A}' \subset \mathcal{A}$ and prove that every leaf "propagates" only to countably many nearby ones at each step.

5.5.2. Distributions. Let M be a smooth n-manifold. Here is another natural geometric definition.

Definition 5.5.6. A *k-distribution* in M is a rank-k subbundle D of the tangent bundle TM.

A distribution is a collection of k-subspaces $D_p \subset T_pM$ that vary smoothly with p, see Exercise 4.6.6. On $\mathbb{R}^n$, a distribution is like a smooth map $\mathbb{R}^n \to \mathrm{Gr}^k(\mathbb{R}^n)$ that places a subspace $D_x \subset \mathbb{R}^n = T_x\mathbb{R}^n$ at every $x \in \mathbb{R}^n$

Example 5.5.7. If $\mathscr{F}$ is a k-dimensional foliation on M, the k-spaces tangent to the leaves of $\mathscr{F}$ form a k-distribution.

A distribution that is tangent to some foliation $\mathscr{F}$ is called *integrable*. Note that a diffeomorphism $\varphi : M \to M'$ transforms a distribution D on M into one D' on M' in the obvious way, by setting $D'_{\varphi(p)} = d\varphi_p(D_p)\ \forall p \in M$. The integrability condition may also be expressed without using foliations:

Proposition 5.5.8. *D is integrable $\Longleftrightarrow \forall p \in M$ there is a chart $\varphi : U \to \mathbb{R}^n$ with $p \in U$ that transforms D into the horizontal distribution.*

The horizontal distribution in $\mathbb{R}^n$ is $D_x = \mathbb{R}^k \times \{0\} \subset \mathbb{R}^k \times \mathbb{R}^{n-k}\ \forall x \in \mathbb{R}^n$.

Proof. ($\Rightarrow$). If D is tangent to a foliation $\mathscr{F}$, any chart compatible with $\mathscr{F}$ transforms D into the horizontal one.

($\Leftarrow$). All these charts define a foliation in the sense of Definition 5.5.5. $\square$

5.5.3. The Frobenius Theorem. We now state and prove a theorem that characterises the integrable distributions via the Lie bracket of vector fields.

A vector field X on a manifold M is *tangent* to a distribution D if $X(p) \in D_p$ for all $p \in M$. A distribution D is *involutive* if whenever X, Y are two vector fields defined in some open set that are tangent to D, their Lie bracket $[X, Y]$ is also tangent.

Theorem 5.5.9 (Frobenius Theorem). *A distribution D on a manifold M is integrable $\Longleftrightarrow$ it is involutive.*

Proof. If D is integrable, at every $p \in M$ there is a chart that transforms it into the horizontal distribution $D_x = \mathbb{R}^k \times \{0\}$ in $\mathbb{R}^n$. If X, Y are vector fields in $\mathbb{R}^n$ tangent to D, they are of the form

$$X = \sum_{i=1}^k X^i \frac{\partial}{\partial x^i}, \qquad Y = \sum_{i=1}^k Y^i \frac{\partial}{\partial x^i}$$

and by Exercise 5.4.4 we get $[X, Y]^i = 0$ for all $i > k$. Therefore $[X, Y]$ is also tangent to D and D is involutive.

Conversely, suppose that D is involutive. For every $p \in M$ we pick a chart near p that transforms p in 0 and D_p into the horizontal space $D_0 = \mathbb{R}^k \times \{0\}$. For sufficiently small x every k-space D_x may not be horizontal, but it still intersects the vertical space $V = \{0\} \times \mathbb{R}^{n-k}$ in the origin. By projecting along V we get canonical isomorphisms $D_0 \to D_x$ that send the basis $e_1, \ldots, e_k$ to a local frame on D of the type

$$X_1 = \frac{\partial}{\partial x^1} + \sum_{i=k+1}^n X_1^i \frac{\partial}{\partial x^i}, \quad \ldots, \quad X_k = \frac{\partial}{\partial x^k} + \sum_{i=k+1}^n X_k^i \frac{\partial}{\partial x^i}.$$

Exercise 5.4.4 gives $[X_i, X_j]^l = 0$ for all $i, j, l = 1, \ldots, k$, hence $[X_i, X_j]$ is tangent to the vertical space V at every point. Since D is involutive, the vector field $[X_i, X_j]$ must be tangent to D and this implies that $[X_i, X_j] = 0$.

We have discovered that $X_1, \ldots, X_k$ are commuting vector fields and by Proposition 5.4.13 we can transform them via a chart into the coordinate ones $X_i = \frac{\partial}{\partial x^i}$. In this chart the distribution is horizontal so Proposition 5.5.8 applies. The proof is complete. $\square$

As an example, the vector fields in $\mathbb{R}^3$

$$X_1 = \frac{\partial}{\partial x}, \qquad X_2 = \frac{\partial}{\partial y} + x\frac{\partial}{\partial z}$$

do not commute since $[X_1, X_2] = \frac{\partial}{\partial z}$. Therefore they generate a non-integrable plane distribution in $\mathbb{R}^3$, drawn in Figure 5.5.

The following criterion may be useful in some cases.

Exercise 5.5.10. A distribution D in M is involutive $\Longleftrightarrow$ for every $p \in M$ there is a local frame $X_1, \ldots, X_k$ for D such that $[X_i, X_j]$ is tangent to D $\forall i, j$.

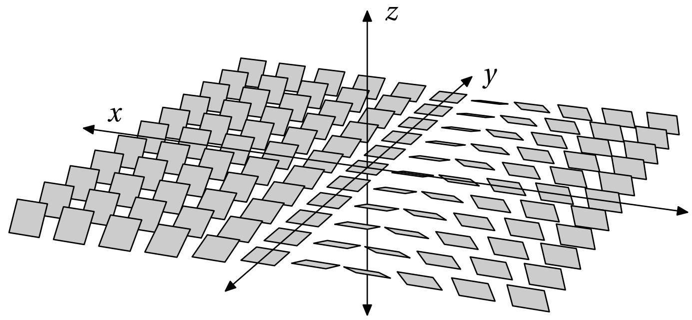

Figure 5.5. A non-integrable plane distribution in $\mathbb{R}^3$.

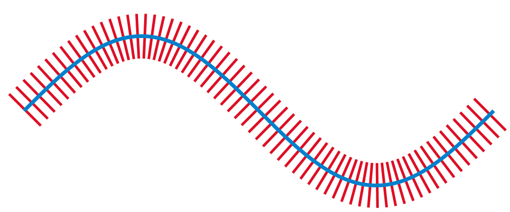

Figure 5.6. A tubular neighbourhood of a curve on the plane.

Hint. To prove $\Leftarrow$, write any vector field X, Y tangent to D locally as a combination of X_i, with coefficients that are smooth functions. Use Exercise 5.4.5 to deduce that $[X, Y]$ is also tangent to D. $\square$

5.6. Tubular neighbourhoods

Let M be a smooth m-manifold. Among all the open neighbourhoods of a given point $p \in M$, the simplest ones are undoubtedly those that are diffeomorphic to $\mathbb{R}^m$. These are certainly not unique, and there is no canonical way to choose a preferred one; however, we will prove in this section that these are unique up to isotopy, thus answering to Question 3.10.7.

More generally, we will show that not only points, but any submanifold $N \subset M$ has a similar kind of nice open neighbourhood, called a *tubular neighbourhood*. The idea that we have in mind is that, for a curve on the plane, a tubular neighbourhood should look like in Figure 5.6, and for a knot $K \subset \mathbb{R}^3$ it should be a little open tube around K. As in Figure 5.6, a tubular neighbourhood should be a bundle over N.

We prove here the existence and uniqueness (up to isotopy) of tubular neighbourhoods for any submanifold $N \subset M$.

5.6.1. Definition. Let M be a m-manifold and $N \subset M$ a n-submanifold. A *tubular neighbourhood* for N is a vector bundle $E \to N$ together with an embedding $i\colon E \hookrightarrow M$ such that:

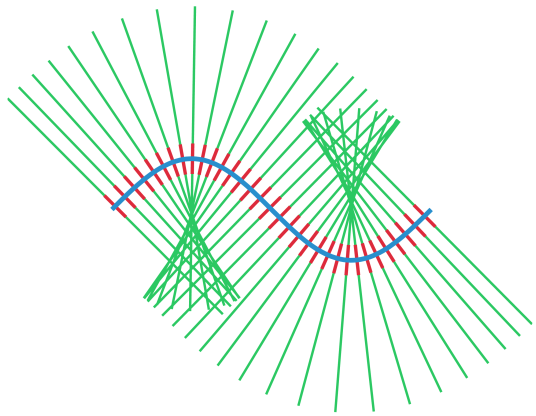

Figure 5.7. To construct a tubular neighbourhood, we map the normal bundle in $\mathbb{R}^n$ and pick a sufficiently small neighbourhood of N so that this map is an embedding.

- $i|_N = \mathrm{id}_N$, where we identify N with the zero-section in E;
- $i(E)$ is an open neighbourhood of N.

We usually call a tubular neighbourhood simply the image $i(E)$ of E in N, but keeping in mind that it has a bundle structure with base N.

The second hypothesis implies that $\dim E = \dim M$, so E must have rank $m - n$. Recall that the normal bundle νN of N inside M has precisely that rank, so it seems a promising candidate.

5.6.2. Existence. We now prove the existence of tubular neighbourhoods in two steps: in the first step we only consider the case $M = \mathbb{R}^m$.

Proposition 5.6.1. *Every submanifold $N \subset \mathbb{R}^m$ has a tubular neighbourhood with $E = \nu N$.*

Proof. As shown in Example 4.3.6, we have

$$\nu N = \big\{(p, v) \bigm| p \in N, v \in \nu_p N\big\} \subset N \times \mathbb{R}^m \subset \mathbb{R}^m \times \mathbb{R}^m.$$

We have identified $\nu_p N$ with $T_p N^\perp$. We now define the smooth map

$$\begin{aligned} f\colon \quad \nu N &\longrightarrow \mathbb{R}^m, \\ (p, v) &\longmapsto p + v. \end{aligned}$$

See Figure 5.7. We now study the differential $df_{(p,0)}$ at each $p \in N$. We have

$$T_{(p,0)}\nu N = T_p N \times \nu_p N \subset \mathbb{R}^m \times \mathbb{R}^m.$$

If we identify $T_p N \times \nu_p N$ with $\mathbb{R}^m$, we discover easily that the differential $df_{(p,0)}$ is the identity. In particular, it is invertible, so f is an immersion at every point in N. We now prove that there is an open neighbourhood U of N in νN where f is an embedding, see Figure 5.7.

Since being an immersion is an open condition, there is an open neighbourhood U of N where f is an immersion. Since $f|_N$ is injective, after possibly taking a smaller U we may also suppose that $f|_U$ is injective and an embedding (exercise).

By shrinking νN as in Lemma 4.5.11 we can embed $i\colon \nu N \hookrightarrow U$ keeping N fixed, and the composition $f \circ i$ is a tubular neighbourhood for N. $\square$

We now turn to a more general case.

Theorem 5.6.2. *Let M be a manifold. Every submanifold $N \subset M$ has a tubular neighbourhood with $E = \nu N$.*

Proof. We may embed M in some $\mathbb{R}^k$ thanks to Whitney's Theorem 3.11.8. Now for every $p \in N$ we have the vector space inclusions

$$T_pN \subset T_pM \subset \mathbb{R}^k.$$

We identify ν_pN with the orthogonal complement of T_pN inside T_pM, so that

$$T_pN \oplus \nu_pN = T_pM \subset \mathbb{R}^k.$$

We consider the smooth map

$$\begin{aligned} F\colon \quad \nu N &\longrightarrow \mathbb{R}^k, \\ (p, v) &\longmapsto p + v. \end{aligned}$$

Let W be a tubular neighbourhood of M in $\mathbb{R}^k$, with bundle projection $\pi\colon W \to M$. We set $U = F^{-1}(W)$ and define the map

$$\begin{aligned} f\colon \quad U &\longrightarrow M, \\ (p, v) &\longmapsto \pi(p + v). \end{aligned}$$

As above, the differential at N is just the identity and we conclude that $f \circ i$ is a tubular neighbourhood for N for some appropriate bundle shrinking i. $\square$

5.6.3. Uniqueness. It is a remarkable and maybe surprising fact that, despite their quite general definition, tubular neighbourhoods are actually unique if one considers them up to isotopy.

We must clarify what we mean by "isotopy" in this context. Let M be a manifold and $N \subset M$ a submanifold. Two tubular neighbourhoods $i_0\colon E^0 \to M$ and $i_1\colon E^1 \to M$ are *isotopic* if there are a bundle isomorphism $\psi\colon E^0 \to E^1$ and an isotopy F relating the embeddings i_0 and $i_1 \circ \psi$ that keeps N pointwise fixed, that is such that $F(p, t) = p$ for all $p \in N$ and all t.

Note that each embedding $F_t = F(\cdot, t)$ is a tubular neighbourhood of N, so F indeed describes a smooth path of varying tubular neighbourhoods.

Theorem 5.6.3. *Let M be a manifold and $N \subset M$ a submanifold. Every two tubular neighbourhoods of N are isotopic.*

To warm up, we start by proving the following.

Proposition 5.6.4. *Every embedding $f : \mathbb{R}^n \hookrightarrow \mathbb{R}^n$ with $f(0) = 0$ is isotopic to its differential df_0 via an isotopy that fixes 0 at each time.*

Proof. The isotopy for $t \in (0, 1]$ is simply defined as follows:

$$F(x, t) = \frac{f(tx)}{t}.$$

We extend it to the time $t = 0$ by writing the first-order Taylor expansion

$$f(x) = h_1(x)x^1 + \ldots + h_n(x)x^n$$

where $h_i(0) = \frac{\partial f}{\partial x^i}(0)$ for all i. For every $t \in (0, 1]$ we get

$$F(x, t) = h_1(tx)x^1 + \ldots + h_n(tx)x^n$$

and this expression makes sense also for $t = 0$, yielding the equality $F(x, 0) = df_0(x)$. The proof is complete.[1] □

We can now prove Theorem 5.6.3.

Proof. Let E^0 and E^1 be two tubular neighbourhoods of N. We see E^1 as embedded directly in M, and we want to modify the given embedding $f : E^0 \to M$ via an isotopy so that it matches with E^1.

We first prove that after an isotopy we may suppose that $f(E^0) \subset E^1$. Indeed, Lemma 4.5.11 provides a shrinkage $g : E^0 \to E^0$ with $f \circ g(E^0) \subset E^1$ isotopic to the identity through a family g_t of embeddings, and by composing it with f we get an isotopy between f and $f \circ g$.

Now that $f(E^0) \subset E^1$, we can construct the isotopy $F : E^0 \times [0, 1] \to M$ by mimicking the proof of Proposition 5.6.4: we simply write

$$F(v, t) = \frac{f(tv)}{t}.$$

Here $f(tv)$ is a particular vector in E^1 and hence its division by t makes sense. This is certainly an isotopy for $t \in (0, 1]$, and we now extend it to $t = 0$ similarly to what we did above.

Consider a $v \in E^0$, with $p = \pi(v) \in N$. The point p has an open neighbourhood U above which E^1 is trivialised as $U \times \mathbb{R}^{m-n}$. There are also a smaller neighbourhood $V \subset U$ and a $r > 0$ such that $E^0|_V$ is also trivialised as $V \times \mathbb{R}^{m-n}$ and moreover

$$f\big(V \times B(0, r)\big) \subset U \times \mathbb{R}^{m-n}.$$

This holds by continuity. See Figure 5.8. We may represent f on $V \times B(0, r)$ as a map

$$f(x, y) = \big(f_1(x, y), f_2(x, y)\big).$$

We have $f(x, 0) = (x, 0)$. Since $f_2(x, 0) = 0$ we can write

$$f_2(x, y) = h_1(x, y)y^1 + \ldots + h_{m-n}(x, y)y^{m-n}$$

[1] To be precise, we should substitute t with $\rho(t)$ via a transition function ρ to get an isotopy defined for all $t \in \mathbb{R}$. We will tacitly assume this in other points in this book.

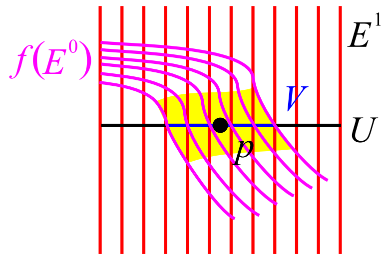

Figure 5.8. By continuity, we can find two neighbourhoods $V \subset U$ of p above which both E^0 and E^1 trivialise, and a $r > 0$ such that $f\big(V \times B(0,r)\big) \subset U \times \mathbb{R}^{m-n}$ (the yellow zone).

with

$$h_i(x,0) = \frac{\partial f_2}{\partial y^i}(x,0).$$

We can then represent F as

$$\begin{aligned} F(x,y,t) &= \left(f_1(x,ty), \frac{1}{t} f_2(x,ty)\right) \\ &= \big(f_1(x,ty), h_1(x,ty)y^1 + \ldots + h_{m-n}(x,ty)y^{m-n}\big). \end{aligned}$$

This map is well-defined and smooth also at $t=0$. The map at $t=0$ is

$$F_0(x,y) = F(x,y,0) = \left(x, \frac{\partial f_2}{\partial y}(x,0)y\right).$$

It sends every fibre of E^0 to a fibre of E^1 via a linear map, which is in fact an isomorphism because f is an embedding and hence

$$df_{(x,0)} = \begin{pmatrix} I_n & * \\ 0 & \frac{\partial f_2}{\partial y}(x,0) \end{pmatrix}$$

is an isomorphism. Therefore $F_0 \colon E^0 \to E^1$ is a bundle isomorphism. □

We have proved that the tubular neighbourhood of a submanifold $N \subset M$ is unique up to isotopy and bundle isomorphisms: in particular, this shows that every tubular neighbourhood of N is isomorphic to the normal bundle νN.

5.6.4. Embedding open balls. The uniqueness theorem for tubular neighbourhoods is quite powerful, and it has some remarkable consequences already when N is a point.

Proposition 5.6.5. *Let M be a connected smooth n-manifold. Two embeddings $f, g \colon \mathbb{R}^n \hookrightarrow M$ are always isotopic, possibly after pre-composing g with a reflection in $\mathbb{R}^n$.*

Proof. We may see both f and g as tubular neighbourhoods of $f(0)$ and $g(0)$. Since connected manifolds are homogeneous (Corollary 5.3.4), after an ambient isotopy we may suppose that $f(0) = g(0)$. By the uniqueness of the tubular neighbourhood, the map f is isotopic to $g \circ \psi$ for some linear isomorphism $\psi\colon \mathbb{R}^n \to \mathbb{R}^n$. By Corollary 3.9.11 we may isotope ψ to be either the identity or a reflection. $\square$

The oriented version is more elegant to state:

Proposition 5.6.6. *Let M be an oriented connected smooth n-manifold. Two orientation-preserving embeddings $f, g\colon \mathbb{R}^n \hookrightarrow M$ are always isotopic.*

5.6.5. Hypersurfaces. Let M be a smooth manifold. A *hypersurface* in M is a submanifold $N \subset M$ of codimension 1.

Proposition 5.6.7. *Let M be orientable. The normal bundle of a hypersurface $N \subset M$ is trivial $\Longleftrightarrow$ N is also orientable.*

Proof. Fix an orientation for M. The normal bundle is a line bundle, and it is trivial $\Longleftrightarrow$ it has a nowhere-vanishing section.

If N is orientable, we fix an orientation on N. The two orientations of M and N induce a locally coherent orientation on the normal line νN_p for every $p \in N$, which distinguishes between "positive" and "negative" normal vectors, see Exercise 2.5.2. Fix a Riemannian metric on νN, and pick all the positive vectors of norm one: they form a nowhere-vanishing section.

On the other hand, if the normal bundle is trivial, the normal orientation and the orientation of M induce similarly an orientation on N. $\square$

5.6.6. Continuous maps are homotopic to smooth maps. By combining the tubular neighbourhoods and Whitney's Embedding Theorem, we may now prove that every continuous map between smooth manifolds is homotopic to a smooth map. Let M and N be two smooth manifolds.

Theorem 5.6.8. *Let $f\colon M \to N$ be a continuous map, whose restriction to some (possibly empty) closed subset $S \subset M$ is smooth. The map f is continuously homotopic to a smooth map $g\colon M \to N$ with $f(x) = g(x)$ for all $x \in S$, via a homotopy that is constant on S.*

Proof. By Whitney's Embedding Theorem 3.11.8 we may suppose that $N \subset \mathbb{R}^n$ for some n. Let νN be a tubular neighbourhood of N. For every $p \in N$ we let $r(p)$ be the distance from p to the boundary of the open set νN.

By Proposition 3.3.9 there is a smooth map $h\colon M \to \mathbb{R}^n$ with $\|h(p) - f(p)\| < r(f(p))$ and $h(p) = f(p) \forall p \in S$. The homotopy $H(p,t) = (1-t)f(p) + th(p)$ lies entirely in νN and hence can be composed with the projection $\pi\colon \nu N \to N$ to give a homotopy $G(p,t) = \pi(H(p,t))$ between f and the smooth $g = \pi \circ h$. $\square$

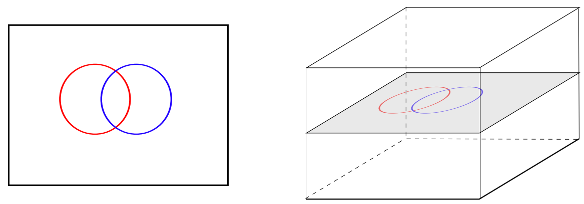

Figure 5.9. Transversality depends on the ambient space: the two curves are transverse in $\mathbb{R}^2$, not in $\mathbb{R}^3$.

The proof shows also that g may be chosen to be arbitrarily close to f, but to express "closeness" rigorously we need to see N embedded in some $\mathbb{R}^n$.

Corollary 5.6.9. *Two smooth maps $f, g\colon M \to N$ are continuously homotopic $\iff$ they are smoothly homotopic.*

Proof. Every continuous homotopy $F\colon M \times [0,1] \to N$ can be extended to a continuous map $F\colon M \times \mathbb{R} \to N$ and then be homotoped to a smooth map $G\colon M \times \mathbb{R} \to N$ by keeping $F|_{M\times\{0\}}$ and $F|_{M\times\{1\}}$ fixed. □

5.7. Transversality

We now show that any two smooth maps (and in particular, submanifolds with their inclusion maps) can be perturbed to cross nicely. The notion of "nice crossing" is surprisingly simple to define and is called *transversality*.

5.7.1. Definition. Let $f\colon M \to N$ and $g\colon W \to N$ be two smooth maps between manifolds, sharing the same target N.

Definition 5.7.1. We say that f and g are *transverse* if for every $p \in M$ and $q \in W$ with $f(p) = g(q)$ we have

$$\operatorname{Im} df_p + \operatorname{Im} dg_q = T_{f(p)}N.$$

In this case we write $f \pitchfork g$.

If $M \subset N$ is a submanifold and $g\colon W \to N$ is a map, we say that g and M are *transverse* if g and the inclusion map $M \hookrightarrow N$ are, and in this case we write $g \pitchfork M$. Two submanifolds $M, W \subset N$ are *transverse* if their inclusions are, and in this case we write $M \pitchfork W$ to denote their intersection $M \cap W$.

Set $m = \dim M$, $w = \dim W$, and $n = \dim N$. Note that if $m + w < n$ then $f \pitchfork g \iff$ the maps f and g have disjoint images. See Figure 5.9.

If $W = \{q\}$ is a point, then $f \pitchfork g \iff g(q)$ is a regular value for f.

5.7.2. Fibre bundles. Here is a basic example.

Proposition 5.7.2. *Let $\pi\colon E \to M$ be a fibre bundle. A map $f\colon N \to E$ is transverse to a fibre $E_q \iff q$ is a regular value for $\pi \circ f$.*

Proof. Pick $p \in N$ with $f(p) \in E_q$. We have $T_{f(p)}E_q = \ker d\pi_{f(p)}$, so

$$\operatorname{Im} df_p + T_{f(p)}E_q = T_{f(p)}E \iff \operatorname{Im} d(\pi \circ f)_p = T_qM.$$

The proof is complete. □

Exercise 5.7.3. A submanifold $W \subset E$ is the image of a section of a bundle $E \to M \iff$ it intersects transversely every fibre E_q in a single point.

5.7.3. Intersections. We now extend a theorem from the context of regular values to the wider realm of transverse maps.

Proposition 5.7.4. *Let $M \subset N$ be a submanifold and $g\colon W \to N$ a smooth map. If $g \pitchfork M$ then $X = g^{-1}(M)$ is a submanifold of codimension $n - m$.*

Proof. Pick $p \in X$. We look at a neighbourhood of $q = g(p) \in M$ and after taking a chart we suppose that $(M, N) = (\mathbb{R}^m \times \{0\}, \mathbb{R}^m \times \mathbb{R}^{n-m})$, $q = 0$.

Consider the projection $\pi\colon \mathbb{R}^m \times \mathbb{R}^{n-m} \to \mathbb{R}^{n-m}$ onto the second factor. Near p we have $X = g^{-1}(\mathbb{R}^m \times \{0\}) = g^{-1}(\pi^{-1}(0)) = (\pi \circ g)^{-1}(0)$ and by Proposition 5.7.2 the composition $\pi \circ g$ is a submersion at p. Therefore X is a submanifold near p, and hence everywhere, of codimension $n - m$. □

In particular, the intersection $X = M \pitchfork W$ of two transverse submanifolds $M, W \subset N$ is a submanifold with $\operatorname{codim} X = \operatorname{codim} M + \operatorname{codim} W$.

5.7.4. Thom's Transversality Theorem. We now state a general theorem, that will allow us to construct many transverse maps. Let M, S, N be manifolds of arbitrary dimension.

Theorem 5.7.5. *Let $F\colon M \times S \to N$ be a smooth map. If F is transverse to some submanifold $Z \subset N$, then $F_s = F(\cdot, s)\colon M \to N$ is also transverse to Z for almost every $s \in S$.*

We mean as usual that the thesis holds for all the values $s \in S$ that lie outside of some zero measure subset.

Proof. Since $F \pitchfork Z$, the preimage $W = F^{-1}(Z) \subset M \times S$ is a smooth submanifold. Consider the projection $\pi\colon M \times S \to S$ and particularly its restriction $\pi|_W\colon W \to S$. We now claim that if s is a regular value for $\pi|_W$ then $F_s \pitchfork Z$. From this we conclude: by Sard's Lemma almost every $s \in S$ is a regular value for $\pi|_W$.

Consider a point $(p, s) \in W$. Since s is regular for $\pi|_W$ we have

$$T_{(p,s)}W + T_{(p,s)}(M \times \{s\}) = T_{(p,s)}(M \times S).$$

Since $F \pitchfork Z$ we have

$$dF_{(p,s)}\big(T_{(p,s)}(M \times S)\big) + T_{F(p,s)}Z = T_{F(p,s)}N.$$

By combining the two equations we get

$$\begin{aligned} T_{F(p,s)}N &= dF_{(p,s)}(T_{(p,s)}W) + dF_{(p,s)}\big(T_{(p,s)}(M \times \{s\})\big) + T_{F(p,s)}Z \\ &= dF_{(p,s)}\big(T_{(p,s)}(M \times \{s\})\big) + T_{F(p,s)}Z \\ &= d(F_s)_p(T_pM) + T_{F(p,s)}Z. \end{aligned}$$

In the second equality we have eliminated the first addendum since it is contained in the third. We have proved that $F_s \pitchfork Z$. $\square$

5.7.5. Consequences. We now draw some consequences from Thom's Transversality Theorem. Here is an amazingly simple application.

Corollary 5.7.6. *Let M be a manifold and $f: M \to \mathbb{R}^n$ be a smooth map. Let $Z \subset \mathbb{R}^n$ be a submanifold. For almost all $s \in \mathbb{R}^n$, the translated map*

$$f_s(p) = f(p) + s$$

is transverse to Z.

Proof. The map $F: M \times \mathbb{R}^n \to \mathbb{R}^n$, $F(p,s) = f(p) + s$ is a submersion and is hence clearly transverse to any submanifold $Z \subset \mathbb{R}^n$. So Thom's Transversality Theorem applies. $\square$

Corollary 5.7.7. *Let $M, N \subset \mathbb{R}^n$ be any two submanifolds. For almost every $s \in \mathbb{R}^n$ the translate $M + s$ and N are transverse.*

Here is a perturbation theorem for a map between two arbitrary manifolds.

Corollary 5.7.8. *Let $f: M \to N$ be a smooth map between manifolds and $W \subset N$ be a submanifold. There is a $g: M \to N$ homotopic to f that is transverse to W.*

Proof. Consider N embedded in some $\mathbb{R}^n$ and pick a tubular neighbourhood $\nu N \subset \mathbb{R}^n$ of N with projection $\pi: \nu N \to N$. Using a partition of unity, pick a smooth positive function $r: N \to \mathbb{R}$ such that $B(q, r(q)) \subset \nu N$ for every $q \in N$. We define the map

$$F: M \times B^n \longrightarrow N, \quad F(p,s) = \pi\big(f(p) + r(f(p))s\big).$$

Here $B^n \subset \mathbb{R}^n$ is the unit ball as usual. The map F is a submersion and is hence transverse to any $W \subset N$. Therefore for some $s \in B^n$ the map $g = F_s$ is transverse to W and is homotopic to f through F_{ts}. $\square$

5.7.6. Stability. If we perturb a map (with compact domain) that is transverse, it keeps being transverse: transversality is a *stable property*. We introduce this concept in more generality.

A property P of a smooth map $f: M \to N$ is *stable* if for every smooth homotopy $f_t: M \to N$, $t \in \mathbb{R}$ with $f_0 = f$ there is an $\varepsilon > 0$ such that all the maps f_t with $|t| < \varepsilon$ share the property P.

Proposition 5.7.9. *Let M be compact and $f: M \to N$ a smooth map. The following properties are stable for f:*

- *f is an immersion,*
- *f is a submersion,*
- *f is an embedding,*
- *f is transverse to a fixed closed submanifold $W \subset N$.*

Proof. Consider

$$F: M \times \mathbb{R} \longrightarrow N \times \mathbb{R}, \qquad F(x,t) = \big(f_t(x), t\big).$$

The map f_t is an immersion or submersion at $p \in M \iff F$ is an immersion or submersion at (p,t). Written in coordinates, this is an open condition, hence it holds on a neighbourhood of $M \times \{0\} \subset M \times \mathbb{R}$, which contains $M \times (-\varepsilon, \varepsilon)$ since M is compact.

Suppose that f is an embedding. Then f_t is an immersion for $t \in (-\varepsilon, \varepsilon)$. We prove that, after possibly taking a smaller $\varepsilon > 0$, each f_t with $t \in (-\varepsilon, \varepsilon)$ is injective: if not, there are sequences $t_i \to 0$, $p_i, q_i \in M$ with $f_{t_i}(p_i) = f_{t_i}(q_i)$. Since M is compact we may suppose that $p_i \to p$ and $q_i \to q$. Since f is injective we have $p = q$. This gives a contradiction because F is an immersion at $(p, 0)$ and is hence injective on a small neighbourhood. Finally, injective immersions are embeddings because M is compact (again).

Stability of transversality is similar and left as an exercise. $\square$

We warn the reader that being an embedding is a stable property (when the base manifold is compact), while being only injective is not! Consider

$$f_t(x): \mathbb{R} \longrightarrow \mathbb{R}, \qquad f_t(x) = (x^2 - t^2)x.$$

Here f_0 is injective while f_t is not so for any $t \neq 0$. One can use this homotopy to construct another homotopy $f_t: S^1 \to S^1$ where f_0 is a homeomorphism and f_t is not injective for any $t \neq 0$. Of course f_0 is not a diffeomorphism, since there must be a $p \in S^1$ with trivial df_p.

5.8. The Ehresmann Theorem

We prove here a theorem that promotes every proper submersion to a fibre bundle. We then apply it to the definition of some new fibre bundles.

5.8.1. The theorem. A fibre bundle is a submersion, but the converse is often not true: if pick a fibre bundle and remove a random closed subset from the domain, we still get a submersion that is probably not a fibre bundle. The converse is however guaranteed if the map is proper.

Theorem 5.8.1 (Ehresmann). *Every proper submersion is a fibre bundle.*

We start with a lemma.

Lemma 5.8.2. *Let $f\colon M \to N$ be a submersion. For every vector field $Y \in \mathfrak{X}(N)$, there is one $X \in \mathfrak{X}(M)$ that is f-related to Y.*

Proof. Every p has an open neighbourhood $U(p)$ where the submersion looks like a projection $\mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, and it is easy to construct a vector field $X_p \in \mathfrak{X}(U(p))$ that is f-related to $Y|_{f(U)}$. We take a partition of unity $\{\rho_p\}$ subordinate to the covering $\{U(p)\}$ and define

$$X(q) = \sum \rho_p(q) X_p(q).$$

We get $df_q(X(q)) = \sum \rho_p(q)Y(f(q)) = Y(f(q))$ and we are done. □

We leave the following as an exercise.

Exercise 5.8.3. Let $f\colon M \to N$ be a proper map, and $X \in \mathfrak{X}(M)$, $Y \in \mathfrak{X}(N)$ two f-related vector fields. The field X is complete $\Longleftrightarrow$ Y is.

Proof of the Ehresmann Theorem. Let $f\colon M \to N$ be a proper submersion. Since the theorem is local in the codomain, we can suppose that $N = \mathbb{R}^n$.

Let $X_1, \ldots, X_n$ be fields in M that are f-related to $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$. These are complete by Exercise 5.8.3. Let $\Phi^1, \ldots, \Phi^n$ be their flows. The map

$$\begin{aligned} f^{-1}(0) \times \mathbb{R}^n &\longrightarrow M \\ (p, x^1, \ldots, x^n) &\longmapsto \Phi^n_{x^n}(\cdots(\Phi^1_{x^1}(p))\cdots) \end{aligned}$$

is a diffeomorphism with inverse

$$q \longmapsto (\Phi^1_{-x^1}(\cdots(\Phi^n_{-x^n}(p))\cdots), f(q))$$

where $f(q) = (x^1, \ldots, x^n)$. Therefore f is a fibration. □

Corollary 5.8.4. *A submersion between compact manifolds is a fibre bundle.*

This corollary is quite useful, because proving that a map is a submersion is much easier than verifying that it is a fibre bundle.

5.8.2. The Hopf fibration. Consider the 3-sphere in $\mathbb{C}^2$, written as

$$S^3 = \{(w, z) \in \mathbb{C}^2 \mid |w|^2 + |z|^2 = 1\}.$$

The *Hopf fibration* is the map

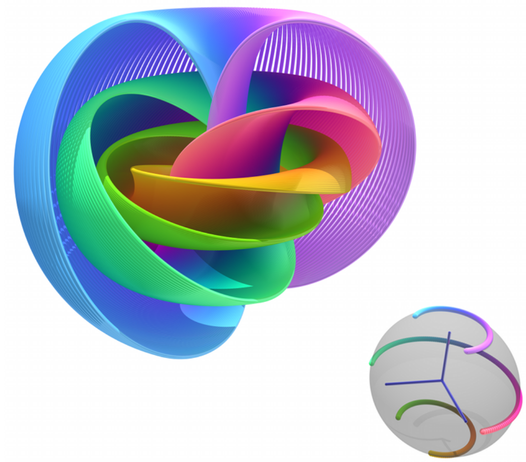

Figure 5.10. Some fibers of the Hopf fibration, projected stereographically on $\mathbb{R}^3$. The counterimage of the circular arcs in S^2 shown on the bottom are portions of tori foliated into circles n 3-space.

$$S^3 \longrightarrow \mathbb{CP}^1$$
$$(w, z) \longmapsto [w, z]$$

Since $\mathbb{CP}^1$ is diffeomorphic to S^2, we may consider it as a map $S^3 \to S^2$.

Exercise 5.8.5. The Hopf fibration is a submersion, and hence a fibration by Ehresmann's Theorem.

If $(w, z) \in S^3$, we find easily that

$$f^{-1}(f(w, z)) = \left\{(e^{i\theta}w, e^{i\theta}z) \mid e^{i\theta} \in S^1\right\}$$

and hence the fibre of the Hopf fibration is a geometric circle. The Hopf fibration is a curious fibration $S^3 \to S^2$ with fibres consisting of circles. It is clearly not trivial, since the total space is not diffeomorphic to $S^2 \times S^1$. We also deduce the non intuitive fact that S^3 may be foliated into circles!

Exercise 5.8.6. The stereographic projection $S^n \setminus \{N\} \to \mathbb{R}^n$ sends circles disjoint from N to circles, and circles containing N to lines.

We deduce that $\mathbb{R}^3$ may be foliated into 1-dimensional submanifolds that consist of one single line and infinitely many circles. Can you visualize it? A portion is shown in Figure 5.10. The line there is vertical.

We can define a Hopf bundle $S^{2n+1} \to \mathbb{CP}^n$ with fibre S^1 for every n.

5.9. Exercises

Exercise 5.9.1. Let X be a vector field on M. Let $\gamma\colon \mathbb{R} \to M$ be an integral curve and $p \in M$ a point such that $\lim_{t\to+\infty} \gamma(t) = p$. Show that $X(p) = 0$.

Exercise 5.9.2. Construct a nowhere-vanishing vector field on each lens space $L(p, q)$.

Exercise 5.9.3. Let M be a manifold and $X, Y \in \mathfrak{X}(M)$ vector fields. Prove that

$$[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_{[X,Y]}.$$

This is an equality of operators on $\Gamma(\mathcal{T}_k^h(M))$. The bracket $[A, B]$ of two such operators is by definition $[A, B] = AB - BA$. Note that if $(h, k) = (1, 0)$ this is equivalent to the Jacobi equality on vector fields.

Exercise 5.9.4. Construct a foliation on the torus $T = S^1 \times S^1$ that has both compact and non-compact leaves.

Exercise 5.9.5. Let D be a rank-k distribution on a manifold M. Show that D is integrable if and only if the following holds: for every $p \in M$ there is a k-submanifold $S \subset M$ containing p with $T_qS = D_q$ for all $q \in S$.

Exercise 5.9.6. Show that each lens space $L(p, q)$ has a foliation in circles.

Exercise 5.9.7. Consider $S^3 = \big\{(z_1, z_2) \in \mathbb{C}^2 \mid |z|^2 + |w|^2 = 1\big\}$. For every $p = (z_1, z_2) \in S^3$, pick the complex line

$$r_p = \big\{(w_1, w_2) \in \mathbb{C}^2 \;\big|\; w_1\bar{z}_1 + w_2\bar{z}_2 = 0\big\}.$$

(1) Show that $r_p \subset T_pS^3$. Therefore $\{r_p\}_{p\in S^3}$ is a plane distribution in S^3 called the *Hopf distribution*.
(2) Is the Hopf distribution integrable?

Exercise 5.9.8. Show that two embedding $f, g\colon \mathbb{R} \hookrightarrow \mathbb{R}^2$ are always isotopic.

Exercise 5.9.9. Let M be a connected manifold. Let $N \subset M$ be a closed hypersurface. Show that $M \setminus N$ has either one or two connected components. Describe some examples in both cases.

Exercise 5.9.10. Let $f : S^1 \hookrightarrow \mathbb{R}^3$ be a knot (that is, a smooth embedding). Show that there is an affine plane $P \subset \mathbb{R}^3$ such that $\pi \circ f : S^1 \hookrightarrow P$ is an immersion, where π is the orthogonal projection onto P.

Exercise 5.9.11. Let $X = M \pitchfork W$ be the transverse intersection of two submanifolds $M, W \subset N$. Show that every point $p \in X$ has a neighbourhood U in N and a chart $\varphi\colon U \to \mathbb{R}^{w-x} \times \mathbb{R}^x \times \mathbb{R}^{m-x}$ that transforms $W \cap U$ and $M \cap U$ into $\mathbb{R}^{w-x} \times \mathbb{R}^x \times \{0\}$ and $\{0\} \times \mathbb{R}^x \times \mathbb{R}^{m-x}$ with $m = \dim M, w = \dim W, x = \dim X$.

CHAPTER 6

Cut and paste

> Cutting and gluing are simple geometrical constructions which, given some smooth manifolds (possibly with boundaries or corners) and additional data where necessary, give rise to new manifolds. On account of their perspicuity, these methods were much used in the days of topology of surfaces, and they remain a very powerful tool
>
> C. T. C. Wall, 1960

In this chapter we address the following question: how can we construct new smooth manifolds? The most effective techniques known consist in building more complicated smooth manifolds out of simpler pieces, glued altogether along smooth maps. A piece is usually a *manifold with boundary*, and the pieces are glued along (portions of) their boundaries.

Among all kinds of decompositions of manifolds into simple pieces, a prominent role is played by *handle decompositions*, some very general constructions that may be used to build any compact smooth manifold in any dimension, tightly related to the theory of *Morse functions*. We will then use handle decompositions to classify all compact surfaces.

6.1. Manifolds with boundary

We introduce a variation of the definition of smooth manifold that allows the presence of some particular *boundary points*. This is a very natural notion and is present everywhere in differential topology and geometry.

Most of the definitions and theorems about smooth manifolds also apply to manifolds with boundary, with appropriate modifications.

6.1.1. Definition. Consider the upper half-space

$$\mathbb{R}^n_+ = \left\{x \in \mathbb{R}^n \;\middle|\; x_n \geq 0\right\}$$

in $\mathbb{R}^n$. Its *boundary* is the horizontal hyperplane $\partial\mathbb{R}^n_+ = \{x_n = 0\}$, while its *interior* is the open subset $\mathbb{R}^n_+ \setminus \partial\mathbb{R}^n_+ = \{x_n > 0\}$.

We now redefine the notions of charts and atlases in a more general context that allows the presence of boundary points: everything will be like in Section 3.1.1, only with $\mathbb{R}^n_+$ replacing $\mathbb{R}^n$.

Let M be a topological space. A $\mathbb{R}^n_+$-*chart* is a homeomorphism $\varphi\colon U \to V$ from an open set $U \subset M$ onto an open set $V \subset \mathbb{R}^n_+$. A *smooth* $\mathbb{R}^n_+$-*atlas* in

M is a set $\{\varphi_i\}$ of $\mathbb{R}^n_+$-charts with $\cup U_i = M$ such that the transition maps φ_{ij} are smooth where they are defined. Note that the domain of φ_{ij} is an open subset of $\mathbb{R}^n_+$ and may not be open in $\mathbb{R}^n$, so the correct notion of smoothness is that stated in Definition 3.3.5.

Definition 6.1.1. A *smooth manifold with boundary* is a Haussdorff second-countable topological space M equipped with a smooth $\mathbb{R}^n_+$-atlas.

We will drop the symbol $\mathbb{R}^n_+$ from the notation. As in Section 3.1.1, two compatible atlases are meant to give the same smooth structure.

6.1.2. The boundary. Let M be a smooth manifold with boundary. The points $p \in M$ that are sent to $\partial\mathbb{R}^n_+$ via some chart form the *boundary* ∂M. There is no possible ambiguity here, since if one chart sends p inside $\partial\mathbb{R}^n_+$, then all charts do (exercise).

The boundary ∂M is naturally a $(n-1)$-dimensional smooth manifold *without* boundary. Indeed by restricting the charts to ∂M we get an atlas for ∂M with values onto some open sets of the hyperplane $\partial\mathbb{R}^n_+$, that we identify with $\mathbb{R}^{n-1}$ in the obvious way.

Example 6.1.2. Every open subset $U \subset \mathbb{R}^n_+$ is a smooth manifold with boundary $\partial U = U \cap \partial\mathbb{R}^n_+$. The atlas consists of just the identity chart.

The *interior* of M is $\operatorname{int}(M) = M \setminus \partial M$. It is a manifold without boundary.

6.1.3. Maps. The notions of smooth maps and diffeomorphisms extend to manifolds with boundary without any modification. When we have a smooth map $f\colon M \to N$ between manifolds with boundary, a boundary or interior point of M may be sent to a boundary or interior point of N: all four combinations may arise, and the reader is invited to construct examples of all four types.

A diffeomorphism $f\colon M \to N$ between two manifolds with boundary restricts to a diffeomorphism $f\colon \partial M \to \partial N$ of their boundaries.

6.1.4. Regular domains. We now describe one important source of examples. Let M be a smooth n-manifold without boundary.

Definition 6.1.3. A *regular domain* is a subset $D \subset M$ such that for every $p \in D$ there is a chart $\varphi\colon U \to V$ with $p \in U$ and $V \subset \mathbb{R}^n$ that sends $U \cap D$ onto $V \cap \mathbb{R}^n_+$.

Every regular domain D has a natural structure of manifold with boundary, obtained by taking as an atlas all the charts φ of this type. The boundary ∂D is a codimension-1 submanifold of M.

Exercise 6.1.4. For every $a < b$, the closed segment $[a, b]$ is a domain in $\mathbb{R}$ and hence a manifold with boundary; the boundary consists of the points a and b.

Here is a concrete way to construct regular domains:

Proposition 6.1.5. *Let M be a manifold without boundary and $f\colon M \to \mathbb{R}$ a smooth function. If y_0 is a regular value, then $D = f^{-1}(-\infty, y_0]$ is a regular domain with $\partial D = f^{-1}(y_0)$.*

Proof. Consider a point $p \in D$. If $f(p) < y_0$, the point p has an open neighbourhood fully contained in D that can be sent inside the interior of $\mathbb{R}^n_+$ via some chart.

If $f(p) = y_0$, by Proposition 3.8.10 there are charts $\varphi\colon U \to \mathbb{R}^n$ and $\psi\colon W \to \mathbb{R}$ with $p \in U$, $\varphi(p) = 0$, $\psi(y_0) = 0$, $f(U) \subset W$ such that $\psi \circ f \circ \varphi^{-1}(x_1, \ldots, x_n) = x_n$ and up to composing with a reversion we may also require that ψ is orientation-reversing. Therefore $\varphi(U \cap D) = \mathbb{R}^n_+$. □

Corollary 6.1.6. *The unit disc*

$$D^n = \left\{ x \in \mathbb{R}^n \;\middle|\; \|x\| \leq 1 \right\}$$

is a domain in $\mathbb{R}^n$ with boundary $\partial D^n = S^{n-1}$.

Proof. We pick $f(x) = \|x\|^2$ and get $D^n = f^{-1}(-\infty, 1]$. Every non-zero value is regular for f. □

Remark 6.1.7. The square $[-1, 1] \times [-1, 1]$ is not a regular domain in $\mathbb{R}^2$, because it has corners. More generally, the product $M \times N$ of two manifolds with boundary is *not* necessarily a manifold with boundary, because if $\partial M \neq \varnothing$ and $\partial N \neq \varnothing$ then some corners arise. However, if $\partial M = \varnothing$ then $M \times N$ is naturally a manifold with boundary and

$$\partial(M \times N) = M \times \partial N.$$

For instance, the cylinder $S^1 \times [-1, 1]$ is a surface with boundary, and the boundary consists of the two circles $S^1 \times \{\pm 1\}$. More generally $S^m \times D^n$ is a manifold with boundary and

$$\partial\big(S^m \times D^n\big) = S^m \times S^{n-1}.$$

6.1.5. Tangent space. The definition of tangent space via derivations also extends *verbatim* to manifolds with boundary.[1] For every point $p \in \mathbb{R}^n_+$, included those on the boundary, we get $T_p\mathbb{R}^n_+ = \mathbb{R}^n$. For a general n-manifold M with boundary, the space T_pM is a n-dimensional vector space $\forall p \in M$.

At every boundary point $p \in \partial M$ the tangent space $T_p\partial M$ is naturally a hyperplane inside T_pM, that divides T_pM into two components, the "interior" and "exterior" tangent vectors, according to whether they point towards the interior of M or the exterior. This subdivision between interior and exterior is obvious in $\mathbb{R}^n_+$ and transferred to M unambiguously via charts.

As in the boundaryless case, every smooth map $f\colon M \to N$ induces a differential $df_p\colon T_pM \to T_{f(p)}N$ at every point $p \in M$.

[1] The definition through curves would need some modifications, so we just abandon it.

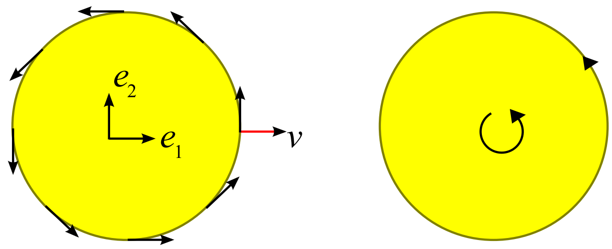

Figure 6.1. The canonical orientation on the disc (given by the canonical basis e_1, e_2) induces the counterclokwise orientation on the boundary circle (left). We may write conveniently the orientations on a surface and on a curve using (curved) arrows (right)

6.1.6. Orientation. The notion of an orientation for a manifold M also extends *as is* to the boundary case, either as a locally coherent assignment of orientations on the tangent spaces T_pM, or equivalently as an oriented atlas. One nice additional feature is that an orientation on M induces one on its boundary ∂M, as we now explain.

Let M be an oriented manifold with boundary of dimension $n \geq 2$. For every $p \in \partial M$, we choose an exterior vector $v \in T_pM$ and note that

$$T_pM = \text{Span}(v) \oplus T_p\partial M.$$

With this subdivision, the orientation on T_pM induces one on $T_p\partial M$: we say that a basis $v_2. \ldots . v_n$ for $T_p\partial M$ is positive $\Longleftrightarrow$ the basis $v, v_2, \ldots, v_n$ is positive for T_pM.[2] By looking on a chart we see that this is a locally coherent assignment that does not depend on the choice of the exterior vector v.

We now consider the one-dimensional case, that is slightly different. First, we define an *orientation* on a point to be the assignment of a sign ± 1. When not mentioned, a point is always equipped with the $+1$ orientation: points are in fact the only manifolds that have a canonical orientation!

If M is an oriented 1-manifold, we orient every boundary point $p \in \partial M$ as $+1$ or -1 depending on whether the vectors pointing outside in the line T_pM are positive or negative.

Every domain in $\mathbb{R}^n$ is naturally oriented by the canonical basis $e_1, \ldots, e_n$, so for instance the disc D^n has a canonical orientation. This canonical orientation induces an orientation on the boundary sphere S^{n-1}. The case $n = 2$ is shown in Figure 6.1.

6.1.7. Immersions, embeddings, submanifolds. Let M, N be manifolds with boundary. We define an *immersion* as usual as a map $f : M \to N$ with injective differentials, and then an *embedding* as an injective immersion $f : M \to N$ that is a homeomorphism onto its image.

[2]This is not the only possible choice one could make, and is usually called the *outward-first* convention.

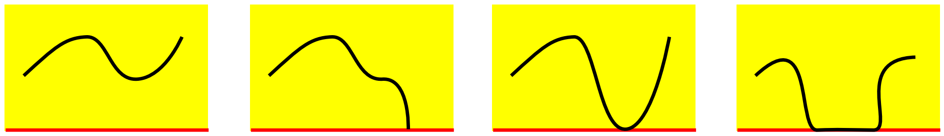

Figure 6.2. Different kinds of compact 1-dimensional submanifolds inside the half-plane $\mathbb{R}^2_+$.

Definition 6.1.8. Let N be a manifold. A *submanifold* is the image of an embedding $f\colon M \hookrightarrow N$.

The reader should note that, as opposite to Definition 3.7.1, we are not saying that a submanifold should look locally like some simple model. This is by far not the case here: Figure 6.2 shows that many different kinds of local configurations arise already when one embeds a segment in the half-plane $\mathbb{R}^2_+$. In higher dimensions things may also get more complicated.

In some cases, we may require the submanifold to satisfy some requirements. For instance, a submanifold $M \subset N$ is *neat* if

- $\partial M = M \cap \partial N$, and
- M meets ∂N transversely, that is at every $p \in \partial M$ we have $T_pM + T_p\partial N = T_pN$.

None of the embedded submanifolds in Figure 6.2 is neat.

6.1.8. Fibre bundles and vector bundles. The theory of bundles extends harmlessly to manifolds with boundary with minor obvious modifications. On a fibre bundle $E \to M$, we can allow M to have boundary, and in that case the trivialising neighborhoods will be diffeomorphic to open subsets of $\mathbb{R}^n_+$, or we can allow the fibre F to have boundary; however, we do not admit *both* M and F to have boundary, because some corners would arise and E would not be a smooth manifold.

In particular the theory of vector bundles also work on manifolds M with boundary. Every manifold M has its tangent bundle TM together with all the other tensor bundles, so we can talk about vector fields and tensor fields in M.

6.1.9. Vector fields. The behaviour of a vector field $X \in \mathfrak{X}(M)$ at the boundary ∂M is often important. We say that X is *tangent*, *pointing outward*, or *inward* at the boundary if $X(p)$ is so at every boundary point $p \in \partial M$.

If a vector field $X \in \mathfrak{X}(M)$ is tangent to the boundary, many facts on vector fields that we proved in the previous chapters are still valid for X with only minor obvious modifications: every point is contained in a maximal integral curve, if X is complete we get a flow on M, and if M is compact then X is necessarily complete. A flow on M restricts to a flow on ∂M. We can define the Lie derivative $\mathcal{L}_X$ of any tensor field along X.

If X is not tangent to the boundary the notions of integral curve and flow are evidently more problematic, but it still possible to define the Lie derivative $\mathcal{L}_X$. Details are left to the reader.

6.1.10. Homotopy, isotopy, ambient isotopy. The notions of smooth homotopy, isotopy, and ambient isotopy extend *verbatim* to manifolds with boundary. An ambient isotopy of a manifold M with boundary restricts to an ambient isotopy of its boundary. An appropriate isotopy extension theorem holds in this context, where the target manifold is allowed to have boundary as long as the isotopy does not cross it:

Theorem 6.1.9. *Let M be compact without boundary, and let N may possibly have boundary. Let $f, g\colon M \to \operatorname{int}(N)$ be embbeddings that are isotopic through embeddings $f_t\colon M \to \operatorname{int}(N)$. There is an ambient isotopy of N relating f and g which is constantly the identity on ∂N.*

Proof. The same proof of Theorem 5.3.3 applies; it suffices to stay away from the boundary ∂N in all the arguments, so that the vector field Y will be constantly vertically tangent to it. $\square$

6.1.11. The unit disc bundle. Let $E \to M$ be a vector bundle over a manifold M without boundary. Fix a Riemannian metric g for E. The *unit disc bundle* is the submanifold with boundary

$$D(E) = \{ v \in E \mid \|v\| \leq 1 \}.$$

The projection π restricts to a projection $\pi\colon D(E) \to M$ and one sees as in Proposition 4.5.7 that this is a disc bundle (a fibre bundle with $F = D^k$) and that it does not depend on g up to isotopy (that is, up to an isomorphism of $E \to M$ that is isotopic to the identity).

The boundary of $D(E)$ is the unit sphere bundle $S(E)$, already considered in Section 4.5.5. The interior of $D(E)$ may be given a bundle structure isomorphic to $E \to M$ by shrinking E.

6.1.12. Closed tubular neighbourhoods. Let $N \subset \operatorname{int}(M)$ be a compact submanifold without boundary contained in the interior of a manifold M possibly with boundary. We know that the submanifold N has a tubular neighbourhood νN in the interior $\operatorname{int}(M)$ of M.

Definition 6.1.10. A *closed tubular neighbourhood* of N in M is the unit disc bundle of any tubular neighbourhood νN of N.

To better distinguish a tubular neighbourhood from a *closed* tubular neighbourhood, we can call the first an *open* tubular neighbourhood. We will use the notation $\bar{\nu} N$ for a closed tubular neighbourhood; note that the interior of a closed tubular neighbourhood may in turn be given the structure of an open tubular neighbourhood, so one can switch easily from open to closed and back.

A closed tubular neighbourhood is a fibre bundle with compact base N and compact fibre D^k, and is hence also compact (exercise). For this reason it is sometimes better to work with closed tubular neighbourhoods; for instance, we may promote isotopy to ambient isotopy:

Theorem 6.1.11. *Let M be a manifold possibly with boundary and $N \subset \mathrm{int}(M)$ be a compact submanifold without boundary. The submanifold N has a unique closed tubular neighbourhood $\bar{\nu}N$ up to ambient isotopy in M.*

Proof. We already know that tubular neighbourhoods are isotopic, and hence also the closed tubular neighbourhoods are. Since these are compact, the isotopy may be promoted to an ambient isotopy. □

The same results apply if M has boundary, as long as N is entirely contained in the interior of M.

6.1.13. Collar. Let M be a manifold with boundary, and N be the union of some connected components of ∂M. A *collar* of N in M is an embedding

$$i \colon N \times [0, 1) \hookrightarrow M$$

such that $i(p, 0) = p$ for every $p \in N$. The collars should be interpreted as the tubular neighbourhoods of the boundary.

Proposition 6.1.12. *The manifold N has a unique collar up to isotopy.*

The proof is the same as that for tubular neighbourhoods, and we omit it. A *closed collar* is the restriction of an open collar to $N \times [0, 1/2]$. If N is compact, the closed collar is unique up to ambient isotopy.

Exercise 6.1.13. For every manifold M the inclusion $\mathrm{int}(M) \hookrightarrow M$ is a homotopy equivalence.

Hint. Use a collar for ∂M to define the homotopy inverse. □

6.1.14. One-dimensional manifolds. We leave to the reader to solve the following exercise, that fully classifies all connected one-dimensional manifolds.

Exercise 6.1.14. Every connected one-dimensional manifold is diffeomorphic to one of the following:

$$S^1, \quad (0, 1), \quad [0, 1), \quad [0, 1].$$

In particular S^1 is the unique connected compact one-dimensional manifold without boundary.

6.1.15. Discs. Let M be a n-manifold, possibly with boundary. A *disc* in M is an embedding $f\colon D^n \hookrightarrow \operatorname{int}(M)$. It may be seen as a closed tubular of $f(0)$. We can now prove this remarkable theorem.

Theorem 6.1.15 (The Disc Theorem). *Let M be a connected smooth n-manifold, possibly with boundary. Two discs $f,g\colon D^n \hookrightarrow \operatorname{int}(M)$ are always ambiently isotopic, possibly after pre-composing g with a reflection.*

Proof. We argue as in Proposition 5.6.5 using Theorem 6.1.11. □

With a little abuse we sometimes call a *disc* the image of an embedding $f\colon D^n \hookrightarrow M$. With this interpretation, which disregards the parametrisation, two discs are always ambiently isotopic. The reader should appreciate how powerful this theorem is, already in the only apparently simpler case $M = \mathbb{R}^n$, for instance in dimension $n = 2$.

The Disc Theorem was proved by Palais in 1960.

6.1.16. Spheres. We end this section by describing how every sphere decomposes beautifully into two simple submanifolds with boundary.

For every $0 < k < n$ we identify $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$ and write a point of $\mathbb{R}^n$ as (x,y) with $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^{n-k}$. By radial expansion we may easily construct a homeomorphism between D^n and $D^k \times D^{n-k}$, which restricts to a homeomorphism between S^{n-1} and the topological boundary of $D^k \times D^{n-k}$. (Recall that $D^k \times D^{n-k}$ is not a smooth manifold because its boundary has corners.) The latter in turn decomposes into two closed subsets

$$S^{k-1} \times D^{n-k}, \qquad D^k \times S^{n-k-1}$$

whose intersection is $S^{k-1} \times S^{n-k-1}$. Having understood this simple topological phenomenon, we write an analogous decomposition of S^{n-1} in the smooth setting. We write

$$S^{n-1} = \big\{(x,y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} \bigm| \|x\|^2 + \|y\|^2 = 1\big\}.$$

We now consider the subsets

$$A = \big\{(x,y) \in S^{n-1} \bigm| \|x\|^2 \leq \tfrac{1}{2}\big\}, \quad B = \big\{(x,y) \in S^{n-1} \bigm| \|y\|^2 \leq \tfrac{1}{2}\big\}.$$

These are both domains, since $\frac{1}{2}$ is a regular value for the maps $(x,y) \mapsto \|x\|^2$ or $\|y\|^2$ restricted on S^{n-1} (exercise). The common boundary

$$A \cap B = \big\{(x,y) \in S^{n-1} \bigm| \|x\|^2 = \|y\|^2 = \tfrac{1}{2}\big\}$$

is diffeomorphic to $S^{k-1} \times S^{n-k-1}$ via the map $(x,y) \mapsto (\sqrt{2}x, \sqrt{2}y)$. We now identify the domains: the map

$$A \longrightarrow D^k \times S^{n-k-1}, \quad (x,y) \longmapsto \left(\sqrt{2}x, \frac{y}{\|y\|}\right)$$

is a diffeomorphism, with inverse $(x,y) \mapsto \frac{\sqrt{2}}{2}\big(x, \sqrt{2-\|x\|^2}y\big)$. We have an analogous diffeomorphism between B and $S^{k-1} \times D^{n-k}$.

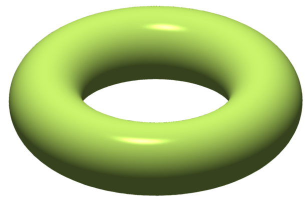

Figure 6.3. A solid torus is a 3-manifold diffeomorphic to $D^2 \times S^1$. The stereographic projection sends the solid torus $B \subset S^3$ to the standard one shown here. Its complement in S^3 is also a solid torus A. Can you see it?

We have discovered that S^{n-1} decomposes into two domains $A \cong D^k \times S^{n-k-1}$ and $B \cong S^{k-1} \times D^{n-k}$ with common boundary $S^{k-1} \times S^{n-k-1}$. We also note that A and B are closed tubular neighborhoods of the spheres

$$S^{n-k-1} \times \{0\} \qquad \{0\} \times S^{k-1}.$$

6.1.17. The 3-sphere. We analyse the 3-sphere with more details. The discussion above shows that S^3 decomposes into $A \cong S^1 \times D^2$ and $B \cong D^2 \times S^1$ along a middle torus $A \cap B \cong S^1 \times S^1$. The stereographic projection

$$\varphi(x, y, z, w) = \frac{2}{1-w}(x, y, z)$$

sends the middle torus

$$A \cap B = \left\{ \frac{\sqrt{2}}{2}(\cos\theta, \sin\theta, \cos\varphi, \sin\varphi) \right\}$$

to the subset

$$\left\{ \frac{2}{\sqrt{2} - \sin\varphi}(\cos\theta, \sin\theta, \cos\varphi) \right\}.$$

A simple computation shows that this is the standard torus of Example 3.8.6 with parameters $a = 2$ and $b = 2\sqrt{2}$. The interior of a standard torus is called a *solid torus* and is the domain $B \cong D^2 \times S^1$ stereographically projected into $\mathbb{R}^3$, see Figure 6.3. Its complement is another solid torus A. The 3-sphere S^3 decomposes into two solid tori.

6.2. Cut and paste

We now introduce some basic cut and paste manipulations that allow to modify the topology of a smooth manifold.

6.2.1. Punctures. Let M be a connected smooth n-manifold, possibly with boundary. The simplest topological modification we can make on M is to remove a point $p \in \text{int}(M)$. By Corollary 5.3.4, the new manifold $M' = M \setminus \{p\}$ does not depend (up to diffeomorphism) on p, and we say that it is obtained by *puncturing* M.

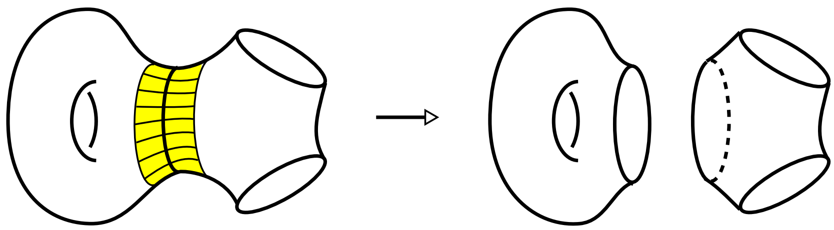

Figure 6.4. How to cut a manifold along a two-sided hypersurface.

A variation of this modification consists of picking a disc $D \subset \text{int}(M)$ and removing its interior: the new manifold

$$M'' = M \setminus \text{int}(D)$$

has the same boundary components as M, plus one new sphere ∂D. The manifold M'' does not depend (up to diffeomorphisms) on the chosen disc D by the Disc Theorem 6.1.15.

Exercise 6.2.1. The manifolds M' and $M'' \setminus \partial D$ are diffeomorphic.

Exercise 6.2.2. If $M = S^n$, we get $M' \cong \mathbb{R}^n$ and $M'' \cong D^n$.

Exercise 6.2.3. If $M = D^n$ then $M'' \cong S^{n-1} \times [-1, 1]$.

Exercise 6.2.4. If $\dim M \geq 3$ then $\pi_1(M') \cong \pi_1(M'') \cong \pi_1(M)$.

Hint. Use Van Kampen. □

6.2.2. Cutting along submanifolds. We now extend the above manipulation from points to arbitrary compact submanifolds.

Let M be a smooth manifold and $N \subset \text{int}(M)$ a compact submanifold of some codimension $k \geq 1$. The complement $M' = M \setminus N$ is a new manifold. As above, a variation consists in taking a closed tubular neighbourhood $\bar{\nu}N$ and considering

$$M'' = M \setminus \text{int}(\bar{\nu}N).$$

The manifold M'' has a new compact boundary component $\partial\bar{\nu}N$, which is a S^{k-1}-bundle over N. The manifold M'' only depends on N and not on the tubular neighbourhood $\bar{\nu}N$ since it is unique up to ambient isotopy.

This operation is particularly interesting if N has codimension 1 and is *two-sided*, that is has trivial normal bundle $\nu N \cong N \times \mathbb{R}$. For instance, this holds if both M and N are orientable: see Proposition 5.6.7. In this case the new manifold M'' has two new boundary components, both diffeomorphic to N. See Figure 6.4. We say that M'' is obtained by *cutting* M along N.

Example 6.2.5. By cutting S^n along its equator S^{n-1} we get two discs.

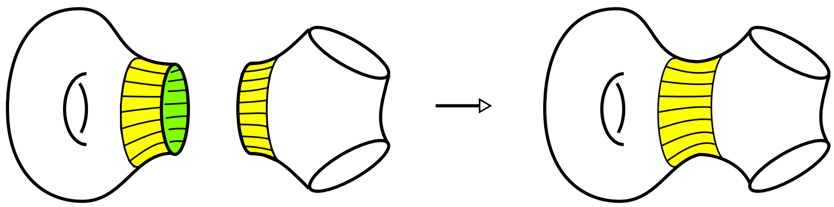

Figure 6.5. How to paste two boundary components N_1 and N_2 via a diffeomorphism φ. To get a new smooth manifold, we pick two collars and we make them overlap.

If $N \subset M$ are both connected and N has codimension one, the new manifold M'' may be connected or not; in the first case, we say that N is *non-separating*, and *separating* in the second.

6.2.3. Pasting along the boundary. Pasting is of course the inverse of cutting. Let M be a (possibly disconnected) manifold, let N_1, N_2 be two boundary components of M, and $\varphi\colon N_1 \to N_2$ be a diffeomorphism. We now define a new manifold M' obtained by *pasting M along φ*.

A naïve construction would be to define M' as $M/_\sim$ where $\sim$ is the equivalence relation that identifies $p \sim \varphi(p)$ for all $p \in N_1$. The result is indeed a topological manifold, but it is not obvious to assign a smooth atlas to $M/_\sim$. So we abandon this route, and we define M' instead by overlapping open collars as suggested by Figure 6.5.

Here are the details. We identify two disjoint closed collars of N_1 and N_2 in M with $N_1 \times [0,1]$ and $N_2 \times [0,1]$, where $N_i = N_i \times \{0\}$. The manifold M' is obtained from M by first removing N_1 and N_2, and then identifying the open subsets $N_1 \times (0,1)$ and $N_2 \times (0,1)$ via the gluing map $\Phi\colon (p,t) \mapsto \big(\varphi(p), 1-t\big)$. The smooth structure on M' is now easily induced by that of M: it suffices to take as an atlas all the charts $\varphi \circ (\pi|_U)^{-1}\colon U' \to V$ where $\pi\colon M \setminus (N_1 \cup N_2) \to M'$ is the projection and $\varphi\colon U \to V$ is a chart of $M \setminus (N_1 \cup N_2)$ such that $\pi|_U\colon U \to U'$ is a diffeomorphism.

Proposition 6.2.6. *The manifold M' depends up to diffeomorphism only on M and on the isotopy class of φ.*

Proof. Different closed collars are ambiently isotopic and hence produce diffeomorphic manifolds M'. Let us identify the glued part of M' with the open product $N_1 \times (0,1)$. Let F_t be an isotopy between $\varphi_0 = F_0$ and $\varphi_1 = F_1$, that we rescale as being constant at $t \in [0, \varepsilon]$ and $t \in [1-\varepsilon, 1]$. A diffeomorphism between the resulting manifolds M'_0 and M'_1 is constructed as follows: it is the identity outside the open product, and $(p,t) \mapsto \big(F_t^{-1}(\varphi_0(p)), t\big)$ on it. $\square$

The manifold M' contains a copy N of $N_1 \cong N_2$, embedded as $N = N_1 \times \{1/2\}$ in the glued product $N_1 \times (0,1) \subset M'$. If we cut M' along N we recover back our original M.

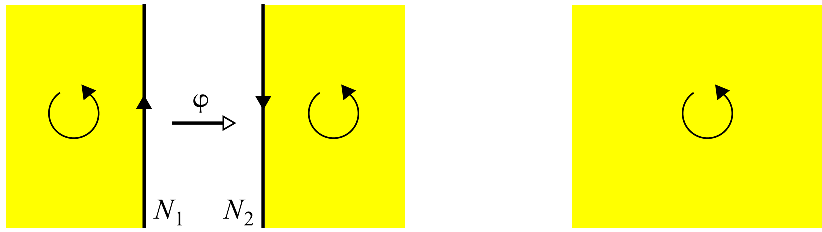

Figure 6.6. If the gluing map φ is orientation-*reversing*, the orientations extend to the new manifold M'.

Remark 6.2.7. Suppose that M is oriented. Both N_1 and N_2 inherit an orientation. If φ is orientation-*reversing*, the gluing map Φ is orientation-*preserving* and hence the orientation of M induces naturally an orientation on M'. So, if you want orientations to extend, you need to glue along orientation-reversing maps φ. See Figure 6.6.

Exercise 6.2.8. The smooth manifold M' is homeomorphic to the topological manifold $M/_\sim$ obtained from M by identifying $p \sim \varphi(p)$ for every $p \in N_1$.

In light of this fact, we will often think of M' simply as the topological space $M/_\sim$, equipped with an appropriate smooth atlas induced by φ.

6.2.4. Self-diffeomorphisms. Proposition 6.2.6 suggests that it is important to understand the self-diffeomorphisms of a manifold up to isotopy. We now solve this (generally difficult) problem for S^1.

Two self-diffeomorphisms of an orientable manifold M are *cooriented* if they either both preserve or both invert the orientation.

Exercise 6.2.9. If two self-diffeomorphisms of an orientable manifold M are isotopic, they are cooriented.

The converse is also sometimes true.

Proposition 6.2.10. *Two cooriented diffeomorphisms of S^1 are isotopic.*

Proof. Let $\varphi_0, \varphi_1 \colon S^1 \to S^1$ be two cooriented diffeomorphisms. They lift to smooth maps $\tilde\varphi_0, \tilde\varphi_1 \colon \mathbb{R} \to \mathbb{R}$ between their universal covers, that are monotone (that is, $\tilde\varphi_0'(t), \tilde\varphi_1'(t) > 0$ (or < 0) $\forall t$) and periodic (that is, $\varphi_i(t + 2\pi) = \varphi_i(t) + 2\pi\ \forall t$). The convex combination

$$\tilde\varphi_t(x) = (1 - t)\tilde\varphi_0(x) + t\tilde\varphi_1(x)$$

is also periodic and monotone, hence it descends to a monotone map $\varphi_t \colon S^1 \to S^1$. Each φ_t is hence a covering, but since it is homotopic to φ_0 it is a diffeomorphism: we get an isotopy between φ_0 and φ_1. □

This fact has important consequences when we want to glue two surfaces along their boundaries. Let Σ be a (possibly disconnected) surface and let

C_1, C_2 be two compact boundary components of Σ. Each C_i is diffeomorphic to S^1 by Exercise 6.1.14. We want to glue them along a diffeomorphism $\varphi\colon C_1 \to C_2$ as in Figure 6.5. The proposition tells us that there are only two possible gluing maps φ up to isotopy.

6.2.5. Doubles. Here is a simple kind of pasting that applies to every manifold with boundary.

The *double* DM of a manifold M with boundary is obtained by taking two identical copies M_1, M_2 of M and defining $\varphi\colon \partial M_1 \to \partial M_2$ as the *identity map*, that is the one that sends every point in ∂M_1 to its corresponding point in ∂M_2. Then DM is obtained by pasting $M_1 \sqcup M_2$ along φ.

The doubled manifold DM has no boundary. If M is compact, then DM also is.

Exercise 6.2.11. The double of D^n is diffeomorphic to S^n. The double of a cylinder $S^1 \times [0,1]$ is diffeomorphic to a torus $S^1 \times S^1$. What is the double of a Möbius strip?

6.2.6. More theorems extended to manifolds with boundary. We note that the double DM contains a copy of M as a closed domain, so in particular we have proved that every manifold with boundary is contained in some manifold without boundary as a closed domain. This fact is often useful to quickly extends theorems from manifolds without boundary to manifolds with boundary. This applies for instance to Whitney's embedding theorem.

Theorem 6.2.12. *For every smooth m-manifold M with boundary there is a proper embedding $M \hookrightarrow \mathbb{R}^{2m+1}$.*

Proof. The Whitney embedding Theorem 3.11.3 furnishes a proper embedding $DM \hookrightarrow \mathbb{R}^{2m+1}$. The inclusion $M \hookrightarrow DM$ is also a proper embedding, so the composition $M \hookrightarrow \mathbb{R}^{2m+1}$ also is. □

It also applies to smoothenings of continuous maps.

Theorem 6.2.13. *Every continuous map $f\colon M \to N$ between manifolds with boundary is homotopic to a smooth one.*

Proof. Using a collar for N we can easily push f inside the interior of N, that is we can homotope f to a map whose image lies in $\mathrm{int}(N)$. The resulting map doubles to a continuous map $DM \to \mathrm{int}(N)$, which is in turn homotopic to a smooth map g by Theorem 5.6.8. We conclude by restricting g to $M \subset DM$. □

Theorem 6.2.14. *Two smooth maps $f, g\colon M \to N$ between manifolds with boundary are continuously homotopic $\Longleftrightarrow$ they are smoothly homotopic.*

Proof. Let f and g be continuously homotopic. Using collars for M and N we can smoothly homotope f and g so that their images lie in $\mathrm{int}(N)$ and each f and g is constant along the fibres of the collar of M. By doubling, these maps extend to smooth maps $DM \to \mathrm{int}(N)$, that are still homotopic. By Corollary 5.6.9 there is a smooth homotopy relating them, that restricts to a smooth homotopy between f and g. □

6.2.7. Exotic spheres. We now investigate the following apparently innocuous construction: we pick a self-diffeomorphism $\varphi\colon S^{n-1} \to S^{n-1}$ and we glue two copies of D^n along φ, thus getting a new manifold M without boundary. What kind of smooth manifold M do we get?

Exercise 6.2.11 says that if $\varphi = \mathrm{id}$ then M is diffeomorphic to S^n. More generally, in the topological category, the answer does not depend on φ.

Proposition 6.2.15. *The manifold M is homeomorphic to S^n. If φ extends to a self-diffeomorphism of D^n, then M is also diffeomorphic to S^n.*

Proof. By Exercise 6.2.8 the manifold M is homeomorphic to the topological manifold $D_1 \cup_\varphi D_2$ obtained by identifying p with $\varphi(p)$. We define a continuous map

$$F\colon D_1 \cup_{\mathrm{id}} D_2 \longrightarrow D_1 \cup_\varphi D_2$$

by coning φ, that is: if $x \in D_1$ then $F(x) = x$, while if $x \in D_2$ we set

$$F(x) = \begin{cases} \|x\|\varphi\big(\frac{x}{\|x\|}\big) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

The map F is a homeomorphism. By Exercise 6.2.11 we have $D_1 \cup_{\mathrm{id}} D_2 \cong S^n$, and this completes the proof that M is homeomorphic to S^n.

If φ extends to a diffeomorphism $\Phi\colon D^n \to D^n$, we can replace $F|_{D_2}$ with Φ and get a diffeomorphism. More precisely, to get a smooth map we need to smoothen it at the equator ∂D^n like when we compose two smooth isotopies (details are left as an exercise). □

Corollary 6.2.16. *If $n = 2$ then M is diffeomorphic to S^2.*

Proof. Up to isotopy, the gluing map $\varphi\colon S^1 \to S^1$ is either the identity or a reflection $z \mapsto \bar{z}$, and they both extend to self-diffeomorphisms of D^2. □

The striking fact here is that when $n \geq 7$ the smooth manifold M may *not* be diffeomorphic to S^n, despite being homeomorphic to it. This implies in particular that there are some crazy self-diffeomorphisms of S^n that are not isotopic neither to the identity nor to a reflection, and moreover they do not extend to self-diffeomorphisms of D^n.

Remark 6.2.17. A smooth manifold homeomorphic but not diffeomorphic to S^n is called an *exotic sphere*. In dimension $n \geq 7$ there are many exotic spheres, and they are all constructed in this way. On the other hand, there are no exotic spheres in dimensions $n = 1, 2, 3, 5, 6$. The dimension 4 remains a total mystery: we do not know if there are exotic spheres, and if there are, they are certainly *not* constructed in this way (that is, by gluing two discs). Even more puzzling, we know that the number of exotic spheres (considered up to diffeomorphism) is finite in every dimension – for instance these are 27 in dimension 7 – except in dimension four, where the number of exotic spheres could be any value from 0 to ∞, both extremes included, as far as we know.

6.3. Connected sums and surgery

We have learned how to cut a manifold along a submanifold, and how to glue two boundary components along a diffeomorphism. We now introduce some more elaborate manipulations that combine cutting and pasting. The most important ones are the *connected sum* that "connects" two manifolds along a tube, and the more general *surgery* that roughly replaces a k-sphere (with trivial normal bundle) with a $(n-k-1)$-sphere. The boundary versions of these manipulations are also important.

6.3.1. Definition. Let M_1 and M_2 be two connected oriented n-manifolds, possibly with boundary. We now define a new oriented manifold $M_1 \# M_2$ called the *connected sum* of M_1 and M_2.

To do so, we consider the orientation-reversing diffeomorphism

$$\alpha\colon \operatorname{int}(D^n)\setminus\{0\} \longrightarrow \operatorname{int}(D^n)\setminus\{0\}, \quad \alpha(v) = \big(1-\|v\|\big)\frac{v}{\|v\|}$$

and two arbitrary embeddings

$$f_1\colon D^n \hookrightarrow \operatorname{int}(M_1), \qquad f_2\colon D^n \hookrightarrow \operatorname{int}(M_2)$$

such that f_1 is orientation-preserving and f_2 is orientation-reversing. Then we glue the punctured manifolds $M_1 \setminus f_1(0)$ and $M_2 \setminus f_2(0)$ via the diffeomorphism

$$f_2\circ\alpha\circ f_1^{-1}\colon f_1\big(\operatorname{int}(D^n)\setminus\{0\}\big) \longrightarrow f_2\big(\operatorname{int}(D^n)\setminus\{0\}\big).$$

The resulting smooth manifold is the *connected sum* of M_1 and M_2 and is denoted as

$$M_1 \# M_2.$$

Since $f_2\circ\alpha\circ f_1^{-1}$ is orientation-preserving, the manifold $M_1\#M_2$ is naturally oriented. You may visualise an example in Figure 6.7. By the Disc Theorem 6.1.15 the manifold $M_1\#M_2$ does not depend, up to orientation-preserving diffeomorphisms, on the maps φ_1 and φ_2.

Remark 6.3.1. The connected sum $M_1\#M_2$ may also be described as a two-steps cut-and-paste operation, where:

(1) first, we remove $f_i\big(\operatorname{int}(D_i)\big)$ from M_i, thus creating a new boundary component $f_i(\partial D_i)$ for M_i, $\forall i=1,2$;
(2) then, we paste the two new boundary components via the diffeomorphism $f_2\circ f_1^{-1}\colon \partial D_1 \to \partial D_2$.

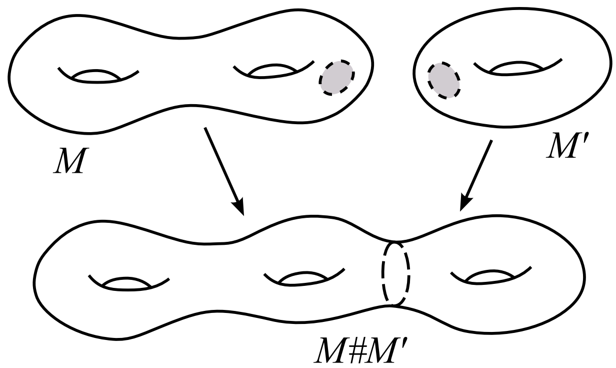

Figure 6.7. The connected sum of two compact surfaces.

We leave as an exercise to prove that this definition of $M_1 \# M_2$ is equivalent to the one given above. In light of the exotic spheres construction, it is important to require the gluing map to be $f_2 \circ f_1^{-1}$ and not any map.

We may see $\#$ as a binary operation on the set[3] of all oriented connected n-manifolds considered up to diffeomorphism.

Proposition 6.3.2. *The connected sum is commutative and associative, and S^n is the neutral element. That is, there are diffeomorphisms*

$$M\#N \cong N\#M, \qquad M\#(N\#P) \cong (M\#N)\#P, \qquad M\#S^n \cong M.$$

Proof. Commutativity is obvious. Associativity holds because we can separate the discs using isotopies, so that both connected sums can be performed simultaneously.

To construct $M\#S^n$ we follow Remark 6.3.1. We choose $\varphi_2 \colon D^n \hookrightarrow S^n$ to be the standard parametrisation of the upper hemisphere. The two-steps operation consists of substituting the upper hemisphere with the lower one along the same map, and this does not change the manifold M. □

The connected sum may be defined also for non-oriented manifolds, but in this case the resulting manifold $M\#N$ is not unique: there are two possibilities, and these may produce non-diffeomorphic manifolds in some cases. We have used orientations here only to simplify the theory.

6.3.2. Compact surfaces. Enough for the theory, we need some examples. One-dimensional manifolds are not very exciting, so we turn to surfaces. We already know some compact connected surfaces, possibly with boundary:

$$S^2, \quad \mathbb{RP}^2, \quad D^2, \quad S^1 \times [0,1], \quad S^1 \times S^1, \quad M, \quad K$$

[3]The suspicious reader may object that smooth manifolds do not form a set. However, if we consider them up to diffeomorphism, we may use Whitney's embedding theorem and see them as subsets of some $\mathbb{R}^n$, and the subsets of $\mathbb{R}^n$ of course form a set.

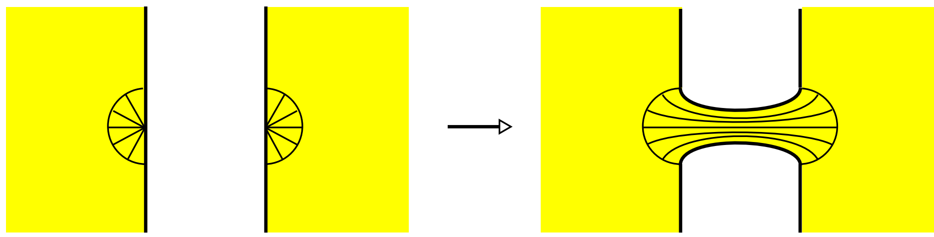

Figure 6.8. The ∂-connected sum of two manifolds with boundary.

where M and K are the compact Möbius strip, considered with its (connected!) boundary, and the Klein bottle. Can we add more surfaces to this list?

Definition 6.3.3. The *genus-g surface* S_g is the connected sum

$$S_g = \underbrace{T\# \ldots \# T}_{g}$$

of g copies of the torus $T = S^1 \times S^1$.

By convention, the surface of genus zero S_0 is the sphere S^2, and that of genus one S_1 is the torus. We have

$$S_g \# S_h \cong S_{g+h}.$$

Figure 6.7 shows that $S_2 \# S_1 \cong S_3$. Note that the torus T is mirrorable, so each time we make a connected sum with T it is not really important which orientation we put on T. In fact each S_g is easily seen to be mirrorable.

6.3.3. ∂-connected sum. A ∂-connected sum is an operation similar to the connected sum, where a bridge is added to connect two portions of the boundaries as in Figure 6.8.

The construction goes as follows. We consider the *half-disc* $D^n_+ = D^n \cap \mathbb{R}^n_+$. We define $D^{n-1} = D^n_+ \cap \{x_n = 0\}$ and $\mathrm{int}(D^n_+) = D^n_+ \cap \{\|x\| < 1\}$. We consider the same orientation-reversing diffeomorphism as above

$$\alpha\colon \mathrm{int}(D^n_+) \setminus \{0\} \longrightarrow \mathrm{int}(D^n_+) \setminus \{0\}, \quad \alpha(v) = \big(1 - \|v\|\big)\frac{v}{\|v\|}.$$

Let M_1 and M_2 be two oriented n-manifolds with boundary. Pick two embedded half-discs

$$f_1\colon D^n_+ \hookrightarrow M_1, \qquad f_2\colon D^n_+ \hookrightarrow M_2$$

such that $f_i^{-1}(\partial M_i) = D^{n-1}$ as in Figure 6.8-(left). We require f_1 to be orientation-preserving and f_2 orientation-reversing. Then we glue the manifolds $M_1 \setminus f_1(0)$ and $M_2 \setminus f_2(0)$ via the diffeomorphism

$$f_2 \circ \alpha \circ f_1^{-1}\colon f_1\big(\mathrm{int}(D^n_+) \setminus \{0\}\big) \longrightarrow f_2\big(\mathrm{int}(D^n_+) \setminus \{0\}\big).$$

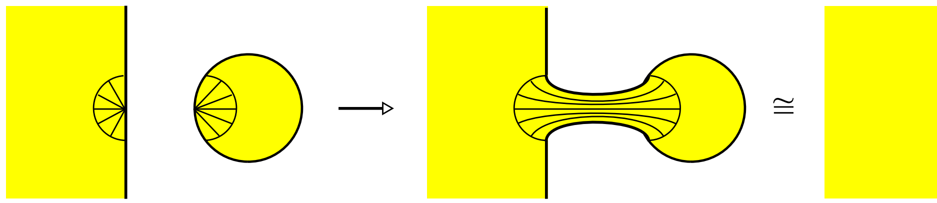

Figure 6.9. The ∂-connected sum with a disc does not change the manifold up to diffeomorphism.

The resulting oriented smooth manifold with boundary is the *∂-connected sum* of M_1 and M_2 and is denoted as

$$M_1 \#_\partial M_2.$$

See Figure 6.8. As above one proves that the resulting manifold depends only on the connected components of ∂M_1 and ∂M_2 intersecting the half-discs. In particular, if both M_1 and M_2 have connected boundary, then $M_1 \#_\partial M_2$ is uniquely determined.

Proposition 6.3.4. *If ∂M_1 and ∂M_2 are connected, we have*

$$\partial(M_1 \#_\partial M_2) \cong \partial M_1 \# \partial M_2.$$

In general we have $M \#_\partial D^n \cong M$.

Proof. The manipulation restricted to the boundaries is a connected sum, so the first isomorphism holds. The second is sketched in Figure 6.9, and we leave the tedious exercise of writing the correct diffeomorphism to the courageous reader. □

6.3.4. Pasting manifolds along submanifolds. We now introduce a generalisation of the connected sum, in which we glue manifolds along disc bundles instead of just discs.

Pick $0 \leq k < n$. Let M_1 and M_2 be two n-manifolds possibly with boundary, and let $N_1 \subset \operatorname{int}(M_1)$ and $N_2 \subset \operatorname{int}(M_2)$ be two diffeomorphic compact k-submanifolds without boundary. We suppose that there is a vector bundle isomorphism $\varphi\colon \nu N_1 \to \nu N_2$ between the two normal bundles, both realised as tubular neighborhoods $\nu N_1 \subset \operatorname{int}(M_1)$ and $\nu N_2 \subset \operatorname{int}(M_2)$. We fix a Riemannian metric on one normal bundle νN_1 and transport it to the other νN_2 along φ, to get an induced isomorphism of the corresponding closed tubular neighbourhoods $\varphi\colon \bar\nu N_1 \to \bar\nu N_2$.

As above, we define the self-diffeomorphism

$$\alpha\colon \operatorname{int}(\bar\nu N_1) \setminus N_1 \longrightarrow \operatorname{int}(\bar\nu N_1) \setminus N_1, \qquad \alpha(v) = \big(1 - \|v\|\big)\frac{v}{\|v\|}.$$

We now glue the manifolds $M_1 \setminus N_1$ and $M_2 \setminus N_2$ via the diffeomorphism

$$\varphi \circ \alpha \colon \mathrm{int}(\bar{\nu} N_1) \setminus N_1 \longrightarrow \mathrm{int}(\bar{\nu} N_2) \setminus N_2.$$

The resulting manifold M is obtained by *pasting* M_1 and M_2 along the submanifolds N_1 and N_2. It is an operation that can be done as soon as the submanifolds N_1 and N_2 have isomorphic normal bundles; note however that, as opposite to connected sum, the choice of the vector bundle isomorphism φ is important here, because two different isomorphisms may not be isotopic in many interesting cases, even if they are co-oriented.

Remark 6.3.5. As in Remark 6.3.1, the construction of M may be described alternatively as a two-steps cut-and-paste operation, where:

(1) first, we remove from M_i the open submanifold $\mathrm{int}(\bar{\nu} N_i)$, thus creating a new boundary component $\partial \bar{\nu} N_i$;
(2) then, we paste the two new boundary components via φ.

6.3.5. Surgery. There is a particular type of pasting that is so important to deserve a separate name.

Let M be a n-manifold, possibly with boundary, and $\Sigma \subset \mathrm{int}(M)$ be a k-sphere (that is, a submanifold diffeomorphic to S^k) with trivial normal bundle, for some $0 \leq k \leq n-1$.

As in Section 6.1.16, we set $\mathbb{R}^{k+1} \times \mathbb{R}^{n-k} = \mathbb{R}^{n+1}$ and consider $S^k \times \{0\} \subset S^n$. We have seen that the normal bundle of $S^k \times \{0\}$ in S^n is also trivial. We can therefore paste M and S^n along the k-spheres Σ and $S^k \times \{0\}$. To do so, we must choose a normal bundle isomorphism $\varphi \colon \nu\Sigma \to \nu S^k$. This operation is called a *surgery* along the sphere $\Sigma \subset \mathrm{int}(M)$. The resulting manifold M' depends on the chosen isomorphism φ.

Remark 6.3.6. We have seen in Section 6.1.16 that S^n decomposes into $S^k \times D^{n-k}$ and $D^{k+1} \times S^{n-k-1}$. Therefore, by Remark 6.3.5, a surgery may also be described as follows: whenever we find a domain in M diffeomorphic to $S^k \times D^{n-k}$, we first remove its interior, thus creating a new boundary $S^k \times S^{n-k-1}$, and then glue $D^{k+1} \times S^{n-k-1}$ to it via the identity map. Shortly: we substitute $S^k \times D^{n-k}$ with $D^{k+1} \times S^{n-k-1}$.

Remark 6.3.7. A surgery along a 0-sphere is like a connected sum: we replace $S^0 \times D^n$, that is two disjoint discs, with $D^1 \times S^{n-1}$, that is a tube. When both points in S^0 are contained in the same connected component, this may be interpreted as a *self*-connected sum of that component.

The inverse operation of a surgery along a k-sphere is naturally a surgery along a $(n-k-1)$-sphere.

Example 6.3.8. Le M be an orientable 3-manifold, possibly with boundary. A *knot* in M is a submanifold $K \subset \mathrm{int}(M)$. Since M is orientable, we will prove that the normal bundle νK of K in M is trivial. The closed tubular TBD alla fine neighbourhood $\bar{\nu}K \subset \text{int}(M)$ is therefore diffeomorphic to a solid torus $S^1 \times D^2$. A surgery along K consists of replacing the solid torus $S^1 \times D^2$ with another solid torus $D^2 \times S^1$. This operation typically modifies drastically the topology of the ambient manifold.

6.3.6. Pasting along submanifolds in the boundary. There is of course a boundary version of pasting along submanifolds, where the submanifolds lie in the boundary. This operation generalises the ∂-connected sum and will be fundamental in the next section.

Let M_1 and M_2 be two n-manifolds with boundary, and let $N_1 \subset \partial M_1$ and $N_2 \subset \partial M_2$ be two compact k-submanifolds without boundary. Let $\varphi\colon \nu N_1 \to \nu N_2$ be an isomorphism of their normal bundles in ∂M_1 and ∂M_2.

We now define a new manifold M' obtained by *pasting* M_1 and M_2 along the submanifolds N_1 and N_2. The operation is the same as above, only with half-discs instead of disc bundles.

Each closed tubular neighbourhood $\bar{\nu}N_i \subset M_i$ is a D^{n-k-1}-bundle over N_i, and using collars we may extend it to a half-disc D^{n-k}_+-bundle $\bar{\nu}_+ N_i$ that is a "half"-tubular neighbourhood of N_i in M_i. The diffeomorphism φ also extends to $\varphi\colon \bar{\nu}_+N_1 \to \bar{\nu}_+N_2$. We glue the manifolds $M_1 \setminus N_1$ and $M_2 \setminus N_2$ via the diffeomorphism

$$\varphi \circ \alpha\colon \text{int}(\bar{\nu}_+N_1) \setminus N_1 \longrightarrow \text{int}(\bar{\nu}_+N_2) \setminus N_2$$

where α and $\text{int}(\bar{\nu}N_i)$ are defined on every fibre D^{n-k}_+ as we did for ∂-connected sums.

The ∂-connected sum corresponds to the case where N_1 and N_2 are points.

6.4. Handle decompositions

We now show that every compact manifold M decomposes into finitely many simple blocks, called *handles*. This important procedure is called a *handle decomposition*.

6.4.1. Handles. We have described in the previous section the operation of pasting two manifolds along submanifolds in their boundaries. We now introduce a particularly important case.

Let M be a n-manifold with boundary. Let $\Sigma \subset \partial M$ be a $(k-1)$-sphere with trivial normal bundle $\nu\Sigma \subset \partial M$, with $0 < k \leq n$. A *k-handle addition* on M is the operation that consists of pasting M with D^n along the $(k-1)$-spheres Σ and $S^{k-1} \times \{0\} \subset S^{n-1}$. As in Section 6.1.16, we see D^n inside $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$.

The result is a new smooth manifold M', which depends on the (isotopy class of the) chosen identification of the trivial normal bundles of Σ and of $S^{k-1} \times \{0\}$.

A n-handle addition is the glueing of a disc D^n to a boundary component $\Sigma \subset \partial M$ diffeomorphic to S^{n-1} via a diffeomorphism $\partial D^n \to \Sigma$. We also

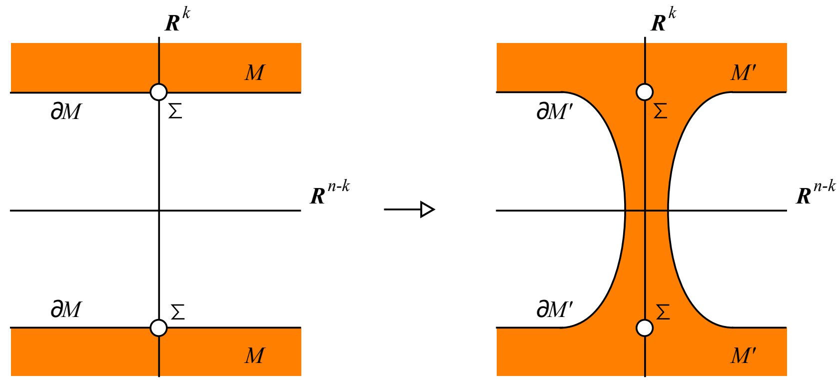

Figure 6.10. An alternative description of the attachment of a k-handle to M.

define a 0-*handle addition* to be simply the addition of a disjoint connected component D^n, with no attachment.

6.4.2. Local model. To better visualise what is going on, we furnish a concrete local model of a k-handle addition, drawn in Figure 6.10.

Let $\Sigma \subset \partial M$ be a $(k-1)$-sphere with trivial normal boundary. It has a half-tubular neighbourhood in M is diffeomorphic to $\Sigma \times \mathbb{R}^{n-k} \times \mathbb{R}_+$ and we identify it with the manifold with boundary

$$U = \left\{(x, y) \in \mathbb{R}^{n-k} \times \mathbb{R}^k \;\middle|\; \|y\| \geq 1\right\}$$

via the map $(u, v, t) \mapsto (v, (t+1)u)$. With this identification we have

$$\Sigma = \{0\} \times S^{k-1}, \qquad \partial U = U \cap \partial M = \mathbb{R}^{n-k} \times S^{k-1}.$$

Let $\rho\colon [-1, 1] \to \mathbb{R}_+$ be a continuous positive function that is smooth on $(-1, 1)$ and such that all derivatives of ρ tend to $\pm\infty$ as $t \to \pm 1$ (corresponding signs). We define a bigger manifold M' by substituting U with the bigger set

$$U' = U \cup \left\{\|y\| < 1, \|x\| < \rho(\|y\|)\right\}.$$

Exercise 6.4.1. The manifold M' is diffeomorphic to M with a k-handle attached to Σ.

See Figure 6.10. Note that with this description the original manifold M is naturally a submanifold of M'.

6.4.3. Topological handles. We can make one further step towards visualization and intuition by using *topological handles*. These capture the topological structure of M' while being a little bit imprecise on its smooth structure. See Figure 6.11.

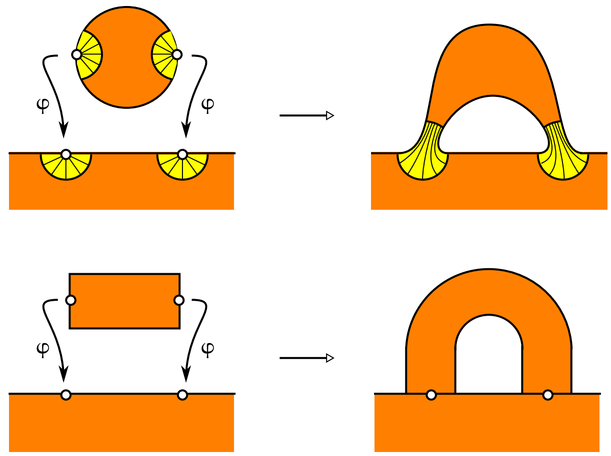

Figure 6.11. The attachment of a 1-handle and of a topological 1-handle along the same map φ. The resulting topological manifold is the same in both constructions, but the smooth structure is well-defined only with the first. For practical purposes, we usually think of a handle as a topological handle whose corners have been somehow "smothened."

A topological handle is what we get if we take $\rho(t) = 1$ constantly in the previous construction. The result is not smooth, but it still works up to homeomorphisms.

In other words, we use $D^k \times D^{n-k}$ instead of D^n. This is not a smooth manifold because of its corners; its topological boundary decomposes into the horizontal $D^k \times S^{n-k-1}$ and the vertical $S^{k-1} \times D^{n-k}$. For every embedding

$$\varphi\colon S^{k-1} \times D^{n-k} \hookrightarrow \partial M$$

we define a new topological space

$$M' = M \cup_\varphi \left(D^k \times D^{n-k}\right)$$

obtained by attaching $D^k \times D^{n-k}$ to M along φ. This operation is the attachment of a *topological k-handle* to M. The attaching of a handle or a topological handle along the same map φ produce homeomorphic manifolds M': the only difference between the two constructions is that in the topological setting the smooth structure on M' is not obvious to see – some new corners arise that should be smoothened, see Figure 6.11. From now on, we will always think as a handle as a topological handle whose corners have been smoothened.

One should think of a topological k-handle $D^k \times D^{n-k}$ as a thickened k-dimensional disc. Here is some useful terminology: the number k is the *index* of the handle; the sphere $S^{k-1} \times \{0\}$ is the *attaching sphere*, while the sphere $\{0\} \times S^{n-k-1}$ is the *belt sphere*. The discs $D^k \times \{0\}$ and $\{0\} \times D^{n-k}$ are the *attaching* and *belt discs*. See some examples in Figure 6.12.

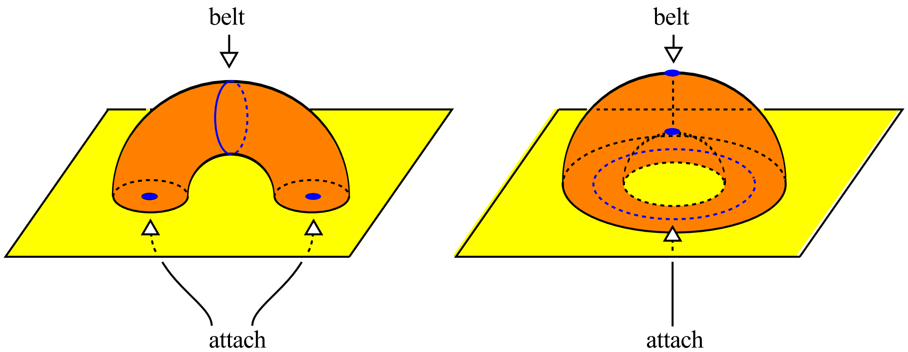

Figure 6.12. A three-dimensional topological 1-handle (left) and 2-handle (right), with the attaching and belt spheres in blue.

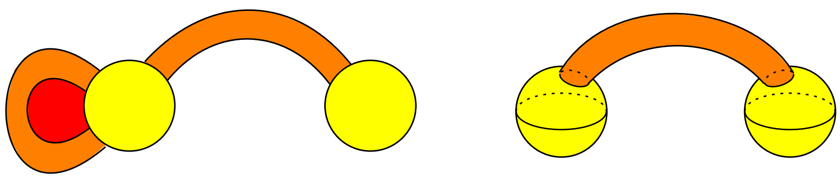

Figure 6.13. Some handle decompositions in dimension two and three. On the left, we have two 0-handles (yellow), one 1-handle (orange), and one 2-handle (red) in dimension two. On the right, we have two 0-handles (yellow) and one 1-handle (orange) in dimension three.

Remark 6.4.2. If M' is obtained from M by the attachment of a k-handle to the $(k-1)$-sphere $S \subset M$, the new boundary $\partial M'$ is obtained from the old ∂M by surgery along the sphere S. This follows readily from the definition.

6.4.4. Handle decomposition. Let M be a compact smooth n-manifold, possibly with boundary. A *handle decomposition* for M is the realisation of M as the result of a finite number of operations

$$\varnothing = M_0 \rightsquigarrow M_1 \rightsquigarrow \cdots \rightsquigarrow M_k = M$$

where each M_{i+1} is obtained by attaching some handle to M_i. Since the only handle that can be attached to the empty set is a 0-handle, the manifold M_1 is the result of a 0-handle attachment to $\varnothing$ and is hence a n-disc.

Example 6.4.3. The sphere S^n, and more generally each of the exotic spheres described in Section 6.2.7, decomposes into two n-discs. We may interpret this decomposition as a n-handle attached to a 0-handle. Therefore S^n has a handle decomposition with one 0-handle and one n-handle.

Conversely, if a compact manifold M without boundary decomposes into two handles only, then these must be a 0- and a n-handle, and so M is either S^n or an exotic sphere (in all cases, it is homeomorphic to S^n).

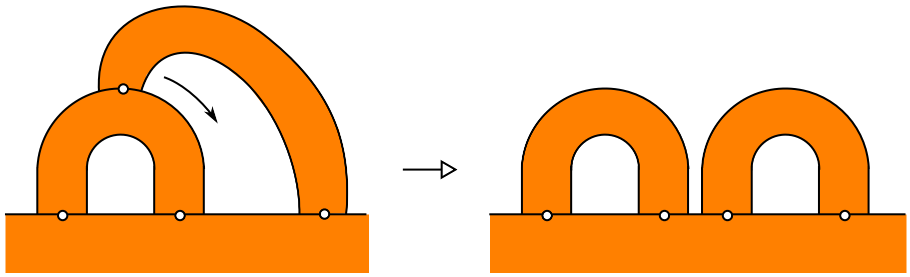

Figure 6.14. If $h \leq k$, we can always slide a k-handle away from a previously attached h-handle. Here $h = k = 1$.

6.4.5. Reordering handles. More examples are shown in Figure 6.13. In both examples in the figure the handle decomposition goes as follows: we first attach some 0-handles (that is, we create discs out of nothing), then we attach some 1-handles, then we attach some 2-handles. We think at the 1-handles in the (left) figure as attached simultaneously. We now show that every handle decomposition can be modified to be of this type.

Proposition 6.4.4. *Every handle decomposition can be modified so that we first attach all 0-handles, then all 1-handles, then all 2-handles ... and so on.*

Proof. Suppose that M_{i+1} is obtained from M_i by attaching a k-handle H^k, and M_{i+2} is obtained from M_{i+1} by attaching a h-handle H^h. We write

$$M_{i+1} = M_i \cup_\varphi H^k, \qquad M_{i+2} = M_{i+1} \cup_\psi H^h.$$

We show below that if $h \leq k$ then H^h can be slid away from H^k as in Figure 6.14. After this move, the handles H^h and H^k are disjoint and hence we can obtain the same manifold M_{i+2} by first attaching H^h and then H^k.

By applying finitely many exchanges of this type we transform every handle decomposition into one where handles are attached with non-decreasing index. Moreover, the handles with the same index can be slid to be disjoint, and hence can be thought to be attached simultaneously. This proves the proposition.

We now show how to slide H^h aways from H^k. The attaching sphere of H^h is a $(h-1)$-sphere $\Sigma \subset \partial M_{i+1}$, while the belt sphere of H^k is a $(n-k-1)$-sphere $\Sigma' \subset \partial M_{i+1}$. If $h \leq k$, we have $(h-1)+(n-k-1) < n-1$. By transversality, we may isotope Σ away from Σ'.

The handles H^k and H^h intersect ∂M into two closed tubular neighbourhoods of Σ' and Σ. Since $\Sigma' \cap \Sigma = \varnothing$, we can isotope the tubular neighbourhood of Σ to be disjoint from that of Σ'. That is, we can slide the handle H^h away from H^k, as stated. □

As stressed in the proof, the handles of the same index are disjoint and can be attached simultaneously, as in Figure 6.13.

Our next goal is to show that every compact smooth manifold decomposes into handles. To this purpose we study the critical points of functions $M \to \mathbb{R}$ and we introduce the *Morse functions*, that are of independent interest.

6.4.6. Hessian at a critical point. Let M be a manifold without boundary and $f : M \to \mathbb{R}$ a smooth function. Let $p \in M$ be a critical point of f. The *Hessian* of f at p is a symmetric bilinear form

$$\mathrm{Hess}\,(f)_p : T_pM \times T_pM \longrightarrow \mathbb{R}$$

defined as follows. Given $v, w \in T_pM$, extend them to two arbitrary vector fields X, Y in some neighbourhood of p. Then we set

$$\mathrm{Hess}\,(f)_p(v, w) = X\big(Y(f)\big)(p).$$

Lemma 6.4.5. *The map* $\mathrm{Hess}\,(f)_p$ *is well-defined, bilinear, and symmetric.*

Proof. Since p is a critical point we have $u(f) = 0$ for every tangent vector $u \in T_pM$. This holds in particular for $u = [X, Y](p)$ and gives $v(Y(f)) = w(X(f))$. The left member does not depend on the extension X, and the right does not depend on the extension Y: therefore both do not depend on the extensions, and the bilinear form is manifestly symmetric. □

It is crucial here that $df_p = 0$. Alternatively, we can also define the Hessian in coordinates: we pick $p = 0$ for simplicity and get

$$f(x) = f(0) + \tfrac{1}{2}\,{}^{t}xHx + o\big(\|x\|^2\big).$$

On some other chart with variables $\bar{x}$, we get $x = J\bar{x} + o\big(\|\bar{x}\|\big)$ where J is the differential of the coordinates change at $x = 0$ and therefore

$$\begin{aligned} f(x) &= f(0) + \tfrac{1}{2}\,{}^{t}\big(J\bar{x} + o\big(\|\bar{x}\|\big)\big)H\big(J\bar{x} + o\big(\|\bar{x}\|\big)\big) + o\big(\|x\|^2\big) \\ &= f(0) + \tfrac{1}{2}\,{}^{t}\bar{x}\,{}^{t}JHJ\bar{x} + o\big(\|\bar{x}\|^2\big). \end{aligned}$$

Therefore H changes to ${}^{t}JHJ$ and hence describes a chart-independent bilinear form on T_pM. The two definitions just given coincide because

$$\mathrm{Hess}\,(f)_0(e_i, e_j) = \frac{\partial^2 f}{\partial x_i \partial x_j} = H_{ij}.$$

6.4.7. Non-degenerate critical points. Let M be a manifold without boundary and $f : M \to \mathbb{R}$ a smooth function. We say that a critical point $p \in M$ for f is *non-degenerate* if the bilinear form $\mathrm{Hess}\,(f)_p$ on T_pM is non-degenerate. We now study the non-degenerate critical points. We start by exhibiting an alternative definition.

Proposition 6.4.6. *A critical point* p *is non-degenerate* $\Longleftrightarrow$ *the section* df *of* T^*M *is transverse to the zero-section at* p*.*

Proof. On a chart, we have $f\colon U \to \mathbb{R}$ for some open set $U \subset \mathbb{R}^n$. We see df as the gradient $\nabla f\colon U \to \mathbb{R}^n$. Now ∇f is transverse to the zero-section at $p \in U \iff$ the differential of ∇f is invertible in p. The differential of ∇f is $\text{Hess}\,(f)_p$, so we are done. $\square$

Corollary 6.4.7. *Non-degenerate critical points are isolated.*

If p is a non-degenerate critical point, then $\text{Hess}\,(f)_p$ is a scalar product on T_pM and has some signature $(k, n-k)$ for some $0 \leq k \leq n$. The integer $n-k$ is the *index* of the critical point p. The following Morse Lemma determines the behaviour of f near p, according to its index.

Lemma 6.4.8 (Morse Lemma). *Let p be a non-degenerate critical point of index $n-k$. On some appropriate chart near p the function f is read as*

$$f(x) = f(p) + x_1^2 + \ldots + x_k^2 - x_{k+1}^2 - \ldots - x_n^2.$$

The chart sends p to 0.

Proof. On a chart we get $f\colon \mathbb{R}^n \to \mathbb{R}$ with $p = 0$. Since 0 is a critical point, Taylor's Theorem 1.3.1 gives

$$f(x) = f(0) + \frac{1}{2}\sum_{i,j=1}^n h_{ij}(x)x_ix_j$$

for some smooth maps h_{ij} such that $h_{ij} = h_{ji}$. The Hessian $H = h_{ij}(0)$ has signature $(k, n-k)$.

To transform f into the desired form, we follow the usual procedure to diagonalise scalar products, and extend it smoothly on a neighbourhood of 0. We proceed by induction: suppose that on some coordinates we write

$$f(x) = \pm x_1^2 \pm \cdots \pm x_{r-1}^2 + \sum_{i,j\geq r} h_{ij}(x)x_ix_j.$$

After a linear change of coordinates we may suppose that $h_{rr}(x) \neq 0$ at $x = 0$ and hence on some small neighbourhood around 0. We pick new coordinates

$$\begin{cases} y_i = x_i \quad \text{for } i \neq r, \\ y_r = \sqrt{\|h_{rr}(x)\|}\left(x_r + \sum_{i>r}\frac{h_{ir}(x)x_i}{h_{rr}(x)}\right). \end{cases}$$

This is indeed a new coordinate systems around 0 by the Inverse Function Theorem. With these new coordinates we easily get

$$f(y) = \pm y_1^2 \pm \cdots \pm y_r^2 + \sum_{i,j>r} h'_{ij}(y)y_iy_j$$

for some functions h'_{ij} defined near p, and we conclude by induction on r. $\square$

6.4.8. Morse functions. Let M be a manifold without boundary. A *Morse function* on M is a function $f: M \to \mathbb{R}$ whose critical points are all non-degenerate. In other words, the differential df is transverse to the zero-section.

We now prove that there are plenty of Morse functions. Via the Whitney embedding theorem we may suppose that $M \subset \mathbb{R}^m$ for some m.

Lemma 6.4.9. *Let $M \subset \mathbb{R}^m$ be a submanifold and $f: M \to \mathbb{R}$ any smooth function. For almost every $v \in \mathbb{R}^m$, the modified function*

$$f_v: M \longrightarrow \mathbb{R}, \qquad f_v(x) = f(x) + \langle v, x \rangle$$

is a Morse function.

Proof. Consider the map

$$\begin{aligned} F: M \times \mathbb{R}^n &\longrightarrow T^*M \\ (p, v) &\longrightarrow d(f_v)_p. \end{aligned}$$

If we prove that F is transverse to the zero-section $s_0 \subset T^*M$, the Thom Transversality Theorem 5.7.5 implies that df_v is transverse to s_0 for almost every $v \in \mathbb{R}^m$ and we conclude.

To prove that F is transverse to s_0, we first note that

$$d(f_v)_p = df_p + \langle v, \cdot \rangle.$$

We deduce that $F(p, \cdot): \mathbb{R}^n \to T^*_pM$ is affine and surjective for every $p \in M$. This implies easily that F is transverse to any section s of T^*M. $\square$

Corollary 6.4.10. *Let $f: M \to \mathbb{R}$ a smooth function. For every $\varepsilon > 0$ there is a Morse function $g: M \to \mathbb{R}$ with $|f(p) - g(p)| < \varepsilon$ for all $p \in M$.*

Proof. Embed M in a ball of $\mathbb{R}^m$, and then apply Lemma 6.4.9 with sufficiently small $\|v\|$. $\square$

We have proved in particular that every M has some Morse function $f: M \to \mathbb{R}$. It is sometimes useful to add the following requirement.

Theorem 6.4.11. *Every manifold M without boundary has a proper Morse function $f: M \to \mathbb{R}$ where the critical values form a discrete closed subset of $\mathbb{R}$ and distinct critical points have distinct critical values.*

Proof. Start with a proper function $M \to \mathbb{R}$ and perturb it to a proper Morse function using Corollary 6.4.10. The critical points form a closed discrete subset of M, hence there are only finitely many of them in the compact subset $f^{-1}([a, b])$ for every $a < b$. Therefore the critical values form a discrete closed subset of $\mathbb{R}$. By choosing a generic small v in the proof of Corollary 6.4.10 we also get that distinct critical points have distinct values. $\square$

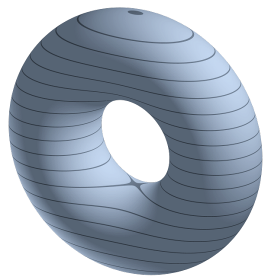

Figure 6.15. On this torus, the height function $f(x, y, z) = z$ is a Morse function with four non-degenerate critical points of index 0, 1, 1, and 2. The level sets $f^{-1}(t)$ are manifolds, except when t is a critical value.

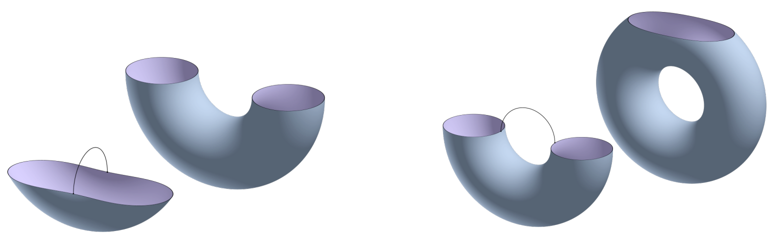

Figure 6.16. Each time a non-degenerate critical point of index k is crossed, a k-handle is added. We show here the two critical points of index 1, and the core segment of the 1-handle in each case.

6.4.9. Existence of handle decompositions. We have introduced Morse functions as a fundamental tool to prove the following remarkable theorem.

Theorem 6.4.12. *Every compact manifold M without boundary has a handle decomposition.*

Proof. Let $f\colon M \to \mathbb{R}$ be a Morse function where critical points have distinct images. Since M is compact, it has finitely many critical points. For instance, Figure 6.15 shows a Morse function on the torus with four critical points. For every $a \in \mathbb{R}$ we define

$$M_a = f^{-1}(-\infty, a].$$

When a is regular, M_a is a domain in M, that is a submanifold with boundary. Consider two regular values $a < b$. We now prove two facts:

(1) If $[a, b]$ contains no critical values, then M_a and M_b are diffeomorphic.
(2) If $[a, b]$ contains a single critical value, image of a critical point of index k, then M_b is diffeomorphic to M_a with a k-handle attached.

An example is shown in Figure 6.16. When a crosses a critical point of index k, a k-handle is attached to M_a. So the torus decomposes into one 0-handle, two 1-handles, and one 2-handle. The claims (1) and (2) clearly imply that M decomposes into handles, one for each critical point of M.

We first prove (1). Fix an arbitrary Riemannian metric on M, that is on the tangent bundle TM. Every T_pM is equipped with a scalar product $\langle , \rangle$, and we use it to transform the covector field df into a vector field ∇f in the usual way, by requiring that

$$df_p(v) = \big\langle \nabla f(p), v \big\rangle.$$

The field ∇f vanishes at the critical points. On a curve $\gamma\colon I \to M$ we get

$$(f \circ \gamma)'(t) = df_{\gamma(t)}\big(\gamma'(t)\big) = \big\langle \nabla f, \gamma'(t) \big\rangle.$$

Let $\rho\colon M \to \mathbb{R}$ be a smooth function that equals $1/\langle \nabla f, \nabla f \rangle$ on the compact set $f^{-1}[a, b]$ and which vanishes outside some bigger compact subset. We define a new vector field

$$X(p) = \rho(p)\nabla f(p).$$

Since M is compact, the vector field X is complete and generates a flow Φ. Consider a maximal integral curve $\gamma(t) = \Phi(p, t)$. If $\gamma(t) \in f^{-1}[a, b]$ then

$$(f \circ \gamma)'(t) = \big\langle \nabla f, \gamma'(t) \big\rangle = \big\langle \nabla f, X \big\rangle = 1.$$

Therefore the flow defines a diffeomorphism

$$M_a \longrightarrow M_b, \qquad p \longmapsto \Phi(p, b - a).$$

We turn to (2). Let $p \in M$ be the unique critical point in $f^{-1}[a, b]$. We suppose for simplicity that $f(p) = 0$. By (1) we may choose $a = -\varepsilon$ and $b = \varepsilon$ for some small $\varepsilon > 0$. By the Morse Lemma, on a chart $U \cong \mathbb{R}^n = \mathbb{R}^{n-k} \times \mathbb{R}^k$ the function f is

$$f(x, y) = \|x\|^2 - \|y\|^2$$

where $(x, y) \in \mathbb{R}^{n-k} \times \mathbb{R}^k$ and $p = (0, 0)$. The manifolds M_ε and $M_{-\varepsilon}$ intersect the chart $\mathbb{R}^{n-k} \times \mathbb{R}^k$ as in Figure 6.17-(left).

We now substitute M_ε with a diffeomorphic submanifold M' that still contains $M_{-\varepsilon}$, and which has the additional property that $M' \setminus M_{-\varepsilon}$ lies entirely in the chart $\mathbb{R}^{n-k} \times \mathbb{R}^k$ as shown in Figure 6.17-(right). To this purpose, we pick a smooth function $\phi\colon \mathbb{R} \to \mathbb{R}$ such that

$$\phi(0) > \varepsilon, \quad \phi(t) = 0 \ \forall t \geq 2\varepsilon, \quad -1 \leq \phi'(t) \leq 0 \ \forall t.$$

We now define another smooth function $F\colon M \to \mathbb{R}$, by requiring that $F(p) = f(p)$ outside the chart, and

$$F(x, y) = f(x, y) - \phi\big(2\|x\|^2 + \|y\|^2\big)$$

inside the chart. We then set

$$M' = F^{-1}(-\infty, -\varepsilon].$$

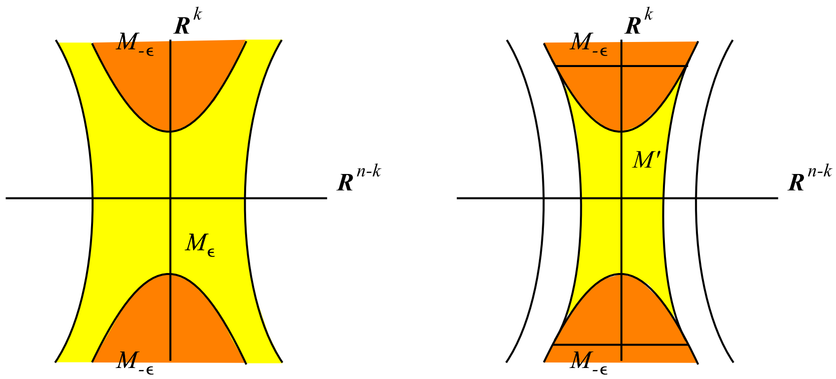

Figure 6.17. The manifolds M_ε and $M_{-\varepsilon}$ intersect the chart $\mathbb{R}^{n-k} \times \mathbb{R}^k$ as shown here (left). We replace M_ε with a diffeomorphic submanifold M', still containing $M_{-\varepsilon}$, so that the yellow zone $M' \setminus M_{-\varepsilon}$ lies entirely in this chart. The yellow zone is a k-handle (right).

Clearly $M' \supset M_{-\varepsilon}$ and $M' \setminus M_{-\varepsilon}$ is contained in the chart. We show that

$$M_\varepsilon = F^{-1}(-\infty, \varepsilon].$$

Indeed, we obviously have $M_\varepsilon \subset F^{-1}(-\infty, \varepsilon]$, and conversely if $F(x, y) \leq \varepsilon$ and $\phi(2\|x\|^2 + \|y\|^2) > 0$ we get $2\|x\|^2 + \|y\|^2 < 2\varepsilon$; therefore

$$f(x, y) = \|x\|^2 - \|y\|^2 \leq \|x\|^2 + \tfrac{1}{2}\|y\|^2 < \varepsilon.$$

In the chart we have

$$\frac{\partial F}{\partial x^i} = 2x^i - \phi'(2\|x\|^2 + \|y\|^2)x^i \geq 2x^i, \tag{9}$$

$$\frac{\partial F}{\partial y^i} = -2y^i - \phi'(2\|x\|^2 + \|y\|^2)y^i \leq -y^i. \tag{10}$$

This implies that dF vanishes only at the origin, so F has the same critical points as f. Since $F(p) < -\varepsilon$, the function F has no critical values in $[-\varepsilon, \varepsilon]$ and (1) implies that M' and M_ε are diffeomorphic.

Finally, we need to show that M' is diffeomorphic to $M_{-\varepsilon}$ with a k-handle attached, painted in yellow in Figure 6.17-(right). To this purpose we fix $y_0 \in \mathbb{R}^k$ and consider

$$F(x, y_0) = \|x\|^2 - \|y_0\|^2 - \phi(2\|x\|^2 + \|y_0\|^2).$$

The horizontal slice $y = y_0$ of M' has the form

$$M' \cap \{y = y_0\} = \{(x, y) \mid F(x, y_0) \leq -\varepsilon\}.$$

Since $\phi(t) > \varepsilon - t$ for all $t \geq 0$, we get

$$F(0, y_0) = -\|y_0\|^2 - \phi(\|y_0\|^2) < -\varepsilon$$

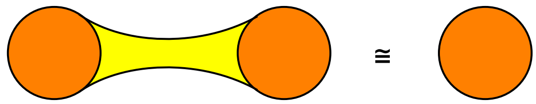

Figure 6.18. A 1-handle attached to two distinct 0-handles: the result is diffeomorphic to a disc.

and hence $(0, y_0)$ belongs to the horizontal slice. The function $F(x, y_0)$ depends only on $\|x\|$ and increases with $\|x\|$ by (9), so the horizontal slice is a disc with some radius $r(y_0) > 0$ depending smoothly on y_0. When $\|y_0\|^2 > 2\varepsilon$ we get $r(y_0) = \sqrt{\|y_0\|^2 - \varepsilon}$.

One concludes by showing that Figure 6.17-(right) is in fact diffeomorphic to Figure 6.10-(right). Therefore M' is $M_{-\varepsilon}$ with a k-kandle attached. The explicit diffeomorphism is left as an exercise. □

6.5. Classification of surfaces

In the previous section we have seen that every compact smooth manifold without boundary decomposes into simple pieces called *handles*. We now use this construction to classify all compact surfaces.

6.5.1. The main theorem. We defined in Section 6.3.2 the genus-g surface S_g as the connected sum of g tori.

Theorem 6.5.1. *Every compact connected and orientable surface S without boundary is diffeomorphic to S_g, for some $g \geq 0$.*

Proof. We pick a handle decomposition of S. This consists of some 0-handles, then 1-handles attached to these 0-handles, and finally 2-handles attached to the result.

We first make an observation that is valid in all dimensions: if we attach a 1-handle to two distinct 0-handles as in Figure 6.18, this is equivalent to making a boundary connected sum of two discs, so the result is again a disc. Therefore we can replace the two 0-handles and the 1-handle altogether with a singe 0-handle, thus simplifying the handle decomposition.

After finitely many such moves, we may suppose that in the handle decomposition of S every 1-handle is attached twice to the same 0-handle. Since S is connected, this easily implies that there is only one 0-handle.

A dual argument works for the 2-handles. Note that every 1-handle is incident to two 2-handles, attached to the two long sides of the 1-handle. If the 2-handles are distinct, then the 1-handle together with the two incident 2-handles form again a picture like in Figure 6.18, and can thus be replaced by a single disc, that is a single 2-handle. After finitely many moves of this type, we easily end with a single 2-handle.

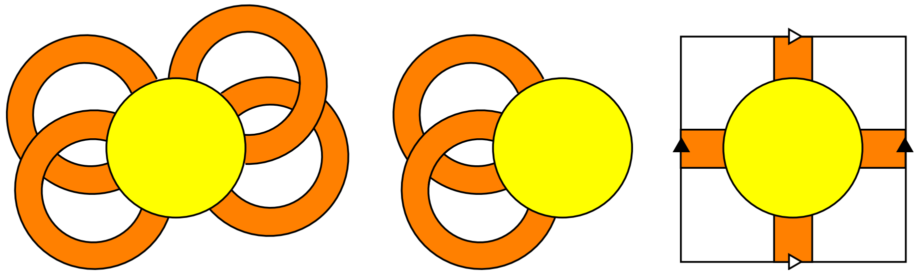

Figure 6.19. The 0-handle and some 1-handles (left). Two interlaced 1-handles (centre). Two interlaced handles form a handle decomposition of a holed torus, seen here as a square with opposite edges identified, with the white hole removed (right).

We have simplified the handle decomposition of S so that it has only one 0- and one 2-handle. If there are no 1-handles, then S decomposes into a 0- and a 2-handle and is hence diffeomorphic to S^2 by Corollary 6.2.16.

Suppose that there are 1-handles. Every 1-handle is a topological rectangle attached to the 0-handle along its short sides, as in Figure 6.19-(left). Up to diffeomorphism, there are two ways of attaching a 1-handle: with or without a twist. However, twists produce Möbius strips, which are excluded since S is orientable. So every 1-handle is attached without a twist, as in the figure.

Since there is only one 2-handle, the union of the 0- and 1-handles is a surface with connected boundary. This implies that every 1-handle must be interlaced with some other 1-handle as in Figure 6.19-(centre). Let $S' \subset S$ be the subsurface consisting of the 0-handle and these two 1-handles. Figure 6.19-(right) shows that S' is diffeomorphic to a torus with a hole. Therefore if we substitute S' with a single 0-handle, that is a disc, we find a simpler handle decomposition of a new surface S'' such that

$$S = S'' \# T.$$

We conclude by induction on the number of 1-handles that S is a connected sum of some g tori. $\square$

In the next chapters we will prove that S_g is not diffeomorphic to $S_{g'}$ if $g \neq g'$, so the genus of a surface fully characterises the surface up to diffeomorphism.

6.6. Exercises

The Euler characteristic of a surface S_g is $\chi(S_g) = 2 - 2g$. This can be taken as a definition here.

Exercise 6.6.1. Pick any positive integers $g, g', d \geq 1$. Show that if $\chi(S_g) = d\chi(S_{g'})$ then there is a degree-d covering $S_g \to S_{g'}$.

Exercise 6.6.2. Let M and N be two connected oriented n-manifolds of dimension $n \geq 3$. Show that

$$\pi_1(M\#N) \cong \pi_1(M) * \pi_1(N)$$

where $*$ is the free product of groups.

CHAPTER 7

Differential forms

Smooth functions from $\mathbb{R}$ to $\mathbb{R}$ can be summed, multiplicated, composed, derived, and integrated. Is there any kind of tensor field on a general manifold M for which these five possible types of manipulations still make sense? Yes! These are the *differential forms*, and they are, together with vector fields, among the most powerful objects one can encounter on a smooth manifold.

Differential forms can be summed, multiplicated, pulled back along any smooth map, derived, and integrated along submanifolds. They can be used to talk about volumes on manifolds. Their derivation generalises the notions of gradient, curl, and divergence in Euclidean space. The interplay between derivation and integration culminates with the *Stokes' Theorem* which generalises various statements that one encounter in analysis relating the integration of objects on domains of $\mathbb{R}^n$ and on their boundaries.

7.1. Differential forms

We introduce the main protagonist of this chapter.

7.1.1. Definition. Let M be a smooth n-manifold, possibly with boundary. A *differential k-form* (shortly, a *k-form*) is a section ω of the alternating bundle

$$\Lambda^k(M)$$

over M, see Section 4.3.4. In other words, for every $p \in M$ we have an antisymmetric multilinear form

$$\omega(p)\colon \underbrace{T_pM \times \cdots \times T_pM}_{k} \longrightarrow \mathbb{R}$$

that varies smoothly with $p \in M$.

Example 7.1.1. A 1-form is a section of $\Lambda^1(M) = T^*M$, that is a covector field. As an important example, the differential df of a smooth function $f\colon M \to \mathbb{R}$ is a 1-form, see Section 4.3.2. This example is not exhaustive: we will see that some 1-forms are not the differential of any function.

By Corollary 2.4.10, every k-form with $k > n$ is necessarily trivial. The vector space of all the k-forms on M is denoted by

$$\Omega^k(M) = \Gamma(\Lambda^k M).$$

7.1.2. Exterior product. Recall from Section 2.4.3 that the exterior algebra $\Lambda^*(V)$ of a real vector space V is equipped with the exterior product $\wedge$. Let now ω and η be a k-form and a h-form on a manifold M. Their *exterior product* is the $(k+h)$-form $\omega \wedge \eta$ defined pointwise by setting

$$(\omega \wedge \eta)(p) = \omega(p) \wedge \eta(p).$$

As in Section 2.4.3, the space

$$\Omega^*(M) = \bigoplus_{k \geq 0} \Omega^k(M)$$

inherits the structure of an anticommutative associative algebra such that

$$\omega \wedge \eta = (-1)^{hk} \eta \wedge \omega$$

and if k is odd we get

$$\omega \wedge \omega = 0.$$

This holds in particular for every 1-form ω.

7.1.3. In coordinates. As usual, differential forms may be written quite conveniently in coordinates.

Let U be an open subset of $\mathbb{R}^n$ or $\mathbb{R}^n_+$. Recall that for some notational reasons it is preferable to denote the canonical basis of $\mathbb{R}^n$ by

$$\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}.$$

For similar reasons, we will now write the dual basis of $(\mathbb{R}^n)^* = \mathbb{R}^n$ as

$$dx^1, \ldots, dx^n.$$

The notation is appropriate because dx^i is the differential of the linear map $x \mapsto x^i$. We have seen in Section 2.4.4 that the vector space $\Lambda^k(\mathbb{R}^n)$ has dimension $\binom{n}{k}$ and a basis consists of all the elements

$$dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

where $1 \leq i_1 < \ldots < i_k \leq n$ vary. Therefore we can write any k-form ω in U in the following way:

$$\omega = \sum_{i_1 < \cdots < i_k} f_{i_1, \ldots, i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

where $f_{i_1, \ldots, i_k}$ is a smooth function on U. We may simplify the notation by considering multi-indices $I = (i_1, \ldots, i_k)$ with $i_1 < \cdots < i_k$ and writing

$$\omega = \sum_I f_I dx^I.$$

Here of course $dx^I = dx^{i_1} \wedge \cdots \wedge dx^{i_k}$.

Example 7.1.2. The differential of a function $f \colon U \to \mathbb{R}$ is

$$df = \frac{\partial f}{\partial x^1} dx^1 + \ldots + \frac{\partial f}{\partial x^n} dx^n.$$

Example 7.1.3. The following are 1-forms in $\mathbb{R}^3$:

$$x^2dy - xe^y dz, \qquad xdx + ydy + zdz$$

and the following are 2-forms:

$$xdx \wedge dy + x^3dy \wedge dz, \qquad xdy \wedge dz - ydx \wedge dz + zdx \wedge dz.$$

Remark 7.1.4. Every n-form in $U \subset \mathbb{R}^n$ is of the type

$$f dx^1 \wedge \cdots \wedge dx^n$$

for some smooth function $f : U \to \mathbb{R}$. Therefore n-forms on open sets $U \subset \mathbb{R}^n$ are somehow like smooth functions on U, but one should not go too far with this analogy, because forms and functions are intrinsically different objects!

It is sometimes useful to write a form as a linear combination of elements $dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ without the strict hypothesis $i_1 < \ldots < i_k$. One has to take care that the notation is not unique in this case, for instance

$$\omega = dx \wedge dy = -dy \wedge dx = \frac{1}{2}dx \wedge dy - \frac{1}{2}dy \wedge dx.$$

It suffices to keep in mind the following relations:

$$dx^i \wedge dx^j = -dx^j \wedge dx^i, \qquad dx^i \wedge dx^i = 0.$$

Example 7.1.5. With these rules in mind, it is also easy to write the wedge product of two differential forms. For instance:

$$(xz^2dy + xdz) \wedge (e^y dx \wedge dz) = -xe^y z^2 dx \wedge dy \wedge dz.$$

7.1.4. Change of coordinates. On a chart, every form ω may be expressed uniquely as a linear combination

$$\omega = \sum_{i_1 < \cdots < i_k} f_{i_1,\ldots,i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

If we use another chart with variables $\bar{x}$ we get

$$\omega = \sum_{i_1 < \cdots < i_k} \bar{f}_{i_1,\ldots,i_k} d\bar{x}^{i_1} \wedge \cdots \wedge d\bar{x}^{i_k}$$

for some new functions $\bar{f}$. How can we pass from one expression to the other? The differentials dx^i are elements of $(\mathbb{R}^n)^*$ and hence change as follows

$$d\bar{x}^i = \frac{\partial \bar{x}^i}{\partial x^j} dx^j.$$

The notation dx^i is designed to help us write this equation correctly. We can then plug this expression in the linear combination to pass from one notation to the other.

Example 7.1.6. Consider the 2-form $\omega = z dx \wedge dy$ on the open set $U = \{x, y, z > 0\}$. We change the coordinates via $x = \bar{x}^2$, $y = \bar{y} + \bar{z}$, $z = \bar{y}$. Then

$$dx = 2\bar{x} d\bar{x}, \qquad dy = d\bar{y} + d\bar{z}, \qquad dz = d\bar{y}$$

and by substituting we see that ω in the new coordinates is read as

$$\omega = (\bar{y})(2\bar{x} d\bar{x}) \wedge (d\bar{y} + d\bar{z}) = 2\bar{x}\bar{y} d\bar{x} \wedge d\bar{y} + 2\bar{x}\bar{y} d\bar{x} \wedge d\bar{z}.$$

An interesting case occurs when we consider n-forms in a n-dimensional manifold. Here on a chart we have

$$\omega = f dx^1 \wedge \cdots \wedge dx^n$$

and Proposition 2.4.15 yields the following simple formula:

$$\omega = f \det \left(\frac{\partial x^i}{\partial \bar{x}^j} \right) d\bar{x}^1 \wedge \cdots \wedge d\bar{x}^n. \tag{11}$$

This equality is very much similar to the change of coordinates formula for integration given in Section 1.3.8, and this is in fact a crucial feature of differential forms: they can be meaningfully integrated on manifolds, as we will soon see.

7.1.5. Support. Let M be a n-manifold and ω be a k-form on M. We define the *support* of ω to be the closure in M of the set of all the points p such that $\omega(p) \neq 0$. Using bump functions, one can easily construct plenty of non-trivial k-forms in $\mathbb{R}^n$ and $\mathbb{R}^n_+$ having compact support.

Moreover, for every k-form ω on M and every open covering U_i of M, we can pick a partition of unity ρ_i subordinate to the covering and write

$$\omega = \sum_i \rho_i \omega.$$

The support of $\rho_i\omega$ is contained in U_i for every i, and this possibly infinite sum makes sense because it is finite at every point $p \in M$. One can in this way write every k-form ω as a (possibly infinite, but locally finite) sum of compactly supported k-forms $\rho_i\omega$. If ω is already compactly supported, the sum is finite.

7.1.6. Pull-back. When we introduced tensors in Chapter 2, the roles of V and V^* were somehow interchangeable, because each space is just the dual of the other. This symmetry is now broken when we talk about manifolds and tensor fields, and it turns out that tensor fields of type $(0, k)$ are sometimes better behaved than those of type $(h, 0)$.

We explain this phenomenon. Let $f: M \to N$ be any smooth map between two manifolds. We have already alluded to the fact that a vector field cannot be transported along f in general, neither forward from M to N nor backwards from N to M. On the other hand, every tensor field α of some type $(0, k)$ on

N may be transported back to a tensor field $f^*\alpha$ of the same type $(0, k)$ on M, by setting

$$f^*\alpha(p)(v_1, \dots, v_k) = \alpha\big(f(p)\big)\big(df_p(v_1), \dots, df_p(v_k)\big) \tag{12}$$

for every $p \in M$ and every $v_1, \dots, v_k \in T_pM$. The tensor field $f^*\alpha$ is the *pull-back* of α along f. If α is (anti-)symmetric, then $f^*\alpha$ also is.

In particular, the pull-back of a k-form ω in N is a k-form $f^*\omega$ in M. We get a morphism of algebras

$$f^*\colon \Omega^*(N) \longrightarrow \Omega^*(M).$$

In particular, we have

$$f^*(\omega \wedge \eta) = f^*(\omega) \wedge f^*(\eta). \tag{13}$$

As usual, we can describe this operation in coordinates: let $f\colon U \to V$ be a smooth map between two open subsets $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$, and

$$\omega = \sum_{i_1 < \ldots < i_k} g_{i_1, \ldots, i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

be a k-form in V. We get

$$f^*\omega = \sum_{i_1 < \ldots < i_k} \big(g_{i_1, \ldots, i_k} \circ f\big) df^{i_1} \wedge \cdots \wedge df^{i_k}$$

where $f^i\colon U \to \mathbb{R}$ is the i-th coordinate of f and df^i its differential. This equality is proved (exercise) by showing that it satisfies (12), using (13).

Example 7.1.7. Consider $f\colon \mathbb{R}^3 \to \mathbb{R}^2$, $f(t, u, v) = (tu, uv)$ and the 2-form $\omega = x dx \wedge dy$ on $\mathbb{R}^2$. We get

$$\begin{aligned} f^*\omega &= tu df_1 \wedge df_2 = tu(u dt + t du) \wedge (v du + u dv) \\ &= tu^2 v dt \wedge du + tu^3 dt \wedge dv + t^2u^2 du \wedge dv. \end{aligned}$$

7.1.7. Contraction. Let M be a manifold and X be a vector field in M. The contraction defined in Section 2.4.6 extends pointwise to a linear map

$$\iota_X\colon \Omega^k(M) \longrightarrow \Omega^{k-1}(M)$$

that sends $\omega \in \Omega^k(M)$ to the $(k-1)$-form $\iota_X(\omega)$ that acts as

$$\iota_X(\omega)(p)(v_1, \dots, v_{k-1}) = \omega(X(p), v_1, \dots, v_{k-1}).$$

7.2. Integration

We now show that k-forms are designed to be integrated along k-submanifolds.

7.2.1. Integration. Consider a n-form

$$\omega = f\, dx^1 \wedge \cdots \wedge dx^n$$

on some open subset V of $\mathbb{R}^n$ or $\mathbb{R}^n_+$, having compact support. We define the *integral* of ω over V simply and naïvely as

$$\int_V \omega = \int_V f.$$

Let now $\psi\colon V \to V'$ be an orientation-preserving diffeomorphism between open sets in $\mathbb{R}^n$ or $\mathbb{R}^n_+$, and denote by $\psi_*\omega = (\psi^{-1})^*\omega$ the n-form transported along ψ. Here is the crucial property that characterises differential forms:

Proposition 7.2.1. *We have*

$$\int_V \omega = \int_{V'} \psi_*\omega.$$

Proof. Combine (11), where $\det > 0$ since ψ is orientation-preserving, with the change of coordinates law for multiple integrals, see Section 1.3.8. □

It is really important that ψ be orientation-preserving: if ψ is orientation-reversing, then a minus sign appears in the equality. Encouraged by this result, we now want to extend integration of forms from open subsets of $\mathbb{R}^n$ to arbitrary oriented manifolds.

Let M be an oriented n-manifold, possibly with boundary, and ω be a n-form on M with compact support. We now define the *integral of* ω *over* M

$$\int_M \omega$$

as follows. If the support of ω is fully contained in the domain U of a chart $\varphi\colon U \to V$ we set

$$\int_M \omega = \int_V \varphi_*\omega.$$

The definition is well-posed because it is chart-independent thanks to Proposition 7.2.1. More generally, if the support of ω is not contained in the domain of any chart, we pick an oriented atlas $\{\varphi_i\colon U_i \to V_i\}$ on M and a partition of unity ρ_i subordinated to the covering U_i. We decompose ω as a finite sum $\omega = \sum_i \rho_i\omega$ and define

$$\int_M \omega = \sum_i \int_M \rho_i\omega.$$

Proposition 7.2.2. *This definition is well-posed.*

Proof. Let $\{\varphi'_j\colon U'_j \to V'_j\}$ be another compatible oriented atlas and ρ'_j a partition of unity subordinated to U'_j. We find

$$\int_M \omega = \sum_i \int_M \rho_i\omega = \sum_i \int_M \Big(\sum_j \rho'_j\Big)\rho_i\omega = \sum_{i,j} \int_M \rho'_j\rho_i\omega.$$

In this expression the roles of ρ_i and ρ'_j can be interchanged, so if we use the partition of unity ρ'_j to calculate the integral we get the same result. $\square$

The following properties follow readily from the definitions. Let ω be a compactly supported n-form on an oriented n-manifold M, possibly with boundary. We denote as $-M$ the manifold M with the opposite orientation.

Proposition 7.2.3. *We have*

$$\int_{-M} \omega = -\int_M \omega.$$

If $f\colon M \to N$ is an orientation-preserving diffeomorphism, then

$$\int_M \omega = \int_N f_* \omega.$$

Remark 7.2.4. We observed in Remark 7.1.4 that on a chart a n-form looks like a function, but we warned the reader that the two notions are quite different on a general manifold M. As opposite to n-forms, functions in M cannot be integrated in any meaningful way; conversely, the value $\omega(p)$ of a n-form ω at $p \in M$ is not a number, in any reasonable sense. Shortly: functions can be evaluated at points, and n-forms can be integrated on sets, but not the converse.

7.2.2. Examples. In practice, nobody uses partitions of unity to integrate a n-form on a manifold, because the partition of unity is typically not explicit. Instead, we prefer to subdivide the manifold into small pieces where the n-form may be integrated easily. We explain briefly the details.

Let M be a smooth n-manifold, possibly with boundary. Recall the notion of Borel subset from Section 3.11.1. If ω is a compactly supported n-form on M, we can define the integral $\int_S \omega$ over a Borel set $S \subset M$ using a partition of unity in the same way as we did above.

Proposition 7.2.5. *If the support of ω is contained in a Borel set S that is a countable disjoint union of Borel sets S_i, then*

$$\int_S \omega = \sum_i \int_{S_i} \omega.$$

Proof. The equality holds for Borel sets in $\mathbb{R}^n$ because it is a property of Lebesgue integration; via a partition of unity we can extend it to M. $\square$

Recall that having measure zero is a well-defined property for Borel subsets of any smooth manifold. If the complement of $S \subset M$ has measure zero, then

$$\int_M \omega = \int_S \omega$$

because the integral over $M \setminus S$ is zero. So we can remove from M any zero-measure set to get a more comfortable domain S and integrate ω there.

Example 7.2.6. Consider the n-dimensional torus $T = S^1 \times \cdots \times S^1$ where every point has some coordinates $(\theta^1, \ldots, \theta^n)$, and the n-form

$$\omega = d\theta^1 \wedge \cdots \wedge d\theta^n.$$

We have

$$\int_T \omega = \int_U \omega = \int_{(0,2\pi)\times\cdots\times(0,2\pi)} 1 = (2\pi)^n$$

by using the open chart $U = (0, 2\pi) \times \cdots \times (0, 2\pi)$ whose complement has measure zero.

We can integrate n-forms on oriented n-manifolds, for all $n \geq 1$. It is sometimes useful to extend this operation to zero-dimensional manifolds. Recall that an orientation for a point p is the assignment of a sign ± 1 and a 0-form on p is just a function f, that is a number $f(p)$. We define the integral of f on p as $\pm f(p)$ according to the orientation of p.

7.2.3. Integration on submanifolds. By combining pull-backs and integration, we get a nice new tool: we can integrate k-forms along k-submanifolds.

Let M be a smooth manifold, possibly with boundary, and ω be a fixed compactly supported k-form on M. For every oriented closed submanifold $S \subset M$ of dimension k, possibly with boundary, we may define the integral of ω along S as follows:

$$\int_S \omega = \int_S i^*\omega$$

where $i\colon S \hookrightarrow M$ is the inclusion map. Quite remarkably, we can use ω to assign a real number to every closed k-submanifold $S \subset M$.

Remark 7.2.7. Since the submanifold S is closed, the support of $i^*\omega$ is compact and the integral makes sense. More generally, it suffices that the intersection of the support of ω with S be compact for the integral to make sense. For instance, this holds for every $\omega \in \Omega^k(M)$ if S is itself compact.

Shortly: functions can be evaluated at points, and k-forms can be integrated along oriented k-submanifolds.

Exercise 7.2.8. Consider the torus $T = S^1 \times S^1$ with coordinates (θ^1, θ^2) and the 1-form $\omega = d\theta^1$. Consider the 1-submanifold $\gamma_i = \{\theta^i = 0\}$ for $i = 1, 2$, oriented like S^1. We have

$$\int_{\gamma_1} \omega = 0, \qquad \int_{\gamma_2} \omega = 2\pi.$$

7.2.4. Submanifolds of (co-)dimension 1 in $\mathbb{R}^n$. The integration of a k-form along a k-submanifold of $\mathbb{R}^n$ may be expressed in a nice geometric way when $k = 1$ or $k = n - 1$, by interpreting the form as a vector field. This discussion is particularly relevant for $\mathbb{R}^3$ since it involves both 1- and 2-forms.

Every 1-form

$$\omega = f_1 dx^1 + \cdots + f_n dx^n$$

in $\mathbb{R}^n$ defines a vector field X with coordinates $X^i = f_i$, and viceversa every vector field X in $\mathbb{R}^n$ defines a 1-form ω. Here we are using implicitly the identification of $\mathbb{R}^n$ with its dual $(\mathbb{R}^n)^*$ furnished by the canonical basis, that is by the Euclidean metric tensor: there is no way to pass from 1-forms to vector fields on a generic smooth manifold (we need a metric tensor for that).

Let $C \subset \mathbb{R}^n$ be an oriented closed 1-submanifold, possibly with boundary (a curve). Let τ be the unit tangent field to C, oriented coherently with C. We suppose that ω has compact support.

Proposition 7.2.9. *We have*

$$\int_C \omega = \int_C X \cdot \tau.$$

Proof. We parametrise locally C as the image of an embedding $\gamma\colon (a,b) \to \mathbb{R}^n$ and write $\gamma(t) = (x^1(t), \ldots, x^n(t))$. We get

$$\begin{aligned}\int_{\gamma(a,b)} \omega = \int_{\gamma(a,b)} f_1 dx^1 + \cdots + f_n dx^n &= \int_a^b \left(f_1 \frac{dx^1}{dt} + \cdots + f_n \frac{dx^n}{dt} \right) dt \\ = \int_a^b X \cdot \gamma'(t) dt = \int_a^b X \cdot t \,\|\gamma'(t)\| dt &= \int_C X \cdot \tau.\end{aligned}$$

The proof is complete. □

We have discovered that the integral of a 1-form on a curve C equals the integral of the tangential component of the corresponding vector field. We now look at the codimension-1 case. A $(n-1)$-form in $\mathbb{R}^n$ may be written as

$$\omega = \sum_{i=1}^n f_i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n$$

where $\widehat{dx^i}$ indicates that this symbol is missing. This also defines a vector field X with coordinates $X^i = (-1)^{i+1} f_i$, and conversely a vector field defines a $(n-1)$-form. (Again, we can do this in $\mathbb{R}^n$, but beware that no natural identification between $(n-1)$-forms and vector fields exists on a generic manifold.)

Let $S \subset \mathbb{R}^n$ be an oriented closed codimension-1 submanifold, possibly with boundary, for instance a surface in $\mathbb{R}^3$. The orientation of S defines a unit normal vector field ν on S, determined by requiring that $\nu, v_1, \ldots, v_{n-1}$ be a positive basis for $\mathbb{R}^n$ if $v_1, \ldots, v_{n-1}$ is a positive basis for T_pS at any $p \in S$. Suppose that ω has compact support.

Proposition 7.2.10. *We have*

$$\int_S \omega = \int_S X \cdot \nu.$$

Proof. We can parametrise S locally as the image of a map $\varphi\colon U \to \mathbb{R}^n$ for some open subset $U \subset \mathbb{R}^{n-1}$. We use the coordinates $t^1, \ldots, t^{n-1}$ for U and $x^1, \ldots, x^n$ for $\mathbb{R}^n$. We write $\varphi(t) = x(t)$ and get

$$\begin{aligned}\int_{\varphi(U)} \omega &= \int_{\varphi(U)} \sum_{i=1}^n f_i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\ &= \int_U \sum_{i=1}^n f_i \frac{\partial x^1}{\partial t^{j_1}} dt^{j_1} \wedge \cdots \wedge \frac{\partial x^n}{\partial t^{j_{n-1}}} dt^{j_{n-1}} \\ &= \int_U \sum_{i=1}^n f_i J_i dt^1 \cdots dt^{n-1} \\ &= \int_U \sum_{i=1}^n X^i (-1)^{i-1} J_i dt^1 \cdots dt^{n-1}\end{aligned}$$

where J_i is the determinant of the matrix obtained by deleting the i-th row of $\frac{\partial x^j}{\partial t^k}$. The vector $J = (J_1, -J_2, \ldots, (-1)^{n-1} J_n)$ is a positive multiple of ν and its norm is the infinitesimal volume of S. Therefore we get

$$\int_{\varphi(U)} \omega = \int_U X \cdot J\, dt^1 \cdots dt^n = \int_U X \cdot \nu \, \|J\| dt^1 \cdots dt^n = \int_{\varphi(U)} X \cdot \nu.$$

The proof is complete. $\square$

We have proved that the integral of a $(n-1)$-form along a hypersurface S equals the integral of the normal component of the corresponding vector field.

7.2.5. Volume form. A smooth manifold is not equipped with any canonical notion of "volume" for its Borel subsets. The most convenient way to introduce one is to select a preferred differential form called a *volume form*.

Let M be an oriented n-manifold, possibly with boundary.

Definition 7.2.11. A *volume form* in M is a n-form ω such that

$$\omega(p)(v_1, \ldots, v_n) > 0$$

for every $p \in M$ and every positive basis $v_1, \ldots, v_n$ of T_pM.

Let ω be a volume form on M and $S \subset M$ be a Borel set with compact closure. It makes sense to define the *volume* of S as

$$\mathrm{Vol}(S) = \int_S \omega.$$

Example 7.2.12. The *Euclidean volume form* on $\mathbb{R}^n$ or $\mathbb{R}^n_+$ is

$$\omega = dx^1 \wedge \cdots \wedge dx^n.$$

The volume that it defines on $\mathbb{R}^n$ or $\mathbb{R}^n_+$ is the ordinary Lebesgue measure.

Here is the crucial property of volume forms:

Proposition 7.2.13. *We have* $\mathrm{Vol}(S) \geq 0$ *for every Borel subset* $S \subset M$ *with compact closure. If* S *has non-empty interior, then* $\mathrm{Vol}(S) > 0$.

Proof. If we use only orientation-preserving charts, the form ω transforms into n-forms $f\,dx^1 \wedge \cdots dx^n$ with $f(x) > 0$ for every x. □

As in ordinary Lebesgue measure theory, we can now define $\mathrm{Vol}(S)$ for every Borel set S, as the supremum of the volumes of the Borel sets with compact closure contained in S. The volume may (or may not) be infinite if S has not compact closure. We have obtained a measure on all the Borel sets in M, that is we have the countable additivity

$$\mathrm{Vol}(S) = \sum \mathrm{Vol}(S_i)$$

whenever S is the disjoint union of countably many Borel sets S_i.

Of course different selections of the volume form ω give rise to different measures, and there is no way to choose a "preferred" volume form ω on an arbitrary oriented manifold M.

Proposition 7.2.14. *If* ω *is a volume form and* $f\colon M \to \mathbb{R}$ *is a strictly positive function, then* $\omega' = f\omega$ *is another volume form. Every volume form* ω' *may be constructed from* ω *in this way.*

Proof. The first assertion is obvious, and the converse follows from the fact that $\Lambda^n(T_pM)$ has dimension 1 and hence for every ω, ω' we may define $f(p)$ as the unique positive number such that $\omega'(p) = f(p)\omega(p)$. □

We also note that volume forms always exist:

Proposition 7.2.15. *If* M *is oriented, there is always a volume form on* M.

Proof. Pick an oriented atlas $\{\varphi_i\colon U_i \to V_i\}$ and a partition of unity ρ_i subordinate to the covering $\{U_i\}$. We define

$$\omega(p) = \sum_i \rho_i(p)\varphi_i^*(dx^1 \wedge \cdots \wedge dx^n)$$

and get a volume form ω. Indeed for every $p \in M$ and positive basis $v_1, \ldots, v_n$ at T_pM the number $\omega(p)(v_1, \ldots, v_n)$ is a finite sum of strictly positive numbers with strictly positive coefficients $\rho_i(p)$, so it is strictly positive. □

7.3. Exterior derivative

At various places in this book we introduce some objects, typically some tensor fields, and then we try to "derive" them in a meaningful way. We now show that differential forms can be derived quite easily, through an operation called *exterior derivative*, that transforms k-forms into $(k+1)$-forms and extends the differential of functions (that transform functions, that is 0-forms, into 1-forms).

7.3.1. Definition. Let ω be a k-form in a smooth manifold M, possibly with boundary. We define the *exterior derivative* $d\omega$, a new $(k+1)$-form on M. We first consider the case where M is an open set in $\mathbb{R}^n$ or $\mathbb{R}^n_+$. Then

$$\omega = \sum_{i_1<\cdots<i_k} f_{i_1,\ldots,i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

and we define

$$d\omega = \sum_{i_1<\cdots<i_k} df_{i_1,\ldots,i_k} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

Recall that $df_{i_1,\ldots,i_k}$ is a 1-form, hence $d\omega$ is a $(k+1)$-form. When ω is a 0-form, that is a function $\omega = f$, then $d\omega$ is the ordinary differential.

Example 7.3.1. Consider the form $\omega = xydx + xydz$ in $\mathbb{R}^3$. We get

$$d\omega = xdy \wedge dx + ydx \wedge dz + xdy \wedge dz.$$

We now extend this definition to an arbitrary smooth manifold M, as usual by considering charts: we just define $d\omega$ on any open chart as above.

Proposition 7.3.2. *The definition of $d\omega$ using charts is well-posed. The derivation induces a linear map*

$$d\colon \Omega^k(M) \longrightarrow \Omega^{k+1}(M)$$

such that, for every $\omega \in \Omega^k(M)$ and $\eta \in \Omega^h(M)$ the following hold:

$$d(\omega\wedge\eta) = d\omega\wedge\eta + (-1)^k\omega\wedge d\eta, \tag{14}$$

$$d(d\omega) = 0. \tag{15}$$

Proof. We first prove the properties (14) and (15) on a fixed chart, and later we use these properties to show that the definition of $d\omega$ is chart-independent and hence well-posed.

Linearity of d is obvious, and using it we may suppose that $\omega = fdx^I$ and $\eta = gdx^J$ where I, J are some multi-indices. We get

$$\begin{aligned} d(\omega\wedge\eta) &= d(fg)\wedge dx^I \wedge dx^J = df\wedge dx^I \wedge gdx^J + dg \wedge fdx^I \wedge dx^J \\ &= d\omega\wedge\eta + (-1)^k\omega\wedge d\eta. \end{aligned}$$

If $\omega = fdx^I$ then

$$d(d\omega) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j} dx^i \wedge dx^j \wedge dx^I = 0$$

because $dx^i \wedge dx^j = -dx^j \wedge dx^i$ so the terms cancel in pairs.

Finally, we can prove that the definition is chart-independent, via the following trick: on open subsets $U \subset \mathbb{R}^n$, the derivation d may be characterised (exercise) as the unique linear map $d\colon \Omega^k(U) \to \Omega^{k+1}(U)$ that is the ordinary differential for $k = 0$ and that satisfies (14) and (15). Therefore two definitions of d on overlapping charts must coincide in their intersection. □

The exterior derivative commutes with the pull-back:

Exercise 7.3.3. If $\varphi\colon M \to N$ is smooth and $\omega \in \Omega^k(N)$, we get

$$d(\varphi^*\omega) = \varphi^*(d\omega).$$

Hint. Prove it when $\omega = f$ is a function, and when $\omega = df$ is the differential of a function. Use Proposition 7.3.2 to extend it to any $\omega = f_I dx^I$. $\square$

7.3.2. Action on vector fields. We may characterise the exterior derivative of k-forms by describing how it interacts with vector fields. For instance, the differential df of a function f acts on vector fields $X \in \mathfrak{X}(M)$ as

$$df(X) = X(f).$$

Concerning 1-forms, we get the following:

Exercise 7.3.4. If $\omega \in \Omega^1(M)$ is a 1-form and and $X, Y \in \mathfrak{X}(M)$ are vector fields, we get

$$d\omega(X,Y) = X\big(\omega(Y)\big) - Y\big(\omega(X)\big) - \omega\big([X,Y]\big).$$

Hint. Again, everything is local, so work in coordinates. $\square$

A similar formula holds also for the differential $d\omega$ of a k-form.

7.3.3. Gradient, curl, and divergence. We now show that the inspiring formula $d(d\omega) = 0$ generalises a couple of familiar equalities about functions and vector fields in $\mathbb{R}^3$.

Let $U \subset \mathbb{R}^3$ be an open set. Recall that the *gradient* of a function $f : U \to \mathbb{R}$ is the vector field

$$\nabla f = \left(\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \frac{\partial f}{\partial x^3}\right).$$

If X is a vector field in U, its *divergence* is the function

$$\mathrm{div} X = \frac{\partial X^1}{\partial x^1} + \frac{\partial X^2}{\partial x^2} + \frac{\partial X^3}{\partial x^3}$$

while its *curl* is the vector field

$$\mathrm{rot} X = \left(\frac{\partial X^3}{\partial x^2} - \frac{\partial X^2}{\partial x^3}, \frac{\partial X^1}{\partial x^3} - \frac{\partial X^3}{\partial x^1}, \frac{\partial X^2}{\partial x^1} - \frac{\partial X^1}{\partial x^2}\right).$$

As in Section 7.2.4, we may interpret a vector field X in U as a 1-form

$$\omega = X^1 dx^1 + X^2 dx^2 + X^3 dx^3$$

and vice-versa. We can also interpret a vector field X as a 2-form

$$\omega = X^1 dx^2 \wedge dx^3 + X^2 dx^3 \wedge dx^1 + X^3 dx^1 \wedge dx^2$$

and viceversa. Finally, we can interpret a 3-form as a function. Beware as usual that this interpretation is not allowed in an arbitrary smooth manifold.

Exercise 7.3.5. With this interpretation, the differential of a 0-, 1-, and 2-form in $\mathbb{R}^3$ corresponds to the gradient, curl, and divergence. That is we get a commutative diagram where vertical arrows are isomorphisms:

$$\begin{tikzcd}
\Omega^0(\mathbb{R}^3) \arrow[r, "d"] \arrow[d, equal] & \Omega^1(\mathbb{R}^3) \arrow[r, "d"] \arrow[d] & \Omega^2(\mathbb{R}^3) \arrow[r, "d"] \arrow[d] & \Omega^3(\mathbb{R}^3) \arrow[d] \\
C^\infty(\mathbb{R}^3) \arrow[r, "\nabla"] & \mathfrak{X}(\mathbb{R}^3) \arrow[r, "\mathrm{rot}"] & \mathfrak{X}(\mathbb{R}^3) \arrow[r, "\mathrm{div}"] & C^\infty(\mathbb{R}^3)
\end{tikzcd}$$

Here $d \circ d = 0$ transforms into the two well-known equalities

$$\mathrm{rot} \circ \nabla = 0, \qquad \mathrm{div} \circ \mathrm{rot} = 0.$$

7.3.4. Cartan's magic formula. Let now M be a manifold, possibly with boundary, and let X a vector field in M. Our toolbox contains an abundance of operators on k-forms, some being determined by X. We find the Lie derivative along X, the contraction along X, and the exterior derivative:

$$\mathcal{L}_X \colon \Omega^k(M) \to \Omega^k(M), \quad \iota_X \colon \Omega^k(M) \to \Omega^{k-1}(M), \quad d \colon \Omega^k(M) \to \Omega^{k+1}(M).$$

These three operators behave similarly with respect to the wedge product:

Proposition 7.3.6. *For every $\omega \in \Omega^k(M)$ and $\eta \in \Omega^h(M)$ we have:*

$$\begin{aligned}
\mathcal{L}_X(\omega \wedge \eta) &= (\mathcal{L}_X\omega) \wedge \eta + \omega \wedge (\mathcal{L}_X\eta), \\
\iota_X(\omega \wedge \eta) &= (\iota_X\omega) \wedge \eta + (-1)^k\omega \wedge (\iota_X\eta).
\end{aligned}$$

Proof. This follows from Exercises 5.4.14 and 2.7.4. $\square$

Compare with Proposition 7.3.2. We say that $\mathcal{L}_X$ is a *derivation*, while ι_X and d are *anti-derivations* because of the $(-1)^k$ sign in the formula. Note also that $\iota_X \circ \iota_X = 0$ and $d \circ d = 0$.

Proposition 7.3.7. *The following operators commute:*

$$\mathcal{L}_X \circ d = d \circ \mathcal{L}_X,$$

$$\mathcal{L}_X \circ \iota_X = \iota_X \circ \mathcal{L}_X.$$

Proof. This first equality holds because the exterior derivative d commutes with diffeomorphisms and with derivations of paths of forms. Hence

$$\begin{aligned}
\mathcal{L}_X(d\omega)(p) &= \frac{d}{dt}\Big|_{t=0} (\Phi_{-t})_*\big(d\omega(\Phi_t(p))\big) \\
&= \frac{d}{dt}\Big|_{t=0} d(\Phi_{-t})_*\big(\omega(\Phi_t(p))\big) \\
&= d\left(\frac{d}{dt}\Big|_{t=0} (\Phi_{-t})_*\big(\omega(\Phi_t(p))\big)\right) = d\big(\mathcal{L}_X(\omega)\big)(p).
\end{aligned}$$

Here Φ_t is the flow associated to X. The second is proved analogously. $\square$

The operators ι_X and d do not commute in general. The three are connected by a nice formula called *Cartan's magic formula*:

Theorem 7.3.8 (Cartan's magic formula). *The following holds:*

$$\mathcal{L}_X = d \circ \iota_X + \iota_X \circ d.$$

Proof. On a function f, the formula holds because $\iota_X(f) = 0$ and

$$\mathcal{L}_X(f) = X(f) = \iota_X(df).$$

On the 1-form df, the formula holds because $d(df) = 0$ and

$$\mathcal{L}_X(df) = d\mathcal{L}_X(f) = d(\iota_X(df)).$$

Propositions 7.3.2 and 7.3.6 show that both operators $\mathcal{L}_X$ and $d \circ \iota_X + \iota_X \circ d$ are derivations (the composition of two antiderivations is a derivation).

Every k-form ω may be written locally as a sum of wedge products $f dx^I$ of functions f and 1-forms dx^i. Cartan's equality holds for each factor f and dx^i. Since both sides of the equality are derivations, it holds also for ω. $\square$

7.4. Stokes' Theorem

We end up this chapter with Stokes' Theorem, that relates elegantly exterior derivatives and integration along manifolds with boundary.

7.4.1. The theorem. We first note that the whole theory of differential forms and integration applies also to manifolds with boundary with no modification. We then highlight a fascinating analogy: when we talk about forms ω we have

$$d(d\omega) = 0,$$

while when we deal with manifolds M with boundary we also get

$$\partial(\partial M) = \varnothing,$$

since the boundary of M is a manifold without boundary. Note also that d transforms a k-form into a $(k+1)$-form, while ∂ transforms a $(k+1)$-manifold into a k-manifold. The operations d and ∂ are beautifully connected by the Stokes' Theorem.

Let M be an oriented $(n+1)$-manifold with (possibly empty) boundary, and equip ∂M with the orientation induced by M.

Theorem 7.4.1 (Stokes' Theorem). *For every compactly supported n-form ω in an oriented $(n+1)$-manifold M possibly with boundary, we have*

$$\int_M d\omega = \int_{\partial M} \omega.$$

Proof. We first prove the theorem for $M = \mathbb{R}^{n+1}_+$. We have

$$\omega = \sum_{i=1}^{n+1} \omega_i$$

with

$$\omega_i = f_i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1}$$

where the hat indicates that the i-th term is missing. By linearity it suffices to prove the theorem for each ω_i individually. We have

$$\begin{aligned} d\omega_i &= df_i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1} \\ &= \sum_j \frac{\partial f_i}{\partial x^j} dx^j \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1} \\ &= \frac{\partial f_i}{\partial x^i} dx^i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1} \\ &= (-1)^{i-1} \frac{\partial f_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^{n+1}. \end{aligned}$$

In the third equality the terms with $j \neq i$ vanish because $dx^j \wedge dx^j = 0$. We now consider two cases separately. If $i \leq n$ we have

$$\begin{aligned} \int_{\mathbb{R}^{n+1}_+} d\omega_i &= (-1)^{i-1} \int_{\mathbb{R}^{n+1}_+} \frac{\partial f_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^{n+1} \\ &= (-1)^{i-1} \int_{\mathbb{R}^{n+1}_+} \frac{\partial f_i}{\partial x^i} dx^1 \cdots dx^{n+1} \\ &= (-1)^{i-1} \int_{\mathbb{R}^n_+} \left(\int_{\mathbb{R}} \frac{\partial f_i}{\partial x^i} dx^i \right) dx^1 \cdots \widehat{dx^i} \cdots dx^{n+1} = 0. \end{aligned}$$

When the $\wedge$ is not present in the expression, it means that we are just doing the usual Lebesgue integration of functions on some Euclidean space. In the last equality we have used that

$$\begin{aligned} \int_{\mathbb{R}} \frac{\partial f_i}{\partial x^i} dx^i = \lim_{t \to \infty} \big[& f_i(x^1, \dots, x^{i-1}, t, x^{i+1}, \dots, x^{n+1}) \\ & - f_i(x^1, \dots, x^{i-1}, -t, x^{i+1}, \dots, x^{n+1}) \big] = 0 - 0 = 0 \end{aligned}$$

because f_i has compact support. On the other hand, we also have

$$\int_{\partial \mathbb{R}^{n+1}_+} \omega_i = 0$$

because ω_i contains dx^{n+1} whose pull-back to $\partial \mathbb{R}^{n+1}_+$ vanishes.

If $i = n+1$ we get

$$\begin{aligned}
\int_{\mathbb{R}^{n+1}_+} d\omega_{n+1} &= (-1)^n \int_{\mathbb{R}^n} \left(\int_0^{+\infty} \frac{\partial f_{n+1}}{\partial x^{n+1}} dx^{n+1} \right) dx^1 \cdots dx^n \\
&= (-1)^n \int_{\mathbb{R}^n} \big(0 - f_{n+1}(x^1, \ldots, x^n, 0) \big) dx^1 \cdots dx^n \\
&= (-1)^{n+1} \int_{\mathbb{R}^n} f_{n+1}(x^1, \ldots, x^n, 0) dx^1 \cdots dx^n \\
&= \int_{\partial \mathbb{R}^{n+1}_+} f_{n+1} dx^1 \wedge \cdots \wedge dx^n = \int_{\partial \mathbb{R}^{n+1}_+} \omega_{n+1}.
\end{aligned}$$

The mysterious disappearance of the $(-1)^{n+1}$ sign at the end is due to the fact that the orientations on $\mathbb{R}^n$ and $\partial\mathbb{R}^{n+1}_+$ match only when n is odd (exercise).

We have proved the theorem for $M = \mathbb{R}^{n+1}_+$. On a general M we pick an atlas $\{\varphi_i \colon U_i \to V_i\}$ with $V_i \subset \mathbb{R}^{n+1}_+$ and a partition of unity ρ_i subordinate to U_i, so that $\omega = \sum_i \rho_i \omega$ is a finite sum (because ω has compact support). By linearity, it suffices to prove the theorem for each addendum $\rho_i\omega$, but in this case we can transport it via φ_i to a form in $\mathbb{R}^{n+1}_+$ and we are done. □

Corollary 7.4.2. *If M is an oriented n-manifold without boundary, for every compactly supported $(n-1)$-form ω we have*

$$\int_M d\omega = 0.$$

7.4.2. Some consequences. Some familiar theorems in multivariate analysis in $\mathbb{R}$, $\mathbb{R}^2$, or $\mathbb{R}^3$ may be seen as particular instances of Stokes' Theorem.

In the line $\mathbb{R}$, Stokes' Theorem is just the fundamental theorem of calculus. A bit more generally, we may consider an embedded oriented arc $\gamma \subset \mathbb{R}^3$ with endpoints p and q and a smooth function f defined on it. Stokes says that

$$\int_\gamma df = f(q) - f(p).$$

So in particular the result depends only on the endpoints of γ, not of γ itself.

In the plane $\mathbb{R}^2$, we may consider a 1-form

$$\omega = f dx + g dy$$

and calculate

$$d\omega = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx \wedge dy.$$

For every compact domain $D \subset \mathbb{R}^2$ bounded by a simple closed curve $C = \partial D$, Stokes' Theorem transforms into *Green's Theorem*:

$$\int_C f dx + g dy = \int_D \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx dy.$$

In the space $\mathbb{R}^3$, the boundary ∂D of a compact domain $D \subset \mathbb{R}^3$ is some surface, and we pick a vector field X on D. After interpreting X as a 2-form as in Sections 7.2.4 and 7.3.3, we apply Stokes' Theorem and get the *Divergence Theorem*:

$$\int_D \mathrm{div} X = \int_{\partial D} X \cdot n$$

where n is the normal vector to ∂D.

Finally, we can also consider an oriented surface $S \subset \mathbb{R}^3$ with some (possibly empty) boundary ∂S, and a vector field X in $\mathbb{R}^3$ supported on S. By interpreting X as a 1-form as in Sections 7.2.4 and 7.3.3 and applying Stokes' Theorem we get the *Kelvin – Stokes Theorem*:

$$\int_S \mathrm{rot} X \cdot n = \int_{\partial S} X \cdot t$$

where n is the unit normal field to S and t is the unit tangent field to ∂S, both oriented coherently with the orientations of S and $\mathbb{R}^3$.

We have proudly proved all these theorems (and many more!) at one time.

7.5. Metric tensors and differential forms

The theory of differential forms on a manifold M may be enriched by the presence of a *metric tensor* g. Metric tensors will be the protagonist of the third part of this book, and they make here only a fleeting appearance.

7.5.1. Metric tensors. A *metric tensor* on a manifold M, possibly with boundary, is a section g of the symmetric bundle

$$S^2(M)$$

such that $g(p)$ is a scalar product (that is, it is non-degenerate) for every $p \in M$. In other words, for every $p \in M$ we have a scalar product

$$g(p) \colon T_pM \times T_pM \longrightarrow \mathbb{R}$$

that varies smoothly with p. This notion will be of fundamental importance when we introduce Riemannian geometry in Chapter 9.

The scalar product $g(q)$ at $q \in M$ has some signature (p, m). One verifies easily that if M is connected the pair (p, m) does not depend on the chosen point $q \in M$ and we simply call it the *signature* of g.

Example 7.5.1. The *Euclidean metric tensor* g_E on $\mathbb{R}^n$ is

$$g_E(x, y) = \sum_{i=1}^n x^i y^i$$

where we have identified $T_p\mathbb{R}^n$ with $\mathbb{R}^n$ as usual.

7.5.2. A metric tensor induces a volume form. If M is oriented, every metric tensor g induces a natural volume form ω on M as follows. At every point $p \in M$, the tangent space T_pM is equipped with an orientation and a scalar product $g(p)$, and as in Section 2.5.3 we define ω unambiguously by requiring

$$\omega(p)(v_1, \ldots, v_n) = 1$$

on every positive orthornormal basis $v_1, \ldots, v_n$ of T_pM. To show that ω varies smoothly with p we calculate ω on coordinates.

Proposition 7.5.2. *If g_{ij} is a metric tensor on $U \subset \mathbb{R}^n$, then*

$$\omega = \sqrt{|\det g_{ij}|}\, dx^1 \wedge \ldots \wedge dx^n.$$

Proof. Let $v^1, \ldots, v^n$ be a positive g-orthonormal basis for $(\mathbb{R}^n)^*$. We get

$$\omega = v^1 \wedge \ldots \wedge v^n = \det A\, dx^1 \wedge \ldots \wedge dx^n$$

where $v^i = A^i_j e^j$. Now $A^l_i g^{ij} A^k_j = \delta^{lk}$ gives $(\det A)^2 \det g^{-1} = 1$ and hence we get $\det A = \sqrt{|\det g|}$. $\square$

In particular the *volume* of a Borel subset $S \subset U$ is

$$\mathrm{Vol}(S) = \int_S \sqrt{|\det g_{ij}|}\, dx^1 \cdots dx^n.$$

7.5.3. Euclidean volume form. The Euclidean metric tensor induces the *Euclidean volume form*

$$\omega_E = dx^1 \wedge \ldots \wedge dx^n$$

on $\mathbb{R}^n$, already encountered in Example 7.2.12, which acts as

$$\omega_E(p)(v_1, \ldots, v_n) = \det \Big(v_1 \cdots v_n \Big)$$

at every $p \in \mathbb{R}^n$.

More generally, we may define a *Euclidean volume form* ω on every oriented k-submanifold $M \subset \mathbb{R}^n$. We do this in two steps: first, we restrict the Euclidean metric tensor from $\mathbb{R}^n$ to its subspace T_pM for every $p \in M$, thus obtaining a (positive definite) metric tensor on M. Then we use this metric tensor on M to get a volume form ω. Again $\omega(p)$ is characterised by the property that $\omega(p)(v_1, \ldots, v_k) = 1$ on every positive orthonormal basis $v_1, \ldots, v_k$ for T_pM. It is also characterised by the fact that the integral of ω along a Borel subset $D \subset M$ is the ordinary k-volume of D as defined in multivariable analysis.

Note that we are using the Euclidean scalar product here to define ω on M. A volume form on a smooth manifold N does *not* induce in general a volume form on a lower-dimensional submanifold M. The metric tensor is needed here.

The codimension-1 case is particularly simple.

Proposition 7.5.3. *Let $M \subset \mathbb{R}^n$ be an oriented $(n-1)$-manifold. The volume form on M is the pull-back of*

$$\omega = \sum_{i=1}^{n} (-1)^{i-1} n^i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n$$

where $n = (n^1, \ldots, n^n)$ is the unit normal vector field on M.

Proof. Proposition 7.2.10 says that

$$\int_D \omega = \int_D n \cdot n = \int_D 1 = \mathrm{Vol}(D)$$

for every Borel subset $D \subset S$, so this is the correct volume form. Alternatively, we may easily verify that for every positive orthonormal basis $v_1, \ldots, v_{n-1}$ of T_pM we have

$$\omega(p)(n, v_1, \ldots, v_{n-1}) = \det(n, v_1, \ldots, v_{n-1}) = 1.$$

In either way, the proof is complete. □

Following the language of Section 7.2.4, the form ω corresponds to the unit normal vector field n. In particular, the Euclidean volume form on S^2 is the pull-back of

$$\omega = dy \wedge dz + dz \wedge dx + dx \wedge dy.$$

More generally, the n-form ω in $\mathbb{R}^{n+1} \setminus \{0\}$ given by

$$\omega = \frac{1}{\|x\|} \sum_{i=1}^{n+1} (-1)^{i-1} x^i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1}$$

pulls back simultaneously to the volume form on the sphere $S(0, r)$ centred in 0 and of radius $r > 0$, for every $r > 0$.

7.5.4. Scalar product on compactly supported k-forms. Let M be a manifold, possibly with boundary, equipped with a metric tensor g. As shown in Section 2.4.11, the scalar product $g(p)$ induces a rescaled scalar product $\langle , \rangle$ on $\Lambda^k(T_pM)$ at each $p \in M$. By letting p vary, we may couple any two k-forms $\alpha, \beta \in \Omega^k(M)$ to get a smooth function $\langle \alpha, \beta \rangle \in C^\infty(M)$.

Let $\Omega^k_c(M)$ be the space of compactly supported k-forms. We can define a bilinear form on $\Omega^k_c(M)$ by setting

$$(\alpha, \beta) = \int_M \langle \alpha, \beta \rangle \omega.$$

Here ω is the volume form induced by g. If $g(p)$ is positive definite for every $p \in M$, then $\langle , \rangle$ and $(,)$ are also both positive definite. In that case we can define the norm $\|\alpha\| = \sqrt{(\alpha, \alpha)}$ of a compactly supported k-form α.

7.5.5. The Hodge star operator. Let M be an oriented n-manifold possibly with boundary, equipped with a metric tensor g. We may identify k-forms and $(n-k)$-forms via the *Hodge star operator*, introduced in Section 2.5.4.

Indeed, if we apply it simultaneously to all points of M, the Hodge star operator becomes a linear map

$$*\colon \Omega^k(M) \longrightarrow \Omega^{n-k}(M).$$

The map is uniquely determined by requiring that

$$\alpha \wedge (*\beta) = \langle \alpha, \beta \rangle \omega$$

for all $\alpha \in \Omega^k(M)$. Here ω is the volume form induced by g.

Example 7.5.4. Let us consider $\mathbb{R}^n$ with its Euclidean metric tensor. It follows from Exercise 2.5.3 that

$$*(dx^1 \wedge \cdots \wedge dx^k) = dx^{k+1} \wedge \cdots \wedge dx^n.$$

More generally,

$$*(dx^{i_1} \wedge \cdots \wedge dx^{i_k}) = \pm dx^{i_{k+1}} \wedge \cdots \wedge dx^{i_n}$$

where the sign is that of the permutation $(i_1, \ldots, i_n)$.

If α, β are compactly supported k-forms, by integrating on M we get

$$\int_M \alpha \wedge *\beta = (\alpha, \beta).$$

7.5.6. (Anti-)self-dual differential forms. If the metric tensor g is positive definite, we deduce from Exercise 2.5.3 that $*\colon \Omega_c^k(M) \to \Omega_c^k(M)$ is an isometry and

$$**\beta = (-1)^{k(n-k)}\beta$$

for every $\beta \in \Omega^k(M)$. In particular, when $n = 2k$ we get an endomorphism

$$*\colon \Omega^k(M) \longrightarrow \Omega^k(M)$$

whose square is $*^2 = (-1)^k$. If k is even (so n is divisible by 4) we get $*^2 = 1$ and as explained in Section 2.5.4 we get a pointwise splitting into eigenspaces $\Lambda^k(T_pM) = \Lambda_+^k(T_pM) \oplus \Lambda_-^k(T_pM)$ and hence a global splitting of bundles

$$\Lambda^k(M) = \Lambda_+^k(M) \oplus \Lambda_-^k(M).$$

This gives a splitting of sections

$$\Omega^k(M) = \Omega_+^k(M) \oplus \Omega_-^k(M).$$

The k-forms in $\Omega_+^k(M)$ and in $\Omega_-^k(M)$ are called respectively *self-dual* and *anti-self-dual*.

7.5.7. Codifferential. Let M be an oriented n-manifold possibly with boundary, equipped with a metric tensor g. We define the *codifferential*

$$\delta\colon \Omega^k(M) \longrightarrow \Omega^{k-1}(M)$$

by setting

$$\delta = (-1)^k *^{-1} d * .$$

It is immediate to prove that $\delta(\delta\omega) = 0$ for any $\omega \in \Omega^k(M)$.

Exercise 7.5.5. Consider $\mathbb{R}^n$ with its Euclidean metric tensor. We have

$$\delta(f\, dx^1 \wedge \cdots \wedge dx^k) = (-1)^i \sum_{i=1}^{k} \frac{\partial f}{\partial x^i} dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k.$$

The following proposition says that when $\partial M = \varnothing$ the operator δ is the formal adjoint[1] of d with respect to the scalar product $(\,,\,)$.

Proposition 7.5.6. *Let M have empty boundary. For every $\alpha \in \Omega^k_c(M)$ and $\beta \in \Omega^{k+1}_c(M)$ we get*

$$(\alpha, \delta\beta) = (d\alpha, \beta).$$

Proof. We note that $\alpha \wedge *\beta$ is a $(n-1)$-form and Stokes gives

$$\begin{aligned} 0 = \int_M d(\alpha \wedge *\beta) &= \int_M (d\alpha \wedge *\beta) + \int_M (-1)^k \alpha \wedge d(*\beta) \\ &= (d\alpha, \beta) + \int_M (-1)^k \alpha \wedge (-1)^{k+1} * \delta\beta = (d\alpha, \beta) - (\alpha, \delta\beta). \end{aligned}$$

The proof is complete. □

Note that we do not require g to be positive-definite. After checking all signs very carefully, we may also write

$$\delta = (-1)^{kn+n+m+1} * d*$$

where (p, m) is the signature of g.

7.5.8. Laplacian. By combining differentials and codifferentials we can define the *Laplacian* of k-forms:

$$\Delta\colon \Omega^k(M) \longrightarrow \Omega^k(M)$$

by setting

$$\Delta = (\delta + d)^2 = \delta d + d\delta.$$

In the second equality we used that $d^2 = 0$ and $\delta^2 = 0$.

[1] The term *formal adjoint* is employed for operators that behave formally like adjoints on spaces that are not necessarily Hilbert spaces.

Exercise 7.5.7. On $\mathbb{R}^n$ equipped with the Euclidean metric tensor, the Laplacian of a function (that is, of a 0-form) is the usual one (with a sign):

$$\Delta f = -\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}.$$

Exercise 7.5.8. The following equalities hold:

$$*\delta = (-1)^k d*, \quad \delta * = (-1)^{k+1} * d, \quad *d\delta = \delta d*, \quad *\delta d = d\delta *, \quad *\Delta = \Delta * .$$

If $\partial M = \varnothing$, the Laplacian is formally self-adjoint:

Exercise 7.5.9. Let M have no boundary. For every $\alpha, \beta \in \Omega_c^k(M)$ we get

(1) $(\Delta\alpha, \beta) = (\delta\alpha, \delta\beta) + (d\alpha, d\beta) = (\alpha, \Delta\beta)$,
(2) $(\Delta\alpha, \alpha) = \|\delta\alpha\|^2 + \|d\alpha\|^2 \geq 0$ if g is positive definite,

7.5.9. Harmonic forms. A k-form $\alpha \in \Omega^k(M)$ is *harmonic* if $\Delta\alpha = 0$.

Proposition 7.5.10. *Let M have no boundary and be equipped with a positive definite metric tensor g. A compactly supported k-form $\alpha \in \Omega_c^k(M)$ is harmonic $\Longleftrightarrow$ $d\alpha = 0$ and $\delta\alpha = 0$.*

Proof. If $d\alpha = 0$ and $\delta\alpha = 0$ then of course $\Delta\alpha = 0$. Conversely, if $\Delta\alpha = 0$ then Exercise 7.5.9-(2) gives $d\alpha = 0$ and $\delta\alpha = 0$. □

Let $\mathcal{H}^k(M) \subset \Omega^k(M)$ denote the vector subspace consisting of all harmonic k-forms. Since $*\Delta = \Delta *$, we deduce that

$$*\colon \mathcal{H}^k(M) \longrightarrow \mathcal{H}^{n-k}(M)$$

is an isomorphism. If M is compact and g is positive definite, the spaces $\mathcal{H}^k(M)$ are equipped with the positive-definite scalar product $(,)$ and $*$ is also an isometry.

Proposition 7.5.11. *Let M be connected. Then*

- *$\mathcal{H}^0(M) \cong \mathbb{R}$ consists of the constant functions.*
- *$\mathcal{H}^n(M) \cong \mathbb{R}$ consists of the n-forms $\lambda\omega$ with $\lambda \in \mathbb{R}$, where ω is the volume form induced by g.*

Proof. A function f on M is harmonic $\Longleftrightarrow$ $df = 0$ $\Longleftrightarrow$ f is locally constant $\Longleftrightarrow$ f is constant (since M is connected). The second assertion follows since $*\colon \mathcal{H}^0(M) \to \mathcal{H}^n(M)$ is an isomorphism that sends the constant function 1 to the volume form ω, see Exercise 2.5.3. □

7.6. Special relativity and electromagnetism

We now use all the mathematical background exposed in the previous pages to introduce two major physical theories, that is Einstein's special relativity and Maxwell's equations of electromagnetism. Both theories have a strong geometric nature and can be described concisely and elegantly using metric tensors and differential forms. We start with the former.

7.6.1. Minkowski space. In special relativity, the spacetime is modeled as the *Minkowski space*. This is simply $\mathbb{R}^4$ with coordinates $t = x^0, x^1, x^2, x^3$, equipped with a specific metric tensor η. Since the tangent plane at every point $x \in \mathbb{R}^4$ is identified with $\mathbb{R}^4$ itself, a metric tensor is specified by a 4×4 invertible symmetric matrix that depends smoothly on $x \in \mathbb{R}^4$. The tensor field η used here is just constantly the matrix

$$\eta = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The positive real number c is the speed of light. From now on, to make life easier we choose some appropriate units such that $c = 1$. The first thing to note is that η is a non positive definite scalar product, having signature $(3, 1)$. The Minkowski space is sometimes denoted as $\mathbb{R}^{3,1}$. We interpret η both as a matrix and as a scalar product, so for every $v, w \in \mathbb{R}^{3,1}$ we write

$$\eta(v, w) = \eta_{ij} v^i w^j.$$

The tangent space at every point $x \in \mathbb{R}^{3,1}$ has a rich structure, that is of fundamental importance in special relativity and in the way we understand our universe. A vector $v \neq 0$ in the tangent space is *timelike*, *lightlike*, or *spacelike* according to whether $\eta(v, v)$ is negative, null, or positive. See Figure 7.1-(left). Timelike vectors v are partitioned into two open cones, depending on the sign of their time component v^0, called *future* and *past*. Timelike (spacelike) vectors v with $\eta(v, v) = -1$ (respectively, $\eta(v, v) = 1$) are called *unit timelike (spacelike) vectors* and form a hyperboloid with two (one) sheets: see Figure 7.2.

A point in $\mathbb{R}^{3,1}$ is called an *event*. A *world path* is any curve in $\mathbb{R}^{3,1}$ whose tangents are all future directed timelike vectors, as in Figure 7.1-(right). In special relativity, nothing can travel faster than light: massless particles (like photons) travel straight with constant speed c, while the velocity of every massive particle is always strictly smaller than c. Therefore photons travel along straight lines with lightlike slope, and massive particles travel along world paths.

Let γ be a world path. Up to reparametrising we may always suppose that the derivative $\gamma'(t)$ is a unit vector for all t, and this will be always assumed tacitly in the following.

A crucial aspect of Minkowski space is that it comes naturally equipped with a group of symmetries called *Lorentz tranformations*, that mix space and time in a counterintuitive way.

7.6.2. Lorentz transformations. A *Lorentz transformation* is a linear isomorphism $f(x) = Ax$ of $\mathbb{R}^4$ that preserves the bilinear form η, that is such

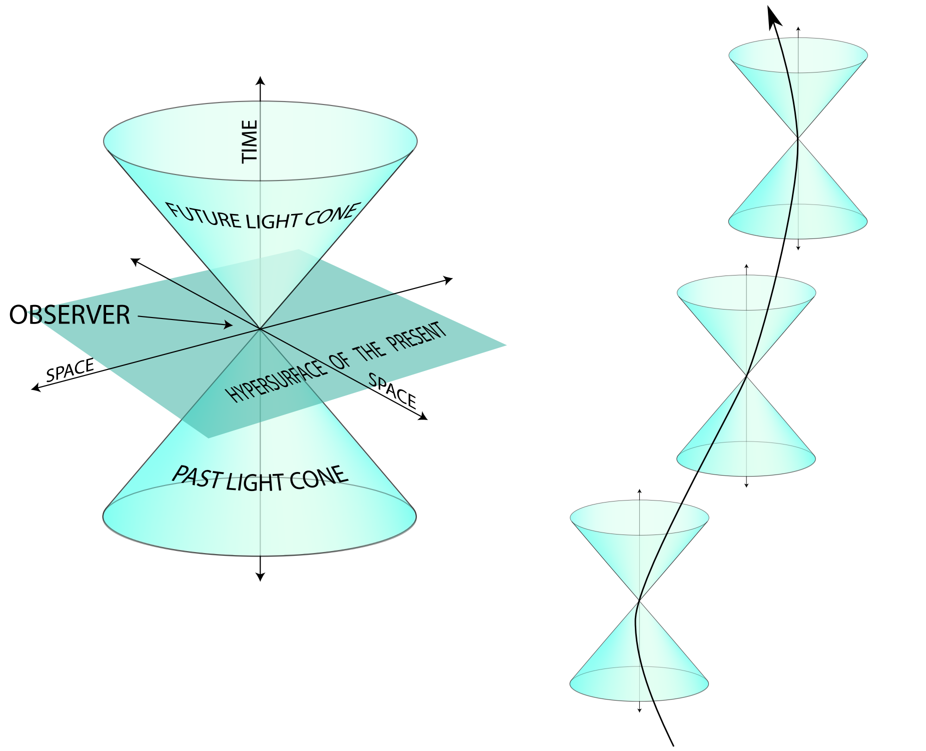

Figure 7.1. The tangent space of every point $x \in \mathbb{R}^{3,1}$ contains points of three types: *timelike*, *lightlike*, and *spacelike*. The timelike points are divided into two components, *future* and *past*. The picture displays the tangent space with one spacial dimension omitted (left). A world line is a curve with future-directed timelike tangent vectors (right)

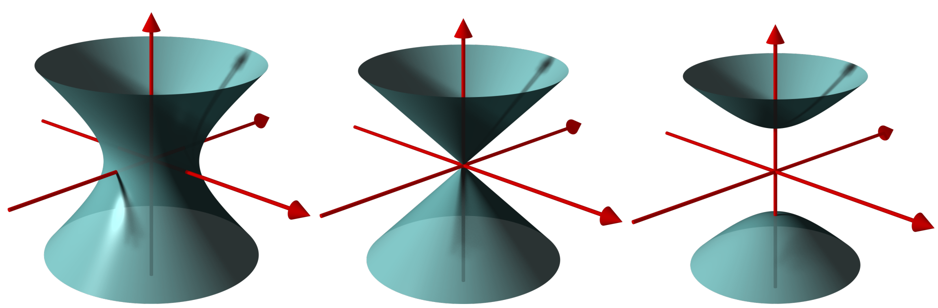

Figure 7.2. The spacelike vectors v with $\eta(v,v) = 1$ form a hyperboloid with one sheet, the lightlike vectors form a cone (called the *light cone*), and the timelike vectors v with $\eta(v,v) = -1$ form a hyperboloid with two sheets (future and past).

that ${}^{\mathrm{t}}A\eta A = \eta$ as matrices. In coordinates we write this as

$$A^i_j \eta_{ik} A^k_l = \eta_{jl}.$$

The group of all Lorentz transformations is denoted by $\mathrm{O}(3,1)$.

A *Lorentz basis* is a basis v_0, v_1, v_2, v_3 of vectors such that $\eta(v_i, v_j) = \eta_{ij}$. The canonical basis e_0, e_1, e_2, e_3 is an example. A matrix A defines a Lorentz transformation $\Longleftrightarrow$ its columns form a Lorentz basis.

Every orthogonal matrix $B \in \mathrm{O}(3)$ gives rise to a Lorentz transformation

$$A = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix}. \tag{16}$$

These matrices represent the usual isometries of three-dimensional space and have no effect on time. For instance one finds the usual rotation of angle θ around a coordinate axis

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

A somehow similar kind of Lorentz transformation is the *Lorentz boost*

$$A = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ \sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This is the simplest kind of Lorentz transformation that mixes space and time. As opposite to rotations, different values of $\zeta \in \mathbb{R}$ yield distinct transformations (no periodicity!). The following exercise can be proved much in the same way as we did in Proposition 3.9.2 for $\mathrm{O}(n)$.

Exercise 7.6.1. The group $\mathrm{O}(3,1)$ is a 6-dimensional submanifold of $M(4)$, hence a Lie group.

Note that $\mathrm{O}(3,1)$ has the same dimension as $\mathrm{O}(4)$. This means that, despite Minkowski space may look less natural than the familiar Euclidean space $\mathbb{R}^4$, it has roughly the same amount of symmetries.

As opposite to $\mathrm{O}(n)$, one sees by looking at Lorentz boosts that $\mathrm{O}(3,1)$ is not compact. Like $\mathrm{O}(n)$, the group $\mathrm{O}(3,1)$ is not connected, and we now check that it has as much as four components (whereas $\mathrm{O}(n)$ has only two).

Since $\eta = {}^{\mathrm{t}}A\eta A$, every matrix $A \in \mathrm{O}(3,1)$ must have $\det A = \pm 1$, and we get a homomorphism $\det\colon \mathrm{O}(3,1) \to \{\pm 1\}$. The kernel is denoted as $\mathrm{SO}(3,1)$. An additional homeomorphism onto the cyclic group of order two is constructed by sending $A \in \mathrm{O}(3,1)$ to the sign of the top-left element A^0_0. The matrix A sends the timelike vector e_0 to a timelike vector that is either future or past directed, depending on the sign of A_0^0. The kernel of this homomorphism is denoted as $\mathrm{O}^+(3,1)$. We also write

$$\mathrm{SO}^+(3,1) = \mathrm{SO}(3,1) \cap \mathrm{O}^+(3,1).$$

The subgroup $\mathrm{SO}^+(3,1)$ consists of all Lorentz transformations that preserve the orientations of both $\mathbb{R}^{3,1}$ and time.

Proposition 7.6.2. *The manifold* $\mathrm{O}(3,1)$ *has four connected components: the normal subgroup* $\mathrm{O}^+(3,1)$*, and its cosets.*

Proof. We prove that $\mathrm{O}^+(3,1)$ is path-connected. This is equivalent to show that a positive Lorentz basis v_0, v_1, v_2, v_3 with future directed v_0 may be continuously deformed through Lorentz basis to the canonical e_0, e_1, e_2, e_3. With a composition of boosts along different axis we may first send continuously v_0 to e_0 (exercise), and then the remaining three spacelike vectors can be moved to e_1, e_2, e_3 continuously (keeping e_0 fixed) since $\mathrm{O}(3)$ is connected.

Points in different cosets cannot be path-connected because the two homomorphisms $\mathrm{O}(3,1) \to \{\pm 1\}$ constructed above are continuous. □

During the proof we have also shown (actually, left as an exercise to prove) that the Lorentz group acts transitively on future-directed time-like vectors v normalized so that $\eta(v,v) = -1$. These form the upper sheet of the hyperboloid shown in Figure 7.2-(right). The stabilizer of one such vector is isomorphic to $\mathrm{O}(3)$. Indeed, we may suppose that this vector is e_0, and the Lorentz transformations that fix e_0 are clearly those of the form (16).

The *Poincaré group* is the group of all affine transformations $f(x) = Ax+b$ of the Minkowski space $\mathbb{R}^{3,1}$ with $A \in \mathrm{O}(3,1)$. These are precisely the affine transformations f that preserve the tensor field η, that is such that $f^*(\eta) = \eta$. The Poincaré group is the natural automorphisms group of $\mathbb{R}^{3,1}$.

7.6.3. Lorentz frame. An important feature of Minkowski space is that its identification with $\mathbb{R}^{3,1}$ is actually not absolute, but it strongly depends on the point of view of the observer. Suppose that you happily travel in Minkowski space along some world path $\gamma\colon (-\varepsilon,\varepsilon) \to \mathbb{R}^{3,1}$. Your tangent vector $\gamma'(t)$ is unit timelike and future directed for all t.

You may complete $\gamma'(0)$ to a Lorentz basis $v_0 = \gamma'(0), v_1, v_2, v_3$. Note that v_0 is determined by your world path, while the spacelike orthonormal basis v_1, v_2, v_3 is not unique: you choose it arbitrarily by indicating three orthogonal directions in space with your arms (and feet).

Having settled a Lorentz basis, you may use it as a new frame for Minkowski space, where you put (quite egoistically) yourself at the center of the universe and v_0, v_1, v_2, v_3 as the new axis. The resulting frame is called a *Lorentz frame* for an observer (you) moving along γ at time $t = 0$.

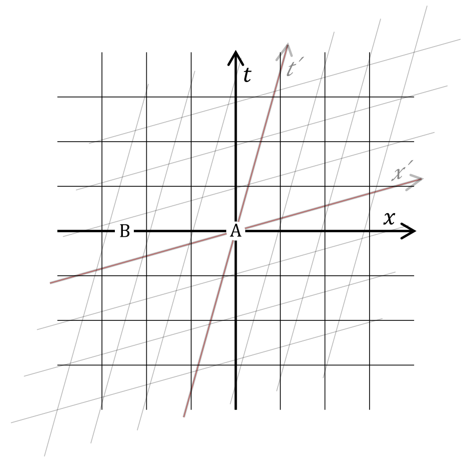

Figure 7.3. Two observers that meet at A with different speed model the universe with different Lorentz frames (x, t) and (x', t'). Foliations do not match: the event B is in the present for one observer, and lies in the future for the other.

7.6.4. Simultaneity is not an absolute notion! There is no way of choosing an absolute frame in the Minkowski universe. Each observer has her own natural Lorentz frame at every instant of her life, whose time axis is tangent to her world path.

An immediate consequence of this viewpoint is the lack of any notion of absolute time, and more dramatically of any notion of *simultaneity* of events. It may look natural to foliate $\mathbb{R}^{3,1}$ by the 3-dimensional sheets $x^0 = k$, and to say that two events are *simultaneous* if they belong to the same sheet. Unfortunately, this foliation is not invariant under Lorentz transformations, because it is not invariant under Lorentz boosts.

An observer traveling on a world path γ may define her foliation by taking all affine 3-spaces that are orthogonal to $\gamma'(0)$ with respect to η. Any observer has thus a well-defined notion of simultaneity for the events occurring in the whole Minkowski universe. However, two observers traveling on distinct world lines with different tangent vectors $\gamma_1'(0) \neq \gamma_2'(0)$ will obtain different foliations, and therefore different notions of simultaneity: see an example in Figure 7.3. A fully egoistic perspective is also not easy to handle, because the foliations that you obtain at different times t_1 and t_2 of your world path γ may differ if $\gamma'(t_1) \neq \gamma'(t_2)$. An event that was "occurring in the past" with respect to your natural frame yesterday may have now jumped to the future after that you accelerated your spaceship this morning.

7.6.5. Spacetime interval, chronology, and causality are absolute notions. The old absolute notions of past and future are not completely destroyed: in Minkowski space we still have the absolute notions of *causality*, *chronology*, and of *spacetime interval* between events.

Given two events A and B in the Minkowski space $\mathbb{R}^{3,1}$, we consider the vector $\overrightarrow{AB} = B - A$ and define the *spacetime interval* $\eta(\overrightarrow{AB}, \overrightarrow{AB})$ between A and B. Since η is preserved by any transformation of the Poincaré group, the spacetime distance between two events is a number that actually does not depend on the particular Lorentz frame chosen to calculate it. So it is an intrinsic invariant of Minkowski space. Note also that the spacetime interval is symmetric – it does not change if we reverse the roles of A and B. The spacetime interval is a positive/null/negative real number $\Longleftrightarrow$ the vector $\overrightarrow{AB}$ is spacelike/lightlike/timelike.

Being spacelike/lightlike/timelike is a well-defined notion for $\overrightarrow{AB}$ that is independent of the chosen Lorentz frame. This allows us to define two partial orderings between events, the *chronological* and the *causal* orderings, both of physical relevance. Let A and B be two events. In the chronological order, we write $A < B$ if and only if $\overrightarrow{AB}$ is a future-directed timelike vector, while in the causal order we write $A < B$ if and only if $\overrightarrow{AB}$ is a future-directed timelike or lightlike vector. In both the chronological and causal settings we really get a partial ordering (exercise).

In the chronological ordering, one sees easily that $A < B \Longleftrightarrow$ there is a world path from A to B. From a physical point of view, this means that it is (at least in principle) possibile for a massive body to travel from A to B. In the causal ordering, we get $A < B \Longleftrightarrow$ there is either a world path or a light line from A to B. This means that it is possible (at least in principle) that the event A has some consequences on the event B, because some particle (with or without mass) might have gone from A to B.

7.6.6. Proper time. Consider a massive body that travels in Minkowski space along some world line γ. We define the *proper time* of the body as the integral of $\sqrt{-\eta(\gamma'(t), \gamma'(t))}$ along the path. This is a Lorentz transformation independent notion and is therefore intrinsic: it measures how time passes as perceived by the massive body.

As already noted, up to reparametrising we can (and usually do) always assume that $\gamma'(t)$ is a unit vector for all t. In this setting, the world-line $\gamma\colon [a, b] \to \mathbb{R}^{3,1}$ is parametrised by proper time: the body perceives that $b - a$ units of time have elapsed between the events $A = \gamma(a)$ and $B = \gamma(b)$. Of course two bodies that meet at A and B passing through different world lines may have perceived different time intervals. It is natural now to ask what is the quickest path between A and B, and the answer should not be surprising.

Exercise 7.6.3. Let A, B be two events such that $\overrightarrow{AB}$ is a future timelike vector. The world path with shortest time length from A to B is the segment.

Every other path from A to B has time length bigger than $\sqrt{-\eta(\overrightarrow{AB}, \overrightarrow{AB})}$.

7.6.7. Four-velocity and four-momentum. Let us consider again a massive body traveling along a world path γ, that we suppose parametrised by proper time. At every time t, the body has a *four-velocity* $\gamma'(t)$. This is a unit time-like tangent vector at $\gamma(t)$.

Mass and energy enter into this picture in the simplest way. Every massive body has a *rest mass* $m > 0$ that is constant along its journey and intrinsic (that is, frame independent). At each time of its world path we define the *four-momentum* of the body as $P(t) = m\gamma'(t)$. Note that $m = \sqrt{-\eta(P(t), P(t))}$.

The four-momentum is of course a tangent vector. Its coordinates may be denoted as $P(t) = (E, p_x, p_y, p_z)$. The quantity E is called the *energy*, and the vector (p_x, p_y, p_z) is the *momentum*. While the four-momentum is an intrinsic object, its components "energy" and "momentum" are not: they strongly depend on the chosen Lorentz frame. If the body is at rest in the frame, we get $P(t) = (E, 0, 0, 0)$ and hence $m = E$. This is the famous Einstein equivalence $E = mc^2$ expressed with $c = 1$. In general, we have

$$m^2 = -\eta(P(t), P(t)) = E^2 - (p_x)^2 - (p_y)^2 - (p_z)^2.$$

If we write $p = \sqrt{p_x^2 + p_y^2 + p_z^2}$ we find

$$E = \sqrt{m^2 + p^2}.$$

Analogously, we write the four-velocity $\gamma'(t) = (v^0, v_x, v_y, v_z)$, the *velocity* component is (v_x, v_y, v_z), and its norm $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$, and we get

$$E = \sqrt{m^2 + m^2v^2} = m\sqrt{1 + v^2} = m + \frac{1}{2}mv^2 + \cdots$$

If the body travels at a velocity v much smaller than $c = 1$, its energy is the rest mass + the kinetic energy + small order terms. The kinetic energy has appeared quite unexpectedly out of the blue!

As we said, the energy E is not an intrinsic quantity. A massive body with four-momentum P, examined by an observer traveling with four-velocity v, has energy $E = -\eta(P, v)$.

7.6.8. The electromagnetic field tensor. Special relativity has been introduced by Einstein to resolve an incompatibility between Maxwell's equations of electrodynamics and the Netwon mechanics. It should then not surprise the reader that Maxwell's equations fit naturally and elegantly within the geometric frame of Minkowski space.

We are used to interpret an electric field E and a magnetic field B as vector fields in $\mathbb{R}^3$. We now see that both E and B may actually be seen as components of an antisymmetric tensor field F of type $(0, 2)$, that is a 2-form.

The *electromagnetic tensor field* is a 2-form F on the Minkowski space $\mathbb{R}^{3,1}$. A 2-form in $\mathbb{R}^{3,1}$ is simply a 4×4 antisymmetric matrix that depends smoothly on the point. The components of F may be written as

$$F = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}.$$

In other words,

$$\begin{aligned} F = &-E_1 dt \wedge dx^1 - E_2 dt \wedge dx^2 - E_3 dt \wedge dx^3 \\ &+ B_3 dx^1 \wedge dx^2 + B_2 dx^3 \wedge dx^1 + B_1 dx^2 \wedge dx^3. \end{aligned}$$

Here $E = (E_1, E_2, E_3)$ and $B = (B_1, B_2, B_3)$ are the usual electric and magnetic fields. The crucial fact is the following: the tensor field F is intrinsic, while E and B strongly depend on the chosen Lorentz frame.

7.6.9. The Lorentz force law. We write the Lorentz force law, that evaluates the acceleration of a particle with mass m and charge q crossing the field with four-velocity v. We use η to transform the $(0, 2)$ tensor field F into a $(1, 1)$ tensor field, that we still denote by F. In coordinates, we are raising an index as $F^i{}_j = F_{kj}\eta^{ik}$. Recall that η^{ik} is the inverse matrix of η_{ik}, so they are the same matrix, and in coordinates the $(1, 1)$ tensor field F is

$$F = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}.$$

Recall that a $(1, 1)$ tensor is an endomorphism. Let a particle with mass m and charge q move along its world line $\gamma(t)$ parametrised by proper time t, and let $v = \gamma'(t)$ be its four-velocity. The *Lorentz force law* states that

$$\frac{dv}{dt} = \frac{q}{m} F(v).$$

In coordinates we get

$$\frac{dv^i}{dt} = \frac{q}{m} F^i{}_j v^j.$$

One checks immediately that this equality is equivalent to the familiar non-relativistic Lorentz force law:

$$\begin{cases} \dfrac{dE}{dt} = qE \cdot u, \\ \dfrac{du}{dt} = \dfrac{q}{m}(E + u \times B) \end{cases}$$

where $u = (v^1, v^2, v^3)$ is the spacial velocity and E is the energy of the particle.

7.6.10. Maxwell's equations. We are ready to write Maxwell's equations. We calculate the exterior derivative

$$\begin{aligned} dF = &\left(\frac{\partial B_3}{\partial t} + \frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2}\right) dt \wedge dx^1 \wedge dx^2 \\ &+ \left(\frac{\partial B_2}{\partial t} + \frac{\partial E_1}{\partial x^3} - \frac{\partial E_3}{\partial x^1}\right) dt \wedge dx^3 \wedge dx^1 \\ &+ \left(\frac{\partial B_1}{\partial t} + \frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3}\right) dt \wedge dx^2 \wedge dx^3 \\ &+ \left(\frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3}\right) dx^1 \wedge dx^2 \wedge dx^3 \end{aligned}$$

and deduce that the two Maxwell equations

$$\begin{cases} \mathrm{rot} E = -\dfrac{\partial B}{\partial t}, \\ \mathrm{div} B = 0 \end{cases}$$

are equivalent to the very concise single equation

$$dF = 0.$$

The remaining two Maxwell equations also reduce to a single equality involving differential forms, but in order to write them we need to formalise charges and currents. It should not be much of a surprise at this point that charges and currents are unified into a single object called *four-current density*, a vector field J on Minkowski space. Its components are $J = (\rho, J^1, J^2, J^3)$. The time component ρ is the *charge density* and $j = (J^1, J^2, J^3)$ is the *current density*. As for the four-momentum, the four-current density is intrinsic, while its components "charge" and "current" density depend on the Lorentz frame. The word "density" is sometimes omitted.

The equation $dF = 0$ is in fact unrelated to the metric tensor η. On the contrary, the next equation that we will write depends on η. The link between differential forms and metric tensors is furnished by the Hodge $*$ operator.

We can write everything explicitly for $\mathbb{R}^{3,1}$. The canonical orientation of $\mathbb{R}^{3,1}$ together with η induce the volume form

$$\omega = dt \wedge dx^1 \wedge dx^2 \wedge dx^3.$$

The Hodge star operator transforms a k-form into a $(4-k)$-form. In particular we deduce from Exercise 2.5.3 the following equalities:

$$\begin{aligned}
*1 &= dt \wedge dx^1 \wedge dx^2 \wedge dx^3, & *(dt \wedge dx^1 \wedge dx^2 \wedge dx^3) &= -1,\\
*dt &= -dx^1 \wedge dx^2 \wedge dx^3, & *dx^1 &= -dt \wedge dx^2 \wedge dx^3,\\
*dx^2 &= -dt \wedge dx^3 \wedge dx^1, & *dx^3 &= -dt \wedge dx^1 \wedge dx^2,\\
*(dt \wedge dx^1) &= -dx^2 \wedge dx^3, & *(dt \wedge dx^2) &= -dx^3 \wedge dx^1,\\
*(dt \wedge dx^3) &= -dx^1 \wedge dx^2, & *(dx^1 \wedge dx^2) &= dt \wedge dx^3,\\
*(dx^2 \wedge dx^3) &= dt \wedge dx^1, & *(dx^3 \wedge dx^1) &= dt \wedge dx^2.
\end{aligned}$$

We have $*^2 = 1$ on 1- and 3-forms and $*^2 = -1$ on 0-, 2-, and 4-forms. By applying these equalities to the electromagnetic field tensor F we get

$$\begin{aligned}
*F = {}& E_1 dx^2 \wedge dx^3 + E_2 dx^3 \wedge dx^1 + E_3 dx^1 \wedge dx^2\\
&+ B_1 dt \wedge dx^1 + B_2 dt \wedge dx^2 + B_3 dt \wedge dx^3.
\end{aligned}$$

The components of the 2-form $*F$ may be written as

$$*F = \begin{pmatrix} 0 & B_1 & B_2 & B_3 \\ -B_1 & 0 & E_3 & -E_2 \\ -B_2 & -E_3 & 0 & E_1 \\ -B_3 & E_2 & -E_1 & 0 \end{pmatrix}.$$

We now turn to the four-current J. The four-current is a vector field, and we can transform it into a 1-form (still denoted by J) by contracting it with the metric tensor η. In coordinates $J_i = J^j \eta_{ij}$ and hence

$$J = -\rho dt + J^1 dx^1 + J^2 dx^2 + J^3 dx^3.$$

By applying the Hodge star operator we find

$$\begin{aligned}
*J = {}& \rho dx^1 \wedge dx^2 \wedge dx^3 - J^1 dt \wedge dx^2 \wedge dx^3\\
&- J^2 dt \wedge dx^3 \wedge dx^1 - J^3 dt \wedge dx^1 \wedge dx^2.
\end{aligned}$$

It is now immediate to prove that the last two Maxwell equations

$$\begin{cases} \operatorname{rot} B = j + \dfrac{\partial E}{\partial t}, \\ \operatorname{div} E = \rho \end{cases}$$

are equivalent to the following:

$$d(*F) = *J.$$

7.6.11. Comments. We have described a magnetic field as a 2-form F in Minkowski space. Maxwell's equations are

$$\begin{cases} dF = 0, \\ d(*F) = *J. \end{cases} \tag{17}$$

We now make some comments on this construction.

Covariant form. The most important property of Maxwell's equations (17) is probably that they are *expressed in covariant form*. This means that they represent some relations between the tensor fields η, F, J that are true independently on the particularly chosen coordinates chart.

Invariants from F. We can extrapolate from the tensor field F some other tensor fields. For instance, the may consider the rescaled scalar product of forms from Section 2.4.11 and get

$$\langle F, F\rangle = \frac{1}{2} F_{ij}\eta^{ik}\eta^{jl}F_{kl} = B^2 - E^2$$

where $B^2 = B_1^2 + B_2^2 + B_3^2$ and $E^2 = E_1^2 + E_2^2 + E_3^2$. The squared norms B^2 and E^2 of the magnetic and electric field are not separately invariant, but the difference $B^2 - E^2$ is. Using the Hodge star we also find

$$\langle F, *F\rangle = \frac{1}{2} F_{ij}\eta^{ik}\eta^{jl}(*F)_{kl} = 2\, E\cdot B$$

and hence also the scalar product $E\cdot B = E_1B_1 + E_2B_2 + E_3B_3$ is invariant. In particular, if E and B are orthogonal in some Lorentz frame, they are so in any Lorentz frame.

Continuity equation. Since $*J = d(*F)$, we deduce that

$$d(*J) = d(d(*F)) = 0.$$

This is the *continuity equation*

$$\frac{\partial \rho}{\partial t} + \operatorname{div} j = 0.$$

Codifferential. Using the codifferential δ defined in Section 7.5.7, Maxwell's equations (17) can be written as follows:

$$\begin{cases} dF = 0, \\ \delta F = J. \end{cases}$$

Stokes. Stokes' Theorem implies that for any domain $D \subset \mathbb{R}^{3,1}$ we have

$$\int_{\partial D} *J = \int_D d(*J) = 0.$$

This very general fact furnishes different kinds of physical information depending on the shape of the four-dimensional domain D. For instance, if $D = D' \times [t_0, t_1]$ for some domain $D' \subset \mathbb{R}^3$ the equation says that the difference between the total charge of D' at the times t_1 and t_0 is equal to the flow of current along $\partial D'$.[2]

Potential. We will see in the next chapter that $dF = 0$ on $\mathbb{R}^n$ implies that $F = dA$ for some 1-form A called *potential*. We write it as

$$A = -V\,dt + A_1 dx^1 + A_2 dx^2 + A_3 dx^3.$$

The potential has the disadvantage of not being unique, and the advantage of containing only 4 parameters instead of the 6 parameters that define F. If we write $a = (A_1, A_2, A_3)$ as a vector, we see that in coordinates V and a are the usual potentials for the electric and magnetic fields, that is

$$E = -\nabla V - \frac{\partial a}{\partial t}, \qquad B = \text{rot}a.$$

Concerning Maxwell's equations, the first $dF = ddA = 0$ is now automatic since $d^2 = 0$. The second one $\delta F = J$ becomes $\delta dA = J$.

Recall that A is not unique: we can modify A to $A' = A + df$ for any function f and get another potential A' for F. If we find a f that satisfies

$$\Delta f = -\delta A \tag{18}$$

then we easily get $\delta A' = 0$. Here $\Delta = d\delta + \delta d$, see Section 7.5.8. With this potential the second Maxwell equation $\delta dA' = J$ can be written using the Laplacian as

$$\Delta A' = J.$$

Exercise 7.6.4. Let f be a function on $\mathbb{R}^{3,1}$. The Laplacian of f is $\Delta f = -\Box f$ where $\Box$ is the *d'Alambertian*

$$\Box f = -\frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_3^2}.$$

The equation (18) is thus of type $\Box f = \delta A$. This PDE is an inhomogeneous wave equation and solutions are known to exist in many cases. Fonte?

[2]The subset D is not strictly a domain because its boundary is not smooth, however corners can be smoothened and Stokes' Theorem still applies.

7.7. Exercises

Exercise 7.7.1. Let M be a n-manifold. Show that $\omega \in \Omega^n(M)$ is a volume form with respect to some orientation of M if and only if $\omega(p)$ is nowhere vanishing. In particular M is orientable $\Longleftrightarrow$ there is a nowhere vanishing n-form.

Exercise 7.7.2. Let M be a n-manifold without boundary. Let $\omega \in \Omega^1(M)$ be nowhere vanishing and with $d\omega = 0$. Since $\omega(p) \neq 0$ for all $p \in M$, the functional $\omega(p)\colon T_pM \to \mathbb{R}$ is non-trivial and hence we can define a distribution of hyperplanes:

$$D(p) = \ker \omega(p).$$

Show that D is integrable and so gives rise to a foliation on M, uniquely determined by ω.

Exercise 7.7.3. Let D be a 2-dimensional distribution on a 3-manifold M. Show that D is integrable $\Longleftrightarrow$ for every nowhere-vanishing 1-form α defined on some open set with $\ker \alpha = D$ we have $\alpha \wedge d\alpha = 0$.

Exercise 7.7.4. Consider the n-torus $M = S^1 \times \cdots \times S^1$ with its coordinates $(\theta^1, \ldots, \theta^n)$. Each tangent space is canonically identified with $\mathbb{R}^n$ and we assign it the Euclidean metric tensor $g_{ij} = \delta_{ij}$. Show that each harmonic k-form on the n-torus is a linear combination of the k-forms

$$dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

and hence

$$\dim \mathcal{H}^k(M) = \binom{n}{k}.$$

CHAPTER 8

De Rham cohomology

We now exploit the relation $d(d\omega) = 0$ on differential forms to build an algebraic construction called *De Rham cohomology*. This algebraic construction has some similarities with the fundamental group: it assigns groups to manifolds, and it is functorial, that is smooth maps induce groups homomorphisms. It can be used in particular to distinguish manifolds.

Cohomology is however different from fundamental groups, and may be used to accomplish some tasks that the fundamental group is unable to carry out. For instance, we will use it to prove that the smooth manifolds

$$S^4, \qquad S^2 \times S^2, \qquad \mathbb{CP}^2$$

are pairwise non-homeomorphic, and not even homotopy equivalent, although they are all simply-connected compact four-manifolds.

8.1. Definition

8.1.1. Closed and exact forms. Let M be a smooth n-manifold, possibly with boundary.

Definition 8.1.1. A k-form ω on M is *closed* if $d\omega = 0$, and is *exact* if there is a $(k-1)$-form η such that $\omega = d\eta$.

Since $d(d\eta) = 0$, every exact form is also closed, but the converse does not always hold, and this is the key point that motivates everything that we are going to say in this chapter. We now list some motivating examples.

Example 8.1.2. Every n-form ω in M is closed, since $d\omega$ is a $(n+1)$-form, and every $(n+1)$-form is trivial on M. On the other hand, if M is compact, oriented, and without boundary, and ω is a volume form, then ω is not exact: if $\omega = d\eta$ by Stokes' Theorem we would get

$$\int_M \omega = \int_M d\eta = 0$$

but the integral of a volume form is always strictly positive, a contradiction.

More generally, the following holds.

Proposition 8.1.3. *If $\omega \in \Omega^k(M)$ is exact, for every compact oriented k-submanifold $S \subset M$ without boundary we have*

$$\int_S \omega = 0.$$

Proof. If $\omega = d\eta$ then Stokes applies. □

Example 8.1.4. On the torus $T = S^1 \times S^1$ with coordinates θ^1, θ^2, the 1-form $\omega = d\theta^1$ of Exercise 7.2.8 is closed but is not exact: indeed note that θ^1 is only a locally defined function (whose value has a $+2k\pi$ indeterminacy), so ω is locally exact, which suffices for getting closedness $d(d\theta^1) = 0$ but not for global exactness, since the integral of ω over the curve γ_2 does not vanish.

Example 8.1.5. Pick $U = \mathbb{R}^2 \setminus \{0\}$. Using polar coordinates ρ, θ we may define the closed non-exact form $\omega = d\theta$ on U, like in the previous example. In Euclidean coordinates the form is

$$\omega = \frac{xdy - ydx}{x^2 + y^2}$$

and the skeptic reader may check that $d\omega = 0$ via direct calculation. As above, the 1-form is not exact because its integral above the curve $S^1 \subset U$ is $2\pi \neq 0$.

In the last example, it is tempting to think that ω is not exact because there is a "hole" in U where the origin has been removed (note that ω does not extend to the origin). We will confirm this intuition in the next pages: closed non-exact forms detect some kinds of topological holes in the manifold M, and this precious information is efficiently organised into the more algebraic *De Rham cohomology*.

8.1.2. De Rham cohomology. Let M be a smooth manifold, possibly with boundary. We define

$$Z^k(M), \qquad B^k(M)$$

respectively as the vector subspaces of $\Omega^k(M)$ consisting of all the closed and all the exact k-forms.

As we said, we have the inclusion $B^k(M) \subset Z^k(M)$ and hence we may define the *De Rham cohomology group* as the quotient

$$H^k(M) = Z^k(M)/B^k(M).$$

This is actually a vector space, but the term "group" is usually employed in analogy with some more general constructions where all these spaces are modules over some ring.

8.1.3. The Betti numbers. The *k-th Betti number* of M is the dimension

$$b^k(M) = \dim H^k(M).$$

Of course this number may be infinite, but we will see that it is finite in the most interesting cases. This is a remarkable and maybe unexpected fact, since both $Z^k(M)$ and $B^k(M)$ are typically infinite-dimensional.

For every $k > \dim M$ we have $b^k(M) = 0$, since there are no non-trivial k-forms on M for $k > n$.

8.1.4. The Euler characteristic. Let M be a smooth n-manifold whose Betti numbers b^k are all finite. The *Euler characteristic* of M is the integer

$$\chi(M) = \sum_{i=0}^{n} (-1)^i b^i(M).$$

This is an ubiquitous invariant, defined also for more general topological spaces.

8.1.5. The zeroest group. As a start, we may easily identify $H^0(M)$ for any smooth manifold M. We first make a general remark: if M has finitely many connected components $M_1, \ldots, M_h$, we naturally get

$$H^k(M) = H^k(M_1) \oplus \cdots \oplus H^k(M_h).$$

For this reason, we usually suppose that M be connected.

Proposition 8.1.6. *If M is connected, there is a natural isomorphism*

$$H^0(M) \cong \mathbb{R}.$$

Proof. The space $Z^0(M)$ consists of all the functions $f : M \to \mathbb{R}$ such that $df = 0$, and $B^0(M)$ is trivial. By taking charts, we see that $df = 0 \Longleftrightarrow f$ is locally constant (that is, every $p \in M$ has a neighbourhood where f is constant) $\Longleftrightarrow f$ is constant, since M is connected. Therefore $H^0(M) = Z^0(M)$ consists of the constant functions and is hence naturally isomorphic to $\mathbb{R}$. □

For a possibly disconnected M, we get the following.

Corollary 8.1.7. *The Betti number $b^0(M)$ equals the number of connected components of M.*

8.1.6. The cohomology algebra. Let M be a smooth manifold, possibly with boundary. We may define the vector space

$$H^*(M) = \bigoplus_{k \geq 0} H^k(M).$$

Proposition 8.1.8. *The exterior product $\wedge$ descends to $H^*(M)$ and gives it the structure of an associative algebra.*

Proof. If $\omega \in Z^k(M)$ and $\eta \in Z^h(M)$ then

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta = 0$$

and hence $\omega \wedge \eta \in Z^{k+h}(M)$. If moreover $\omega \in B^k(M)$, that is $\omega = d\zeta$, we get

$$\omega \wedge \eta = d\zeta \wedge \eta = d(\zeta \wedge \eta) - (-1)^{k-1}\zeta \wedge d\eta = d(\zeta \wedge \eta)$$

and hence $\omega \wedge \eta \in B^{k+h}(M)$. Therefore the wedge product passes to the quotients $H^k(M)$ and $H^h(M)$. □

If $\omega \in H^p(M)$ and $\eta \in H^q(M)$, then $\omega \wedge \eta \in H^{p+q}(M)$. As for $\Omega^*(M)$, the algebra $H^*(M)$ is anticommutative, that is

$$\omega \wedge \eta = (-1)^{pq} \eta \wedge \omega.$$

In particular, if p is odd we get

$$\omega \wedge \omega = 0.$$

8.1.7. Functoriality. Every smooth map $f : M \to N$ induces a linear map

$$f^* : \Omega^k(N) \longrightarrow \Omega^k(M)$$

by pull-back. The map commutes with d and hence it sends close forms to close forms, and exact forms to exact forms. Therefore it induces a map

$$f^* : H^k(N) \longrightarrow H^k(M)$$

and more generally a morphism of algebras

$$f^* : H^*(N) \longrightarrow H^*(M).$$

We may say that cohomology is a *contravariant functor*, where *contravariant* means that arrows are reversed (we go backwards from $H^k(N)$ to $H^k(M)$), and *functor* means that $(f \circ g)^* = g^* \circ f^*$ and $\mathrm{id}_M^* = \mathrm{id}_{H^*(M)}$.

Every diffeomorphism $f : M \to N$ induces an isomorphism $f^* : H^*(N) \to H^*(M)$. In particular M and N have the same Betti numbers.

The reader should compare this functor with the *covariant functor* furnished by the fundamental group, that sends pointed topological spaces (X, x_0) to groups $\pi_1(X, x_0)$.

8.1.8. The line. The De Rham cohomology of $\mathbb{R}$ can be calculated easily.

Proposition 8.1.9. *We have $H^0(\mathbb{R}) = \mathbb{R}$ and $H^k(\mathbb{R}) = 0$ for all $k > 0$.*

Proof. There are no k-forms with $k \geq 2$, so the only thing to prove is that $H^1(\mathbb{R}) = 0$. Given a 1-form $\omega = f(x)dx$, we can define

$$F(x) = \int_0^x f(t)dt$$

and we get $\omega = dF$. Therefore every 1-form is exact and $H^1(\mathbb{R}) = 0$. □

We say that the cohomology of a manifold M is *trivial* if $H^0(M) = \mathbb{R}$ and $H^k(M) = 0$ for all $k > 0$. We will soon discover that the cohomology of $\mathbb{R}^n$ is also trivial for every n.

8.1.9. Integration along submanifolds. Let M be a n-manifold and $S \subset M$ an oriented compact k-submanifold without boundary. Remember that every k-form $\omega \in \Omega^k(M)$ may be integrated over S, so furnishing a linear map

$$\int_S : \Omega^k(M) \longrightarrow \mathbb{R}.$$

By Stokes' Theorem, the integral of an exact form vanishes, and hence this linear map descends to a map in cohomology

$$\int_S : H^k(M) \longrightarrow \mathbb{R}.$$

If M is itself compact, oriented, and without boundary, the map

$$\int_M : H^n(M) \longrightarrow \mathbb{R}$$

is surjective, since as we already remarked every volume form ω has a non-trivial image. This implies that $b^n(M) \geq 1$. We will prove (as a consequence of Poincaré's duality) that $b^n(M) = b^0(M)$ equals the number of connected components of M in this case. It is very important that M be compact, oriented, and without boundary to get this equality.

8.2. The Poincaré Lemma

One important feature of the fundamental group is that it is unaffected by homotopies. We prove here the same thing for the De Rham cohomology. As a consequence, we will show that the cohomology of $\mathbb{R}^n$ is trivial, as that of any contractible manifold. This fact is known as the *Poincaré Lemma* and can be stated as follows: every closed k-form in $\mathbb{R}^n$ is exact if $k \geq 1$. Despite the simplicity of this sentence, its proof is quite involved.

8.2.1. Cochain complexes. Some of the properties of De Rham cohomology may be deduced by purely algebraic means, and work in more general contexts. For these reasons we now reintroduce cohomologies with a purely algebraic language.

A *cochain complex* C is a sequence of vector spaces $C^0, C^1, C^2, \ldots$ with linear maps $d^k \colon C^k \to C^{k+1}$ such that $d^{k+1} \circ d^k = 0$ for all k. We usually indicate d^k by d and write the cochain complex as

$$C^0 \xrightarrow{d} C^1 \xrightarrow{d} C^2 \xrightarrow{d} \ldots$$

The elements in $Z^k = \ker d^k$ are called *cocycles*, and those in $B^k = \operatorname{Im} d^{k-1}$ are the *coboundaries*. The *cohomology* of C is constructed as $H^k = Z^k / B^k$

for every $k \geq 0$. We may indicate it as $H^k(C)$ to stress its dependence on the cochain complex C.

Of course when $C^k = \Omega^k(M)$ we obtain the De Rham cohomology of M, but this general construction applies to many other contexts, so it makes sense to consider it abstractly.

Remark 8.2.1. A *chain complex* is a sequence of vector spaces $C_0, C_1, \ldots$ equipped with maps $d_k \colon C_k \to C_{k-1}$ such that $d \circ d = 0$. The theory of chain complexes is similar and somehow dual to that of *co*chain complexes: one defines the *cycles* as $Z_k = \ker d_k$, the *boundaries* as $B_k = \operatorname{Im} d_{k+1}$, and the *homology group* $H_k = Z_k / B_k$.

A *morphism* between two cochain complexes C and D is a map $f^k \colon C^k \to D^k$ for all $k \geq 0$ such that the following diagram commutes

$$\begin{tikzcd}
\cdots \arrow[r, "d"] & C^{k-1} \arrow[r, "d"] \arrow[d, "f"] & C^k \arrow[r, "d"] \arrow[d, "f"] & C^{k+1} \arrow[r, "d"] \arrow[d, "f"] & \cdots \\
\cdots \arrow[r, "d"] & D^{k-1} \arrow[r, "d"] & D^k \arrow[r, "d"] & D^{k+1} \arrow[r, "d"] & \cdots
\end{tikzcd}$$

We have denoted f^k simply by f. Since f commutes with d, it sends cocycles to cocycles and coboundaries to coboundaries, and hence induces a homomorphism $f_* \colon H^k(C) \to H^k(D)$ for every k.

8.2.2. Cochain homotopy. We introduce an algebraic notion of homotopy that will reflect the notion of homotopy between maps. Let $f, g \colon C \to D$ be two morphisms between cochain complexes. A *cochain homotopy* between them is a linear map $h^k \colon C^k \to D^{k-1}$ for all $k \geq 1$ such that

$$f^k - g^k = d^{k-1} \circ h^k + h^{k+1} \circ d^k$$

for all $k \geq 0$. Shortly, we may write

$$f - g = dh + hd. \tag{19}$$

It is useful to visualise everything by drawing the following diagram:

$$\begin{tikzcd}
\cdots \arrow[r, "d"] & C^{k-1} \arrow[r, "d"] \arrow[d, shift right, "g"'] \arrow[d, shift left, "f"] \arrow[dl, "h"'] & C^k \arrow[r, "d"] \arrow[d, shift right, "g"'] \arrow[d, shift left, "f"] \arrow[dl, "h"'] & C^{k+1} \arrow[r, "d"] \arrow[d, shift right, "g"'] \arrow[d, shift left, "f"] \arrow[dl, "h"'] & \cdots \arrow[dl, "h"'] \\
\cdots \arrow[r, "d"] & D^{k-1} \arrow[r, "d"] & D^k \arrow[r, "d"] & D^{k+1} \arrow[r, "d"] & \cdots
\end{tikzcd}$$

Two morphisms f, g are *homotopic* if there is a cochain homotopy between them. The relevance of homotopies relies in the following fact.

Proposition 8.2.2. *If two cochain maps f, g are homotopic, they induce the same maps in cohomology.*

Proof. For every $a \in C^k$ we have

$$f(a) - g(a) = d(h(a)) + h(d(a)).$$

If $a \in Z^k(C)$ we get $d(a) = 0$ and hence

$$f(a) - g(a) = d(h(a)) \in B^k(D).$$

Therefore f and g induce the same maps on cohomology. $\square$

Having settled the basic algebraic machinery, we now turn back to De Rham cohomology.

8.2.3. Homotopy invariance. We prove the homotopy invariance of De Rham cohomology. Let M and N be smooth manifolds of dimension m and n, possibly with boundary.

Theorem 8.2.3. *Two homotopic smooth maps $f, g \colon M \to N$ induce the same homomorphisms $f^* = g^* \colon H^*(N) \to H^*(M)$ in De Rham cohomology.*

Proof. Let $F \colon M \times [0,1] \to N$ be the homotopy between f and g. We may suppose that F is smooth by Corollary 5.6.9. We build a cochain homotopy

$$h \colon \Omega^k(N) \longrightarrow \Omega^{k-1}(M)$$

between the morphisms $f^*, g^* \colon \Omega^*(N) \to \Omega^*(M)$. This will imply that $f^* = g^*$ in cohomology. The map h is defined as follows: for every $\omega \in \Omega^k(N)$ we define $h(\omega) \in \Omega^{k-1}(M)$ by setting

$$h(\omega)(p) = \int_0^1 i_t^*\Big(\big(\iota_{\frac{\partial}{\partial t}} F^*(\omega)\big)(p,t)\Big)dt$$

for every $p \in M$. Here $\frac{\partial}{\partial t}$ is the constant vector field along $t \in [0,1]$ and $i_t \colon M \to M \times [0,1]$ is the embedding $i_t(p) = (p,t)$. In other words:

$$h(\omega)(p)(v_1, \ldots, v_{k-1}) = \int_0^1 F^*(\omega)(p,t)\Big(\frac{\partial}{\partial t}, v_1, \ldots, v_{k-1}\Big)dt.$$

We now prove that h is indeed a cochain homotopy between f^* and g^*. We drop the point p from the notation. We get:

$$\begin{aligned}
g^*(\omega) - f^*(\omega) = F_1^*(\omega) - F_0^*(\omega) &= \int_0^1 \frac{\partial}{\partial t} F_t^*(\omega)dt = \int_0^1 i_t^* \mathcal{L}_{\frac{\partial}{\partial t}} F^*(\omega)dt \\
&= \int_0^1 i_t^* d\iota_{\frac{\partial}{\partial t}} F^*(\omega)dt + \int_0^1 i_t^*\iota_{\frac{\partial}{\partial t}} dF^*(\omega)dt \\
&= d\int_0^1 i_t^*\iota_{\frac{\partial}{\partial t}} F^*(\omega)dt + \int_0^1 i_t^*\iota_{\frac{\partial}{\partial t}} F^*(d\omega)dt \\
&= dh\omega + hd\omega.
\end{aligned}$$

The third equality follows from the definition of Lie derivative, the fourth is Cartan's magic formula, and the fifth holds because the differential commutes with pull-backs and with integrating along a path of forms. $\square$

Here is an important consequence.

Corollary 8.2.4. *Two homotopically equivalent manifolds have isomorphic De Rham cohomologies.*

Proof. Let $f: M \to N$ and $g: N \to M$ be homotopy equivalences. We may suppose that they are smooth, and the homotopy is also smooth, by Theorems 6.2.13 and 6.2.14. Then $f \circ g \sim \mathrm{id}_N$ and $g \circ f \sim \mathrm{id}_M$ and hence $f^* \circ g^* = \mathrm{id}$ and $g^* \circ f^* = \mathrm{id}$. □

In particular, two homeomorphic manifolds have the same De Rham cohomology. This is a quite remarkable fact: the cohomology groups $H^*(M)$ are defined in an analytic way through k-forms, but the result is in fact independent of the smooth structure.

Corollary 8.2.5. *Every contractible manifold has trivial cohomology.*

Proof. The point (or $\mathbb{R}$, if you prefer) has trivial cohomology. □

Corollary 8.2.6 (Poincaré's Lemma). *Every closed k-form in $\mathbb{R}^n$ is exact, for every $k \geq 1$.*

If M is a manifold with boundary, the inclusion $\mathrm{int}(M) \hookrightarrow M$ is a homotopy equivalence by Exercise 6.1.13, and therefore induces isomorphisms between the cohomology groups of M and $\mathrm{int}(M)$.

8.2.4. Compact orientable manifolds without boundary. We now use the De Rham cohomology to prove a non-trivial topological fact.

Proposition 8.2.7. *A compact oriented manifold M without boundary with $\dim M \geq 1$ is never contractible.*

Proof. The manifold M has a volume form ω by Proposition 7.2.15, and Example 8.1.2 shows that ω is closed but not exact. Therefore $H^n(M) \neq 0$ for $n = \dim M$. In particular the cohomology of M is not trivial. □

Note that the hypothesis "compact" and "without boundary" are both necessary, as the counterexamples $\mathbb{R}^n$ and D^n show. The orientability hypothesis may be removed, but more work is needed for that (for instance, one may use a different kind of cohomology).

With the same techniques, we can in fact prove more.

Proposition 8.2.8. *A compact oriented manifold M without boundary is never homotopy equivalent to any manifold N with $\dim N < \dim M$.*

Proof. If $m = \dim M$, we have $H^m(M) \neq 0$ and $H^m(N) = 0$. □

8.3. The Mayer – Vietoris sequence

We have calculated the De Rham cohomology of contractible spaces, and we are ready for more complicated manifolds. The main tool for calculating $H^*(M)$ for general manifolds M is the *Mayer – Vietoris sequence*, and we introduce it here.

8.3.1. Exact sequences.

8.3.1. Exact sequences. We now introduce some algebra. A (finite or infinite) sequence of real vector spaces and linear maps

$$\ldots \longrightarrow V_{i-1} \xrightarrow{f_{i-1}} V_i \xrightarrow{f_i} V_{i+1} \longrightarrow \ldots$$

is *exact* if $\operatorname{Im} f_i = \ker f_{i+1}$ for all i such that f_i and f_{i+1} are both defined. We let 0 denote a 0-dimensional vector space; the maps $0 \to U$ and $W \to 0$ are of course trivial. For instance, the following sequence

$$0 \longrightarrow V \xrightarrow{f} W$$

is exact $\Longleftrightarrow$ f is injective, and

$$V \xrightarrow{g} W \longrightarrow 0$$

is exact $\Longleftrightarrow$ g is surjective. The sequence

$$0 \longrightarrow U \xrightarrow{f} V \xrightarrow{g} W \longrightarrow 0$$

is exact $\Longleftrightarrow$ f is injective, g is surjective, and $\operatorname{Im} f = \ker g$. An exact sequence of this last type is called a *short exact sequence*.

Exercise 8.3.1. If a sequence

$$\ldots \longrightarrow V_{i-1} \xrightarrow{f_{i-1}} V_i \xrightarrow{f_i} V_{i+1} \longrightarrow \ldots$$

is exact, the following sequences are also exact:

$$\ldots \longleftarrow V_{i-1}^* \xleftarrow{f_{i-1}^*} V_i^* \xleftarrow{f_i^*} V_{i+1}^* \longleftarrow \ldots$$

$$\ldots \longrightarrow V_{i-1} \otimes W \xrightarrow{f_{i-1}\otimes \mathrm{id}} V_i \otimes W \xrightarrow{f_i \otimes \mathrm{id}} V_{i+1} \otimes W \longrightarrow \ldots$$

for every vector space W.

Exercise 8.3.2. For every finite exact sequence of finite-dimensional spaces

$$0 \longrightarrow V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} \ldots \xrightarrow{f_{k-1}} V_k \longrightarrow 0$$

we have

$$\sum_{i=1}^{k} (-1)^i \dim V_i = 0.$$

8.3.2. The long exact sequence. The notion of exact sequence applies also to other algebraic notions like groups, modules, etc. and also to cochain complexes: a short exact sequence of cochain complexes is an exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

where A, B, C are cochain complexes and f, g are morphisms. Exactness means that f is injective, g is surjective, and $\operatorname{Im} f = \ker g$. That is, we have a big planar commutative diagram of morphisms

$$\begin{tikzcd}
 & \vdots \arrow[d, "d"] & \vdots \arrow[d, "d"] & \vdots \arrow[d, "d"] & \\
0 \arrow[r] & A^{k-1} \arrow[r, "f"] \arrow[d, "d"] & B^{k-1} \arrow[r, "g"] \arrow[d, "d"] & C^{k-1} \arrow[r] \arrow[d, "d"] & 0 \\
0 \arrow[r] & A^{k} \arrow[r, "f"] \arrow[d, "d"] & B^{k} \arrow[r, "g"] \arrow[d, "d"] & C^{k} \arrow[r] \arrow[d, "d"] & 0 \\
0 \arrow[r] & A^{k+1} \arrow[r, "f"] \arrow[d, "d"] & B^{k+1} \arrow[r, "g"] \arrow[d, "d"] & C^{k+1} \arrow[r] \arrow[d, "d"] & 0 \\
 & \vdots & \vdots & \vdots &
\end{tikzcd} \tag{20}$$

where every horizontal line is a short exact sequence of vector spaces.

Theorem 8.3.3. *Every short exact sequence of cochain complexes*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0 \tag{21}$$

induces naturally an exact sequence in cohomology

$$\cdots \longrightarrow H^k(A) \xrightarrow{f_*} H^k(B) \xrightarrow{g_*} H^k(C) \xrightarrow{\delta} H^{k+1}(A) \longrightarrow \cdots \tag{22}$$

for some appropriate morphism δ.

Proof. The morphism

$$\delta \colon H^k(C) \longrightarrow H^{k+1}(A)$$

is defined as follows. Given a cocycle $\gamma \in C^k$, by surjectivity of g there is a $\beta \in B^k$ with $g(\beta) = \gamma$. We have

$$g(d\beta) = dg(\beta) = d\gamma = 0$$

because γ is a cocycle. Since $\operatorname{Im} f = \ker g$ there is an $\alpha \in A^{k+1}$ such that $f(\alpha) = d\beta$, and we set

$$\delta([\gamma]) = [\alpha].$$

There are now a number of things to check, and we leave to the reader the pleasure of proving all of them through "diagram chasing." Here are they:

- α is a cocycle, that is $d\alpha = 0$;
- the class $[\alpha] \in H^{k+1}(A)$ does not depend on the choices of β and α; moreover, it only depends on the class $[\gamma]$ of γ;
- if γ is a coboundary then α also is.

This shows that δ is well-defined. Finally, we have to show that the sequence (22) is exact. Have fun! □

The sequence (22) is called the *long exact sequence* induced by the short exact sequence (21).

8.3.3. The Mayer – Vietoris sequence. It is now time to go back to smooth manifolds and their De Rham cohomology.

Let M be a smooth manifold, possibly with boundary, and $U, V \subset M$ be two open subsets covering M, that is with $U \cup V = M$. The inclusions

$$\begin{tikzcd} & U \arrow[dr, "l"] & \\ U \cap V \arrow[ur, "i"] \arrow[dr, "j"'] & & M \\ & V \arrow[ur, "m"'] & \end{tikzcd}$$

induce the morphisms in cohomology

$$\begin{tikzcd} & H^k(U) \arrow[dl, "i^*"'] & \\ H^k(U \cap V) & & H^k(M) \arrow[ul, "l^*"'] \arrow[dl, "m^*"] \\ & H^k(V) \arrow[ul, "j^*"] & \end{tikzcd}$$

Theorem 8.3.4 (Mayer – Vietoris Theorem). *There is an exact sequence*

$$\cdots \longrightarrow H^k(M) \xrightarrow{(l^*, m^*)} H^k(U) \oplus H^k(V) \xrightarrow{i^* - j^*} H^k(U \cap V) \xrightarrow{\delta} H^{k+1}(M) \longrightarrow \cdots$$

for some canonically defined map δ.

Proof. This is the long exact sequence obtained via Theorem 8.3.3 from the short exact sequence of cochain complexes

$$0 \longrightarrow \Omega^*(M) \xrightarrow{(l^*, m^*)} \Omega^*(U) \oplus \Omega^*(V) \xrightarrow{i^* - j^*} \Omega^*(U \cap V) \longrightarrow 0.$$

We only need to check that this short sequence is indeed exact. Note that the morphisms l^*, m^*, i^*, and j^* are just restrictions of k-forms to open subsets. There are four things to check, and these are easily proved by keeping in mind that these morphisms are just restrictions:

- The map (l^*, m^*) is injective: if $l^*(\omega) = m^*(\omega) = 0$, then $\omega = 0$.
- Since $i^* \circ l^* = j^* \circ m^*$, we get $(i^* - j^*) \circ (l^*, m^*) = 0$.
- If (α, β) is such that $i^*(\alpha) = j^*(\beta)$, then α and β agree on $U \cap V$ and hence are restrictions of a global form in M.
- To prove that $i^* - j^*$ is surjective, pick a partition of unity ρ_U, ρ_V subordinate to $\{U, V\}$. Given $\omega \in \Omega^k(U \cap V)$, note that $\rho_V \omega$ extends smoothly to U simply by setting it constantly zero on $U \backslash V$. Therefore $\rho_V \omega \in \Omega^k(U)$ and $\rho_U \omega \in \Omega^k(V)$ and we can write
$$(i^* - j^*)(\rho_V \omega, -\rho_U \omega) = (\rho_U + \rho_V)\omega = \omega.$$

The proof is complete. □

The exact sequence resulting from Theorem 8.3.4 is called the *Mayer – Vietoris long exact sequence* induced by the covering $\{U, V\}$ of M. Recall that $H^k(M) = 0$ whenever $k > n = \dim M$, so the Mayer – Vietoris sequence may be long but is certainly finite. It starts and ends as follows:
$$0 \longrightarrow H^0(M) \longrightarrow H^0(U) \oplus H^0(V) \longrightarrow \cdots \longrightarrow H^n(U \cap V) \longrightarrow 0.$$

Note that U and V are not necessarily connected. The four morphisms i^*, j^*, l^*, m^* are simply restrictions of k-forms. The morphism δ is a bit more complicated, and for most applications we do not need to understand it, so the reader may decide to jump to the next section. Just in case, here is a description of δ. Let ρ_U, ρ_V be a partition of unity subordinated to $\{U, V\}$. Given a closed k-form $\omega \in \Omega^k(U \cap V)$, we may consider the $(k+1)$-form
$$\eta = d\rho_V \wedge \omega = -d\rho_U \wedge \omega \in \Omega^{k+1}(U \cap V).$$
The two expressions coincide since $d\rho_U + d\rho_V = 0$. The 1-forms $d\rho_V$ and $d\rho_U$ are actually defined on M and with support in $U \cap V$. This implies that we can extend η to a form $\eta \in \Omega^{k+1}(M)$ by setting it to be zero on any point in $M \setminus (U \cap V)$.

Proposition 8.3.5. *We have* $\delta([\omega]) = [\eta]$.

Proof. The proofs of Theorems 8.3.3 and 8.3.4 show that $\delta([\omega])$ is constructed by picking the preimage $(\rho_V \omega, -\rho_U \omega)$ of ω, then differentiating
$$\big(d(\rho_V \omega), -d(\rho_U \omega)\big) = (d\rho_V \wedge \omega, -d\rho_U \wedge \omega)$$
using $d\omega = 0$, and finally noting that the pair is the image of η. □

8.3.4. Cohomology of spheres. As a reward for all the effort that we made with short and long exact sequences, we can now easily calculate the De Rham cohomology of spheres.

Proposition 8.3.6. *For every $n \geq 1$ we have*
$$H^0(S^n) \cong H^n(S^n) \cong \mathbb{R}, \qquad H^k(S^n) = 0 \ \ \forall k \neq 0, n.$$

Proof. Using stereographic projections along opposite poles we may cover S^n as $S^n = U \cup V$ with $U \cong V \cong \mathbb{R}^n$ and also $U \cap V \cong S^{n-1} \times \mathbb{R}$. By homotopy equivalence, we have $H^*(U \cap V) \cong H^*(S^{n-1})$.

We first examine the case $n = 1$. Remember that $H^k(M) = 0$ whenever $k > \dim M$ and $H^k(\mathbb{R}^n) = 0$ for all $k > 0$. The Mayer – Vietoris sequence is

$$0 \longrightarrow H^0(S^1) \longrightarrow H^0(\mathbb{R}^1) \oplus H^0(\mathbb{R}^1) \longrightarrow H^0(S^0) \xrightarrow{\delta} H^1(S^1) \longrightarrow 0$$

which translates as

$$0 \longrightarrow \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} \longrightarrow H^1(S^1) \longrightarrow 0.$$

since S^0 has two connected components. Exercise 8.3.2 gives $H^1(S^1) \cong \mathbb{R}$.

We now consider the case $n \geq 2$. The Mayer – Vietoris sequence breaks into pieces since $H^k(\mathbb{R}^n) \oplus H^k(\mathbb{R}^n) = 0$ for all $k > 0$. It starts with

$$0 \longrightarrow H^0(S^n) \longrightarrow H^0(\mathbb{R}^n) \oplus H^0(\mathbb{R}^n) \longrightarrow H^0(S^{n-1}) \xrightarrow{\delta} H^1(S^n) \longrightarrow 0$$

which translates as

$$0 \longrightarrow \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} \longrightarrow \mathbb{R} \longrightarrow H^1(S^n) \longrightarrow 0.$$

Therefore $H^1(S^n) = 0$. Then for every $2 \leq k \leq n$ we get

$$0 \longrightarrow H^{k-1}(S^{n-1}) \xrightarrow{\delta} H^k(S^n) \longrightarrow 0$$

and therefore $H^k(S^n) \cong H^{k-1}(S^{n-1})$. We conclude by induction on n. $\square$

8.3.5. Complex projective spaces. The De Rham cohomology of the complex projective spaces is quite different from that of the spheres, and is in fact very interesting:

Proposition 8.3.7. *We have*

$$H^k(\mathbb{CP}^n) = \begin{cases} \mathbb{R} & \text{if } k \text{ is even and } k \leq 2n, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. When $n = 1$ we have $\mathbb{CP}^1 \cong S^2$ and the theorem is proved. So we proceed by induction on n and suppose $n \geq 2$. Let $H \subset \mathbb{CP}^n$ be a hyperplane. Of course H is diffeomorphic to $\mathbb{CP}^{n-1}$, and the inclusion $i\colon H \hookrightarrow \mathbb{CP}^n$ induces a morphism

$$i^*\colon H^k(\mathbb{CP}^n) \longrightarrow H^k(H) \cong H^k(\mathbb{CP}^{n-1}).$$

We will prove that i^* is an isomorphism for all $k < 2n$ and that $H^{2n}(\mathbb{CP}^n) \cong \mathbb{R}$. From this we conclude by our induction hypothesis.

Pick a point $p \in \mathbb{CP}^n$ not contained in H. Choose the open sets

$$U = \mathbb{CP}^n \setminus H, \qquad V = \mathbb{CP}^n \setminus \{p\}.$$

We have the diffeomorphisms

$$U \cong \mathbb{R}^{2n}, \qquad U \cap V \cong \mathbb{R}^{2n} \setminus \{p\} \cong S^{2n-1} \times \mathbb{R}.$$

The pencil of complex lines passing through p gives V the structure of a $\mathbb{C}$-bundle over $H \cong \mathbb{CP}^{n-1}$. In fact for our purposes it suffices to note that V deformation retracts onto H. In particular, we have the homotopy equivalences

$$U \sim \{\mathrm{pt}\}, \qquad U \cap V \sim S^{2n-1}, \qquad V \sim \mathbb{CP}^{n-1}.$$

The Mayer – Vietoris sequence starts as

$$\begin{aligned} 0 \longrightarrow & H^0(\mathbb{CP}^n) \xrightarrow{(i^*,j^*)} H^0(\mathbb{CP}^{n-1}) \oplus H^0(\mathrm{pt}) \longrightarrow H^0(S^{2n-1}) \\ \xrightarrow{\delta} & H^1(\mathbb{CP}^n) \xrightarrow{i^*} H^1(\mathbb{CP}^{n-1}) \longrightarrow 0. \end{aligned}$$

This is isomorphic to

$$0 \longrightarrow \mathbb{R} \xrightarrow{(i^*,j^*)} \mathbb{R} \oplus \mathbb{R} \longrightarrow \mathbb{R} \longrightarrow H^1(\mathbb{CP}^n) \xrightarrow{i^*} H^1(\mathbb{CP}^{n-1}) \longrightarrow 0$$

and we deduce easily that both i^* are isomorphisms. If $1 < k < 2n-1$ we get

$$0 = H^{k-1}(S^{2n-1}) \longrightarrow H^k(\mathbb{CP}^n) \xrightarrow{i^*} H^k(\mathbb{CP}^{n-1}) \longrightarrow H^k(S^{2n-1}) = 0$$

and therefore i^* is again an isomorphism. The end of the sequence is

$$\begin{aligned} 0 = H^{2n-2}(S^{2n-1}) \longrightarrow & H^{2n-1}(\mathbb{CP}^n) \xrightarrow{i^*} H^{2n-1}(\mathbb{CP}^{n-1}) \\ \longrightarrow H^{2n-1}(S^{2n-1}) \longrightarrow & H^{2n}(\mathbb{CP}^n) \longrightarrow 0. \end{aligned}$$

This is isomorphic to

$$0 \longrightarrow H^{2n-1}(\mathbb{CP}^n) \xrightarrow{i^*} 0 \longrightarrow \mathbb{R} \longrightarrow H^{2n}(\mathbb{CP}^n) \longrightarrow 0.$$

Therefore i^* is a (trivial) isomorphism also in this case and $H^{2n}(\mathbb{CP}^n) \cong \mathbb{R}$. □

Corollary 8.3.8. *The manifolds S^{2n} and $\mathbb{CP}^n$ are not diffeomorphic, and in fact not even homotopy equivalent, when $n > 1$.*

Proof. The Betti numbers of S^{2n} and $\mathbb{CP}^n$ are respectively

$$\begin{gathered} 1, 0, 0, 0, \ldots, 0, 0, 1; \\ 1, 0, 1, 0, \ldots, 1, 0, 1. \end{gathered}$$

These sequences differ when $n > 1$. □

Along the proof of Proposition 8.3.7 we have also shown that the inclusion $i \colon H \hookrightarrow \mathbb{CP}^n$ of any hyperplane H induces a surjective map i^* in cohomology. By iteration, this remains true if we substitute H with a projective subspace S of any dimension.

8.4. Compactly supported forms

We now introduce a variation of De Rham cohomology that considers only forms with compact supports. We will see that this variation has a somehow dual behaviour with respect to the ordinary De Rham cohomology.

8.4.1. Definition. Let M be a smooth manifold, possibly with boundary. For every $k \geq 0$ we define the vector subspace

$$\Omega^k_c(M) \subset \Omega^k(M)$$

that consists of all the k-forms having compact support. Of course if M is compact we have $\Omega^k_c(M) = \Omega^k(M)$. The differential restrict to a map

$$d \colon \Omega^k_c(M) \longrightarrow \Omega^{k+1}_c(M)$$

with $d^2 = 0$. As above, we get a cochain complex $\Omega^*_c(M)$, and its cohomology is called the *De Rham cohomology with compact support*

$$H^k_c(M).$$

The wedge product of two compactly supported forms is also compactly supported, hence the operation $\wedge$ is defined also in this context and gives $H^*_c(M) = \oplus_k H^k_c(M)$ the structure of an associative algebra.

Of course when M is compact we get nothing new, but $H^k_c(M)$ may differ considerably from $H^k(M)$ when M is not compact, as we now show.

8.4.2. The zeroeth group. We now study $H^0_c(M)$ and notice immediately a difference between the compact and the non compact case.

As with De Rham cohomology, if M has finitely many connected components $M_1, \ldots, M_k$ we get $H^0_c(M) = H^0_c(M_1) \oplus \cdots \oplus H^0_c(M_k)$, so one usually considers only connected manifolds.

Proposition 8.4.1. *Let M be connected. If M is compact then $H^0_c(M) = \mathbb{R}$, while if M is not compact then $H^0_c(M) = 0$.*

Proof. The space $H^0_c(M)$ consists of all the compactly supported constant functions. Non-trivial such functions exist only if M is compact. □

Let $b^k_c(M) = \dim H^k_c(M)$ be the compactly supported k-th Betti number. We have discovered that $b^0_c(M)$ equals the number of compact connected components of M. As in ordinary De Rham cohomology, we have $b^k_c(M) = 0$ for every $k > \dim M$.

8.4.3. The n-th group. Let M be an oriented n-manifold without boundary. A compactly supported n-form in M can be integrated, so we get a map

$$\int_M \colon \Omega^n_c(M) \longrightarrow \mathbb{R}.$$

By Stokes' Theorem, the integral of an exact form vanishes (since $\partial M = \varnothing$) hence this linear map descends to a map in cohomology

$$\int_M \colon H^n_c(M) \longrightarrow \mathbb{R}.$$

This map is non-trivial (pick a bump n-form), hence it is surjective. This shows in particular that $b^n_c(M) > 0$. As opposite to the ordinary De Rham cohomology, the compactly supported one detects the dimension n of the manifold M, that is equal to the maximum k such that $b_c^k(M) \neq 0$.

8.4.4. The line. As usual we start by considering the line $\mathbb{R}$.

Proposition 8.4.2. *We have $H_c^1(\mathbb{R}) \cong \mathbb{R}$ and $H_c^k(\mathbb{R}) = 0$ for all $k \neq 1$.*

Proof. We already know that $H_c^k(\mathbb{R}) = 0$ for $k = 0$ and $k \geq 2$, so we turn to the case $k = 1$. The integration map

$$\int_{\mathbb{R}} : H_c^1(\mathbb{R}) \longrightarrow \mathbb{R}$$

is surjective. If $\omega = g(x)dx$ is such that $\int \omega = 0$, we may define $f(x) = \int_{-\infty}^x g(t)dt$ and get a compactly supported f with $\omega = df$. Therefore the integration map is also injective. □

We note that $H_c^k(\mathbb{R}) \cong H^{1-k}(\mathbb{R})$. This is not an accident, as we will see.

8.4.5. Functoriality? If $f: M \to N$ is a *proper* map, the pull-back $f^*\omega$ of $\omega \in \Omega_c^k(N)$ is compactly supported also in M and we get a morphism

$$f^*: \Omega_c^k(N) \longrightarrow \Omega_c^k(M)$$

that commutes with d and hence passes to cohomology groups

$$f^*: H_c^k(N) \longrightarrow H_c^k(M).$$

However, if f is not proper the pull-back is not defined in this context. So we can say that contravariant functoriality holds only for proper maps.

On the other hand, the compactly supported cohomology displays some covariant behaviour: every inclusion map $i: U \hookrightarrow M$ of some open subset U induces the *extension morphism*

$$i_*: \Omega_c^k(U) \longrightarrow \Omega_c^k(M)$$

defined simply by extending k-forms to be zero outside of U. This does not work for general k-forms (extensions would not be smooth, nor continuous, in general). Extensions commute with d and hence we get

$$i_*: H_c^k(U) \longrightarrow H_c^k(M).$$

8.4.6. Poincaré Lemma. We now prove a version of the Poincaré Lemma for compactly supported k-forms in $\mathbb{R}^n$.

Theorem 8.4.3. *We have $H_c^n(\mathbb{R}^n) = \mathbb{R}$ and $H_c^k(\mathbb{R}^n) = 0$ for all $k \neq n$.*

Proof. We identify $\mathbb{R}^n$ with $S^n \setminus \{p\}$ for some $p \in S^n$. We first consider the case $0 < k < n$. Let $\omega \in \Omega_c^k(\mathbb{R}^n)$ be a closed k-form with $0 < k < n$. We need to prove that it is exact.

Since ω has compact support we may extend it to a form in S^n. Since $H^k(S^n) = 0$, we have $\omega = d\eta$ for some $\eta \in \Omega^{k-1}(S^n)$. The support of η may not be contained in $\mathbb{R}^n$, so we now modify η to another form η' with support in $\mathbb{R}^n$ that still satisfies $d\eta' = \omega$.

Let $B \subset \mathbb{R}^n$ be a ball of some radius containing the support of ω. Note that $d\eta = 0$ on $S^n \setminus B$. If $k = 1$, then η is a function on S^n that has some constant value c on $S^n \setminus B$. By setting $\eta' = \eta - c$ we get $d\eta' = d\eta = \omega$ and η' has support in B, so we are done.

We suppose that $k > 1$. Since $S^n \setminus B$ is contractible, there is an $\alpha \in \Omega^{k-2}(S^n \setminus B)$ such that $d\alpha = \eta$. Pick a bump function ρ with support in $S^n \setminus B$ that equals 1 on a neighbourhood of p. Now $\rho\alpha$ extends to S^n and

$$\eta' = \eta - d(\rho\alpha) \in \Omega^{k-1}(S^n)$$

vanishes near p, so it gives a compactly supported form in $\mathbb{R}^n$. We have $d\eta' = d\eta = \omega$ and hence we are done.

In the case $k = n$ we need to prove that the integration map

$$\int_{\mathbb{R}^n} \colon H^n_c(\mathbb{R}^n) \longrightarrow \mathbb{R}$$

is injective. Let $\omega \in \Omega^n_c(\mathbb{R}^n)$ be a closed form with $\int_{\mathbb{R}^n} \omega = 0$. We extend it to a form in S^n. We already know that $\int_{S^n} \colon H^n(S^n) \to \mathbb{R}$ is an isomorphism. Since $\int_{S^n} \omega = 0$, the form ω is exact in S^n and we conclude as above. $\square$

Corollary 8.4.4. *Every compactly supported closed k-form in $\mathbb{R}^n$ with $k \neq n$ is the differential of a compactly supported $(k-1)$-form.*

Example 8.4.5. Every electromagnetic tensor field $F \in \Omega^2_c(\mathbb{R}^{3,1})$ with compact support is closed by Maxwell's equations; hence it is the differential $F = dA$ of a potential 1-form $A \in \Omega^1_c(\mathbb{R}^{3,1})$ with compact support.

We keep observing that $H^k_c(\mathbb{R}^n) = H^{n-k}(\mathbb{R}^n)$ for all n and k. We also note that the compactly supported cohomology is evidently *not* invariant under homotopy equivalence.

8.4.7. The Mayer – Vietoris sequence. The compactly supported version of De Rham cohomology also has a Mayer – Vietoris sequence, which however presents some important differences with respect to the ordinary one.

Let M be a smooth manifold possibly with boundary, and $U, V \subset M$ be two open subsets covering M. The inclusions

$$\begin{tikzcd}
 & U \arrow[dr, "l"] & \\
U \cap V \arrow[ur, "i"] \arrow[dr, "j"'] & & M \\
 & V \arrow[ur, "m"'] &
\end{tikzcd}$$

induce the extension morphisms in cohomology

$$\begin{tikzcd}
 & H_c^k(U) \arrow[dr, "l_*"] & \\
H_c^k(U \cap V) \arrow[ur, "i_*"] \arrow[dr, "j_*"'] & & H_c^k(M) \\
 & H_c^k(V) \arrow[ur, "m_*"'] &
\end{tikzcd}$$

Theorem 8.4.6 (Mayer – Vietoris Theorem). *There is an exact sequence*

$$\cdots \longrightarrow H_c^k(U\cap V) \xrightarrow{(i_*,-j_*)} H_c^k(U)\oplus H_c^k(V) \xrightarrow{l_*+m_*} H_c^k(M) \xrightarrow{\delta} H_c^{k+1}(U\cap V) \longrightarrow \cdots$$

for some canonically defined map δ*.*

Proof. The sequence of complexes

$$0 \longrightarrow \Omega_c^*(U \cap V) \xrightarrow{(i_*,-j_*)} \Omega_c^*(U) \oplus \Omega_c^*(V) \xrightarrow{l_*+m_*} \Omega_c^*(M) \longrightarrow 0$$

is easily seen to be exact: use a partition of unity to show that $l_* + m_*$ is surjective. $\square$

Note that this Mayer – Vietoris sequence is different in nature from the one that we obtained from Theorem 8.3.4.

Exercise 8.4.7. Use the Mayer – Vietoris sequence to confirm that

$$H_c^k(S^n) = H^k(S^n) = \begin{cases} \mathbb{R} & \text{if } k = 0, n, \\ 0 & \text{otherwise.} \end{cases}$$

We cannot refrain from noting again that $H_c^k(S^n) = H^{n-k}(S^n)$. As in ordinary De Rham cohomology, we can write δ explicitly. Let ρ_U, ρ_V be a partition of unity subordinate to $\{U, V\}$. Given $\omega \in \Omega_c^k(M)$ we can define

$$\eta = d\rho_V \wedge \omega = -d\rho_U \wedge \omega \in \Omega_c^{k+1}(U \cap V)$$

and as in the ordinary case we find that $\delta([\omega]) = [\eta]$.

8.4.8. Countably many connected components. We point out another difference between $H^k(M)$ and $H_c^k(M)$. Remember that $\prod_i V_i$ is the space of all sequences $(v_1, v_2, \ldots)$ with $v_i \in V_i$ while $\oplus_i V_i$ is the subspace of all sequences having only finitely many non-zero elements.

Proposition 8.4.8. *Let* M *have countably many connected components* M_1*,* $M_2, \ldots$ *We have the canonical isomorphisms*

$$H^k(M) = \prod_i H^k(M_i), \qquad H_c^k(M) = \bigoplus_i H_c^k(M_i).$$

Proof. We have

$$\Omega^k(M) = \prod_i \Omega^k(M_i), \qquad \Omega^k_c(M) = \bigoplus_i \Omega^k_c(M_i).$$

The point is that a k-form ω in M is like a k-form ω_i in each M_i, and ω has compact support $\iff$ each ω_i has and only finitely many ω_i are non-zero. $\square$

8.4.9. Integration along fibres. Let $\pi\colon M \to N$ be a submersion between oriented manifolds without boundary of dimension $m \geq n$.

For every $p \in N$ the fibre $F = \pi^{-1}(p)$ is a manifold of dimension $h = m-n$, with an orientation induced by that of M and N as follows: for every $q \in F$ we say that $v_1, \ldots, v_h \in T_qF$ is a positive basis if it may be completed to a positive basis $v_1, \ldots, v_m$ of T_qM such that $v_{h+1}, \ldots, v_m$ project to a positive basis of T_pN.

We now define a map

$$\pi_*\colon \Omega^k_c(M) \longrightarrow \Omega^{k-h}_c(N)$$

called *integration along fibres*, as follows. For every $p \in N$ and $v_1, \ldots, v_{k-h} \in T_p(N)$ we set

$$\pi_*(\omega)(p)(v_1, \ldots, v_{k-h}) = \int_F \beta$$

where $F = \pi^{-1}(p)$ and $\beta \in \Omega^h_c(F)$ is defined as

$$\beta(q)(w_1, \ldots, w_h) = \omega(w_1, \ldots, w_h, \tilde{v}_1, \ldots, \tilde{v}_{k-h})$$

where $\tilde{v}_i$ is any vector in $T_q(F)$ such that $d\pi_q(\tilde{v}_i) = v_i$.

Proposition 8.4.9. *The form β is well-defined.*

Proof. For any other lift $\tilde{v}'_i$ we get $\tilde{v}'_i = \tilde{v}_i + \lambda_1 w_1 + \ldots + \lambda_h w_h$ and hence

$$\omega(w_1, \ldots, w_h, \ldots, \tilde{v}'_i, \ldots) = \omega(w_1, \ldots, w_h, \ldots, \tilde{v}_i, \ldots)$$

since $\omega(w_1, \ldots, w_h, \ldots, \lambda_j w_j, \ldots) = 0$. $\square$

The definition is a bit abstract, so we describe an important example more explicitly. Consider the projection

$$\pi\colon \mathbb{R}^n \times \mathbb{R}^h \longrightarrow \mathbb{R}^n$$

on the first factor, and use variables $x^1, \ldots, x^n$ for $\mathbb{R}^n$ and $y^1, \ldots, y^h$ for $\mathbb{R}^h$. We examine the k-form

$$\omega = f\, dx^I \wedge dy^J$$

where f has compact support and I, J are multi-indices.

Exercise 8.4.10. If $J = (1, \ldots, h)$ then

$$\pi_*(\omega) = \left(\int_{\mathbb{R}^h} f(x,y) dy^J\right) dx^I,$$

while if $J \neq (1, \ldots, n)$ we have $\pi_*(\omega) = 0$.

Proposition 8.4.11. *The linear map π_* commutes with differentials and hence descends to a map in cohomology*

$$\pi_*\colon H^k_c(M) \longrightarrow H^{k-h}_c(N).$$

Proof. We must prove that $\pi_*(d\omega) = d\pi_*(\omega)$ for every $\omega \in \Omega^k_c(M)$. By the normal form of submersions every $p \in M$ has an open neighbourhood $U(p)$ such that $\pi|_{U(p)}$ looks like a projection

$$\pi\colon \mathbb{R}^n \times \mathbb{R}^h \longrightarrow \mathbb{R}^n$$

onto the first factor. Using a partition of unity we may write ω as a finite sum of k-forms, each with support contained in some $U(p)$. By linearity it suffices to prove the theorem for one such k-form. Summing up, it suffices to consider the case where $\pi\colon \mathbb{R}^n \times \mathbb{R}^h \to \mathbb{R}^n$ and $\omega \in \Omega^k_c(\mathbb{R}^n \times \mathbb{R}^h)$.

Again by linearity we may restrict to the case

$$\omega = f\,dx^I \wedge dy^J$$

already studied above. If $J = (1, \dots, h)$ we get

$$\pi_*(\omega) = \left(\int_{\mathbb{R}^h} f(x,y)dy^J\right) dx^I$$

and hence

$$\begin{aligned} d\pi_*(\omega) &= \sum_{i=1}^n \frac{\partial}{\partial x^i}\left(\int_{\mathbb{R}^h} f(x,y)dy^J\right) dx^i \wedge dx^I \\ &= \left(\int_{\mathbb{R}^h} \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x,y)dy^J\right) dx^i \wedge dx^I = \pi_* d(\omega). \end{aligned}$$

If $J \neq (1, \dots, h)$ we get $\pi_*(\omega) = 0$ and $\pi_*(d\omega) = 0$. The only non-trivial case is when $|J| = h-1$, so consider for simplicity $J = (2, \dots, h)$. We get

$$\pi_*(d\omega) = \left(\int_{\mathbb{R}^h} \frac{\partial f}{\partial y^1}(x,y)dy^J\right) dx^I = 0$$

because f has compact support. (Here we need $\partial M = \varnothing$.) □

We have discovered that every submersion $f\colon M \to N$ between oriented manifolds induces a linear map

$$\pi_*\colon H^k_c(M) \longrightarrow H^{k-h}_c(N).$$

The map π_* is called *integration along fibres*. It may be characterised by the following geometric property. Let $S \subset N$ be an oriented $(k-h)$-submanifold. Then $W = f^{-1}(S)$ is an oriented k-submanifold of M.

Exercise 8.4.12. For every $\omega \in \Omega^k_c(M)$ we have

$$\int_S \pi_*(\omega) = \int_W \omega.$$

In particular when $k = h$ we get a linear map $\pi_*\colon \Omega^h_c(M) \to C^\infty_c(N)$ that sends ω to the function

$$f(p) = \pi_*(\omega)(p) = \int_F \omega$$

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

8.4.10. Smooth coverings. Let $M \to N$ be a smooth covering between smooth n-manifolds without boundary. A covering is a submersion, and the integration along fibres is a map

$$\pi_*\colon H^k_c(M) \longrightarrow H^k_c(N).$$

In this case the integration along the fibres is just a summation, that is

$$\pi_*(\omega)(p)(v_1, \ldots, v_n) = \sum_{\pi(q)=p} \omega(q)(\tilde v_1, \ldots, \tilde v_n)$$

where $v_i \in T_pN$ and $\tilde v_i = d\pi_q^{-1}(v_i)$. Here is a remarkable application.

Proposition 8.4.13. *If $\pi\colon M \to N$ is a covering of finite degree d, then $\pi^*\colon H^k_c(N) \to H^k_c(M)$ is injective.*

Proof. We have $\frac{1}{d}\pi_* \circ \pi^* = \mathrm{id}$ on $H^k_c(N)$. □

If the covering has infinite degree the maps in cohomology need not to be injective, as the universal covering $\mathbb{R} \to S^1$ easily shows.

8.5. Poincaré duality

We have already noted that $H^k(M) \cong H^{n-k}_c(M)$ on many n-manifolds M, and we now prove this equality in a much wider generality. We emphasize that all the manifolds considered in this section are oriented and without boundary!

8.5.1. The Poincaré bilinear map. Let M be an oriented smooth manifold without boundary. We define the *Poincaré bilinear map*

$$\Omega^k(M) \times \Omega^{n-k}_c(M) \longrightarrow \mathbb{R}$$

by sending the pair (ω, η) to the real number

$$\langle \omega, \eta \rangle = \int_M \omega \wedge \eta.$$

The map is well-defined since $\omega \wedge \eta$ has compact support. We can see easily using Stokes and $\partial M = \varnothing$ that it passes to cohomology groups

$$H^k(M) \times H^{n-k}_c(M) \longrightarrow \mathbb{R}.$$

As every bilinear form, it induces a map

$$\mathrm{PD}\colon H^k(M) \longrightarrow H^{n-k}_c(M)^*$$

that sends $[\omega]$ to the functional $[\eta] \mapsto \langle [\omega], [\eta] \rangle$. We dedicate this section to proving the following.

Theorem 8.5.1 (Poincaré duality). *The map* PD *is an isomorphism.*

As usual, we will need a bit of preliminary homological algebra.

8.5.2. The Five Lemma. The following lemma is solved by diagram chasing, and we leave it to the reader as an exercise – there is certainly much more fun in trying to solve it alone than in reading a boring sequence of implications.

Exercise 8.5.2 (The Five Lemma). Given the following commutative diagram of abelian groups and morphisms

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "\alpha"] & B \arrow[r, "g"] \arrow[d, "\beta"] & C \arrow[r, "h"] \arrow[d, "\gamma"] & D \arrow[r, "i"] \arrow[d, "\delta"] & E \arrow[d, "\epsilon"] \\
A' \arrow[r, "j"'] & B' \arrow[r, "k"'] & C' \arrow[r, "l"'] & D' \arrow[r, "m"'] & E'
\end{tikzcd}$$

in which the rows are exact, if $\alpha, \beta, \delta, \epsilon$ are isomorphisms then γ also is.

8.5.3. Induction on open subsets. Let M be a smooth manifold. We want to prove the Poincaré duality Theorem by induction on open subsets of M, starting with those diffeomorphic to $\mathbb{R}^n$ and then passing to more complicated ones in a controlled way. We will need the following.

Let $\mathcal{A}$ be the collection of open subsets in M determined by the rules:

(1) $\mathcal{A}$ contains all the open subsets diffeomorphic to $\mathbb{R}^n$,
(2) if $U, V, U \cap V \in \mathcal{A}$, then $U \cup V \in \mathcal{A}$,
(3) if $U_i \in \mathcal{A}$ are pairwise disjoint, then $\cup U_i \in \mathcal{A}$.

Note that in the last point there can be infinitely many disjoint sets U_i (they are always countable, since M is second countable).

Lemma 8.5.3. *Every open subset of M belongs to $\mathcal{A}$.*

Proof. The proof is subdivided into steps.

(1) If $U_1, \ldots, U_k \in \mathcal{A}$ and all their intersections lie in $\mathcal{A}$, then also $U_1 \cup \cdots \cup U_k \in \mathcal{A}$.
(2) If $\{U_i\} \subset \mathcal{A}$ is a locally finite countable family, with $\overline{U_i}$ compact for all i, and such that all the finite intersections also lie in $\mathcal{A}$, then $\cup U_i \in \mathcal{A}$.
(3) If $U \subset M$ is diffeomorphic to an open subset $V \subset \mathbb{R}^n$, then $U \in \mathcal{A}$.
(4) Every open subset of M belongs to $\mathcal{A}$.

Point (1) is a simple exercise (prove it by induction on k). Concerning (2), we may suppose that $U = \cup U_i$ is connected, and note that every U_i intersects only finitely many U_j.

We define some new open subsets by setting $W_0 = U_0$ and defining W_{i+1} as the union of all the U_j that intersect W_i and are not contained in $\cup_{a \leq i} W_a$.

Each W_i contains finitely many U_j and hence $W_i \in \mathcal{A}$ by (1). Note that $W_i \cap W_{i+2} = \varnothing$ for all i. We set

$$Z_0 = \sqcup_i W_{2i}, \qquad Z_1 = \sqcup_i W_{2i+1}.$$

We easily get $Z_0, Z_1 \in \mathcal{A}$ and also $Z_0 \cap Z_1 \in \mathcal{A}$, so $U = Z_0 \cup Z_1 \in \mathcal{A}$.

Concerning (3), we note that V is covered by multi-rectangles $(a_1, b_1) \times \cdots \times (a_n, b_n)$ whose closure is contained in V. Every finite intersection is again a product, so all these sets and their intersections are diffeomorphic to $\mathbb{R}^n$ and hence lie in $\mathcal{A}$. This cover can be made locally finite using an exhaustion of V by compact sets. Now (2) applies and we get $U \in \mathcal{A}$.

Finally, by taking an adequate atlas for M (see Proposition 3.3.2) we find a locally finite covering U_i such that every U_i is diffeomorphic to $\mathbb{R}^n$ and has compact closure. The intersections are diffeomorphic to open subsets of $\mathbb{R}^n$ and hence are in $\mathcal{A}$ by (3). We conclude again by (2). □

We have proved in particular that $M \in \mathcal{A}$.

8.5.4. Proof of the Poincaré duality. We can now prove Theorem 8.5.1.

Proof. Let $\mathcal{B}$ be the collection of the open subsets U of M where Poincaré duality holds, that is such that $\mathrm{PD}\colon H^k(U) \to H_c^{n-k}(U)^*$ is an isomorphism for all k. Our aim is of course to prove that $M \in \mathcal{B}$, so we prove that $\mathcal{B}$ fulfills the 3 rules necessary to apply induction on open sets.

(1). If $U \cong \mathbb{R}^n$ then $U \in \mathcal{B}$. We only have to prove that $\mathrm{PD}\colon H^0(\mathbb{R}^n) \to H_c^n(\mathbb{R}^n)^*$ is an isomorphism. Both spaces have dimension one, so it suffices to check that the map is not trivial: if η is a compactly supported n-form over $\mathbb{R}^n$ with $\int \eta = 1$ and 1 is the constant function we get $\langle 1, \eta \rangle = 1$ and hence $1 \in H^0(\mathbb{R}^n)$ is mapped to a nontrivial element $\mathrm{PD}(1) \in H_c^n(\mathbb{R}^n)^*$.

(2). If $U, V, U \cap V \in \mathcal{B}$, then $U \cup V \in \mathcal{B}$. To show this, we consider the following diagram that contains both Mayer – Vietoris sequences:

$$\begin{tikzcd}
{} \arrow[r] & H^{k-1}(U \cap V) \arrow[r, "\delta"] \arrow[d, "\mathrm{PD}"'] & H^k(U \cup V) \arrow[r] \arrow[d, "\mathrm{PD}"'] & H^k(U) \oplus H^k(V) \arrow[r] \arrow[d, "\mathrm{PD}"'] & {} \\
{} \arrow[r] & H_c^{n-k+1}(U \cap V)^* \arrow[r, "\delta^*"'] & H_c^{n-k}(U \cup V)^* \arrow[r] & H_c^{n-k}(U)^* \oplus H_c^{n-k}(V)^* \arrow[r] & {}
\end{tikzcd}$$

The bottom row is obtained by dualising the Mayer – Vietoris exact sequence in the compactly supported cohomology. This diagram commutes up to sign: the two simplest types of squares easily commute, while the more complicated one containing δ commutes only up to sign. Indeed if $[\omega] \in H^{k-1}(U \cap V)$ and ρ_U, ρ_V is a partition of unity subordinated to $\{U, V\}$, by Proposition 8.3.5

for every $[\alpha] \in H_c^{n-k}(U \cup V)$ we get

$$\mathrm{PD}\big(\delta([\omega])\big)([\alpha]) = \int_{U \cup V} d\rho_V \wedge \omega \wedge \alpha,$$

$$\delta^*\big(\mathrm{PD}([\omega])\big)([\alpha]) = \mathrm{PD}([\omega])(\delta([\alpha])) = \int_{U \cap V} \omega \wedge d\rho_V \wedge \alpha$$

and we get the same number up to a sign $(-1)^{k-1}$. We can modify the sign of some maps δ to get a genuine commutative diagram and then apply the Five Lemma, which says that if PD is an isomorphism for U, V, and $U \cap V$, then it is so also for $U \cup V$.

(3). If $U = \sqcup_i U_i$ and $U_i \in \mathcal{B}$, then $U \in \mathcal{B}$. This is a simple consequence of Proposition 8.4.8 and of the natural equality $(\oplus_i V_i)^* = \prod_i V_i^*$. We get

$$H^k(U) \longrightarrow \prod_i H^k(U_i) \stackrel{\prod_i \mathrm{PD}}{\longrightarrow} \prod_i H_c^{n-k}(U_i)^* \longrightarrow \big(\oplus_i H_c^{n-k}(U_i)\big)^* \longrightarrow H_c^{n-k}(U)^*$$

where all arrows are natural isomorphisms and their composition is PD.

By Proposition 8.5.3 we have $M \in \mathcal{B}$ and the proof is complete. □

8.5.5. Betti numbers. As a first consequence of Poicaré Duality, for every orientable manifold M without boundary we have

$$b^k(M) = b_c^{n-k}(M).$$

When M is also compact we get $b^k(M) = b^{n-k}(M)$. In particular we have $b^0(M) = b^n(M) = 1$ when M is compact and connected.

We can prove the finiteness of Betti numbers for any compact manifold.

Theorem 8.5.4. *Let M be a compact manifold, possibly with boundary and/or non orientable. The Betti numbers $b^k(M)$ are finite for all k.*

Proof. We consider first the case where M is orientable and without boundary. We have the canonical Poincaré isomorphisms

$$H^k(M) \cong H^{n-k}(M)^*, \qquad H^{n-k}(M) \cong H^k(M)^*.$$

By combining them we deduce that the canonical embedding $H^k(M) \hookrightarrow H^k(M)^{**}$ is an isomorphism, and we know that this holds if and only if the vector space is finite-dimensional. If M non-orientable and without boundary, it has an orientable double cover and we conclude using Proposition 8.4.13.

If M has boundary, we consider the double DM. We can cover DM with two open subsets U, V such that $U \cong V \cong \mathrm{int}(M) \sim M$ and $U \cap V \cong \partial M \times (0,1) \sim \partial M$. Here $\sim$ denotes homotopy equivalence. The Betti numbers of $M = U \cup V$ and $U \cap V$ are finite, and hence also those of U and V are, as one deduces easily from the Mayer–Vietoris sequence. □

In particular the Euler characteristic $\chi(M) = \sum_{i=0}^n b^i(M)$ of a compact manifold M, possibly with boundary, is always a well-defined finite number.

Corollary 8.5.5. *If M is a compact orientable manifold without boundary of odd dimension n, then $\chi(M) = 0$.*

Proof. We have $b^i(M) = b^{n-i}(M)$, so everything cancels out. $\square$

8.5.6. Orientability. We now show that the De Rham cohomology distinguishes between orientable and non-orientable manifolds. This works only for manifolds without boundary.

Proposition 8.5.6. *Let M be a connected smooth n-manifold without boundary. We have*

$$H^n_c(M) = \begin{cases} \mathbb{R} & \text{if } M \text{ is orientable,} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. If M is oriented, the map

$$\int_M : H^n_c(M) \longrightarrow \mathbb{R}$$

is an isomorphism: we know that $\int_M$ is surjective and $b^n_c(M) = b^0(M) = 1$.

If M is not orientable, it has an orientable double cover $\pi\colon \tilde{M} \to M$, with orientation-reversing deck involution $\iota\colon \tilde{M} \to \tilde{M}$. The induced map

$$\pi^*\colon H^n_c(M) \to H^n_c(\tilde{M})$$

is injective by Proposition 8.4.13. Moreover, for every n-form $\omega \in \Omega^n(M)$, the pull-back $\pi^*\omega$ is well-defined since π is proper, and is ι-invariant; since ι reverses the orientation of $\tilde{M}$ we get

$$\int_{\tilde{M}} \pi^*\omega = \int_{-\tilde{M}} \iota^*\pi^*\omega = -\int_{\tilde{M}} \pi^*\omega.$$

Hence this integral vanishes, and by what said above we get $[\pi^*\omega] = 0$ in cohomology. Since π^* is injective we get $H^n_c(M) = 0$. $\square$

We note that the identification $H^0(M) = \mathbb{R}$ is canonical, while $H^n_c(M) = \mathbb{R}$ depends on the chosen orientation for M (so it is canonical up to sign).

8.5.7. Real projective spaces. We can now easily calculate the De Rham cohomology of $\mathbb{RP}^n$.

Proposition 8.5.7. *We have*

$$H^k(\mathbb{RP}^n) = \begin{cases} \mathbb{R} & \text{if } k = 0 \text{ or } k = n \text{ is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. This works for every manifold M that is covered by S^n. Since the pull-back $\pi^*\colon H^k(M) \to H^k(S^n)$ is injective, the only indeterminacy is for $k = n$ and is determined by whether M is orientable or not. Projective spaces are orientable only in odd dimension. $\square$

We introduced the lens spaces $L(p,q)$ in Section 3.5.6. The same proof shows the following.

Corollary 8.5.8. *We have*

$$H^k(L(p,q)) = \begin{cases} \mathbb{R} & \text{if } k = 0,3, \\ 0 & \text{otherwise.} \end{cases}$$

8.5.8. Signature. If M is an oriented compact manifold without boundary of even dimension $2n$, Poincaré duality furnishes a non-degenerate bilinear form

$$H^n(M) \times H^n(M) \longrightarrow \mathbb{R}$$

that is symmetric or antisymmetric according to whether n is even or odd. This is because of the formula $\omega \wedge \eta = (-1)^{n^2} \eta \wedge \omega$.

When M has dimension $4k$, the non-degenerate bilinear form on H^{2k} is symmetric and hence has a *signature* (p,m), see Section 2.3.1. The *signature* of M is the integer

$$\sigma(M) = p - m.$$

A nice feature of this invariant is that it reacts to orientation reversals.

Proposition 8.5.9. *We have $\sigma(-M) = -\sigma(M)$*

Proof. We have $\int_M \omega = -\int_{-M} \omega$, hence the orientation reversal modifies the bilinear form by a sign and its signature changes from (p,m) to (m,p). $\square$

Recall that an orientable manifold M is *mirrorable* if it has an orientation-reversing diffeomorphism.

Corollary 8.5.10. *A mirrorable orientable $4k$-manifold M has $\sigma(M) = 0$.*

We deduce that for every $k \geq 1$ the complex projective space $\mathbb{CP}^{2k}$ is not mirrorable: its middle Betti number is $b^{2k} = 1$ and hence its signature is $\sigma = \pm 1$. In particular the complex projective plane $\mathbb{CP}^2$ is not mirrorable (while the complex projective line $\mathbb{CP}^1 \cong S^2$ is mirrorable).

8.5.9. The Künneth formula. We now prove an elegant formula that relates the cohomology of a product $M \times N$ with the cohomologies of the factors. This formula is known as the *Künneth formula*.

Let M and N be two smooth manifolds, possibly with boundary (not both). The two projections

$$\pi_M \colon M \times N \longrightarrow M, \qquad \pi_N \colon M \times N \longrightarrow N$$

give rise to a bilinear map

$$\begin{aligned} \Omega^k(M) \times \Omega^h(N) &\longrightarrow \Omega^{k+h}(M \times N) \\ (\omega, \eta) &\longmapsto \pi_M^* \omega \wedge \pi_N^* \eta \end{aligned}$$

that passes to a bilinear map

$$H^k(M) \times H^h(N) \longrightarrow H^{k+h}(M \times N).$$

By the universal property of tensor products, this induces a linear map

$$H^k(M)\otimes H^h(N)\longrightarrow H^{k+h}(M\times N).$$

These linear maps when k and h vary can be grouped altogether as

$$\Psi\colon H^*(M)\otimes H^*(N)\longrightarrow H^*(M\times N).$$

We will henceforth suppose that the Betti numbers of N are all finite: this holds for instance if N is compact, but also for many other manifolds.

Vedere se riusciamo a dimostrare che è un isomorfismo di algebre?

Theorem 8.5.11 (Künneth's formula). *The map Ψ is an isomorphism.*

Before entering into the proof, we note that this implies that

$$H^k(M\times N)\cong\bigoplus_{p+q=k}H^p(M)\otimes H^q(N).$$

Proof. As in the proof of Poincaré Duality, we define $\mathcal{B}$ to be the set of all the open subsets $U\subset M$ such that the theorem holds for the product $U\times N$. Our aim is to show that $M\in\mathcal{B}$ and to this purpose we verify the three rules that are necessary to apply the induction along open subsets.

(1). If $U\cong\mathbb{R}^n$, then $U\in\mathcal{B}$. The manifold $U\times N$ is homotopically equivalent to N and everything holds by homotopy invariance.

(2). If $U,V,U\cap V\in\mathcal{B}$, then $U\cup V\in\mathcal{B}$. To show this, we fix $k\geq 0$, pick $p\leq k$ and consider the Mayer – Vietoris sequence

$$\cdots\longrightarrow H^{p-1}(U\cap V)\longrightarrow H^p(U\cup V)\longrightarrow H^p(U)\oplus H^p(V)\longrightarrow\cdots$$

If we tensor it with $H^{k-p}(N)$ and sum over $p=0,\ldots,k$ we still get an exact sequence by Exercise 8.3.1. Here it is:

$$\cdots\longrightarrow\bigoplus_{p=0}^k\big(H^{p-1}(U\cap V)\otimes H^{k-p}(N)\big)\longrightarrow\bigoplus_{p=0}^k\big(H^p(U\cup V)\otimes H^{k-p}(N)\big)$$

$$\longrightarrow\bigoplus_{p=0}^k\big(H^p(U)\otimes H^{k-p}(N)\big)\bigoplus_{p=0}^k\big(H^p(V)\otimes H^{k-p}(N)\big)\longrightarrow\cdots$$

We now send via Ψ this sequence to the Mayer – Vietoris sequence for $M\times N$:

$$\cdots\to H^{k-1}\big((U\cap V)\times N\big)\to H^k\big((U\cup V)\times N\big)\to H^k(U\times N)\otimes H^k(V\times N)\to\cdots$$

The resulting diagram commutes (exercise) and has two exact rows. Using the Five Lemma we conclude that $U\cup V\in\mathcal{B}$.

(3). If $U=\sqcup_i U_i$ and $U_i\in\mathcal{B}$, then $U\in\mathcal{B}$. This is a consequence of Exercise 2.1.17 and of the fact that $\dim H^p(N)<\infty$ for all p.

By Lemma 8.5.3 we have $M\in\mathcal{B}$ and we are done. $\square$

Remark 8.5.12. When $M=N=\mathbb{Z}$, the map Ψ is not an isomorphism (exercise). We really need one of the factors to have finite-dimensional cohomology here.

Corollary 8.5.13. *Let M and N be manifolds with finite Betti number (for instance, this holds if they are compact). For every k we have:*

$$b^k(M\times N)=\sum_{i=0}^k b^i(M)b^{k-i}(N).$$

Corollary 8.5.14. *The torus $T=S^1\times S^1$ has Betti numbers*

$$b^0(T)=1,\qquad b^1(T)=2,\qquad b^2(T)=1.$$

Corollary 8.5.15. *The Betti numbers of $T^n=\underbrace{S^1\times\cdots\times S^1}_{n}$ are*

$$b^k(T^n)=\binom{n}{k}.$$

Corollary 8.5.16. *The Betti numbers of $S^2\times S^2$ are*

$$b^0=1,\quad b^1=0,\quad b^2=2,\quad b^3=0,\quad b^4=1.$$

We deduce that the compact four-manifolds

$$S^4,\qquad \mathbb{CP}^2,\qquad S^2\times S^2$$

are pairwise not homotopy equivalent (although they are all simply connected) because their Betti numbers are respectively

$$\begin{gathered}1\ 0\ 0\ 0\ 1,\\ 1\ 0\ 1\ 0\ 1,\\ 1\ 0\ 2\ 0\ 1.\end{gathered}$$

Exercise 8.5.17. If M and N are manifolds with finite Betti numbers, then

$$\chi(M\times N)=\chi(M)\cdot\chi(N).$$

8.5.10. Connected sums. The following exercises can be solved using the Mayer – Vietoris sequence carefully.

Exercise 8.5.18. Let M be a smooth connected n-manifold without boundary and N be obtained from M by removing a point. We have:

$$b^i(N)=b^i(M)\quad \forall i\leq n-2$$

$$b^{n-1}(N)=\begin{cases} b^{n-1}(M) & \text{if } M \text{ is compact and oriented,}\\ b^{n-1}(M)+1 & \text{otherwise,}\end{cases}$$

$$b^n(N)=\begin{cases} b^n(M)-1 & \text{if } M \text{ is compact and oriented,}\\ b^n(M) & \text{otherwise,}\end{cases}$$

Hint. Use the Mayer – Vietoris sequence with $M=U\cup V$, $U=N$, and V an open ball containing the removed point. □

Note that in all cases we get $\chi(N)=\chi(M)-1$ if they are defined.

Exercise 8.5.19. Let $M\#N$ be the connected sum of two oriented connected compact manifolds M and N without boundary. We have

$$b^i(M\#N) = 1 \quad \text{for } i = 0, n,$$
$$b^i(M\#N) = b^i(M) + b^i(N) \quad \text{for } 0 < i < n.$$

We can finally calculate the cohomology of a genus-g surface S_g.

Corollary 8.5.20. *The Betti numbers of S_g are*

$$b^0 = 1, \qquad b^1 = 2g, \qquad b^2 = 1.$$

Therefore $\chi(S_g) = 2 - 2g$.

8.6. Intersection theory

We now combine transversality and De Rham cohomology to build a geometric theory on submanifolds called *intersection theory*.

8.6.1. The Poincaré dual of a closed oriented submanifold. Let M be an oriented smooth n-manifold without boundary. Let $S \subset M$ be an oriented closed k-submanifold without boundary. We have already observed that integration along S yields a linear map

$$\int_S : H^k_c(M) \longrightarrow \mathbb{R}.$$

By Poincaré Duality, this linear map corresponds to some cohomology element $[S] \in H^{n-k}(M)$ called the *Poincaré dual* of S. A closed $(n-k)$-form ω_S that represents $[S]$ is characterised by satisfying the equality

$$\int_M \omega_S \wedge \eta = \int_S \eta$$

for every closed $\eta \in \Omega^k_c(M)$. We have just discovered that we can naturally transform oriented closed submanifolds S into cohomology classes $[S]$.

Exercise 8.6.1. Let M be connected, oriented and without boundary. Then:

- the Poincaré dual of M itself is $[M] = 1$ in $H^0(M) = \mathbb{R}$,
- the Poincaré dual of a point $p \in M$ lies in $H^n(M) = H^0_c(M)^*$, which is canonically isomorphic to $\mathbb{R}$ or 0 depending on whether M is compact or not. In the first case $[p] = 1$, in the second $[p] = 0$.

8.6.2. Homotopy invariance. The Poincaré dual may be defined for every smooth proper map $f : S \to M$ defined on some k-dimensional oriented manifold S without boundary. Every such map f induces a linear functional

$$H^k_c(M) \longrightarrow \mathbb{R}$$
$$[\eta] \longmapsto \int_S f^*\eta$$

which is by Poincaré Duality an element $[f] \in H^{n-k}(M)$.

Proposition 8.6.2. *If f, g are homotopic we have $[f] = [g]$.*

Proof. Let $F\colon S \times [0,1] \to M$ be a smooth homotopy relating $f = F_0$ and $g = F_1$. Set $W = S \times [0,1]$. For every $[\eta] \in H^k_c(M)$ we have

$$\int_S g^*\eta - \int_S f^*\eta = \int_{\partial W} F^*\eta = \int_W dF^*\eta = \int_W F^*(d\eta) = 0$$

by Stokes. □

Corollary 8.6.3. *Isotopic oriented submanifolds have equal Poincaré duals.*

8.6.3. Manifolds of finite type. We say that a manifold M is *of finite type* if the Betti numbers $b^i(M)$ of M are all finite. For instance, a compact manifold M is of finite type. The interior of a compact manifold with boundary is of finite type.

Let M be a finite type oriented smooth n-manifold without boundary. Since all the cohomology groups are finite, the Poincaré duality isomorphisms

$$H^k(M) \cong H^{n-k}_c(M)^*$$

also induce the dual isomorphisms

$$H^{n-k}_c(M) \cong H^k(M)^*$$

that send $[\eta]$ to the functional

$$[\omega] \longmapsto \langle[\omega], [\eta]\rangle = \int_M \omega \wedge \eta.$$

Let $S \subset M$ be a compact oriented k-submanifold without boundary. Then

$$\int_S \colon H^k(M) \longrightarrow \mathbb{R}$$

corresponds to some element $[S] \in H^{n-k}_c(M)$ that we call again the Poincaré dual of S. The novelty with respect to Section 8.6.1 is that $[S]$ may be represented as a *compactly supported* closed form η_S, and this is quite relevant for the discussion that follows. It is characterised by satisfying the equality

$$\int_M \omega \wedge \eta_S = \int_S \omega \tag{23}$$

for every closed $\omega \in \Omega^k(M)$. We now show that the Poincaré dual $[S]$ is natural with respect to inclusions. Let $U \subset M$ be any open set of finite type containing S. The inclusion $i\colon U \hookrightarrow M$ induces a map

$$i_*\colon H^{n-k}_c(U) \to H^{n-k}_c(M).$$

Proposition 8.6.4. *We have $i_*([S]) = [S]$.*

Proof. The Poincaré dual $[S] \in H^{n-k}_c(U)$ is represented by a compactly supported closed form $\eta_S \in \Omega^{m-k}_c(U)$ which satisfies

$$\int_U \omega \wedge \eta_S = \int_S \omega$$

for every closed $\omega \in \Omega^k(U)$. Therefore $i_*(\eta_S) \in \Omega_c^{m-k}(M)$ satisfies (23) for every $\omega \in \Omega^k(M)$ and hence represents $[S] \in H_c^{n-k}(M)$. $\square$

Concretely, this implies that we can shrink the support of a closed form η_S representing the Poincaré dual $[S]$ so that it is contained in an arbitrary open set U of finite type containing S; we may pick for instance a tubular neighbourhood $U = \nu S$, which is of finite type since it is homotopically equivalent to S.

By shrinking their compact support we have reduced the study of Poincaré duals to the case where the ambient manifold is a vector bundle over S.

8.6.4. Poincaré duals on vector bundles. Let $E \to S$ be an oriented rank-r vector bundle over a compact oriented k-manifold without boundary S. We think of S embedded in E as the zero-section. The Poincaré dual $[S]$ in the compactly supported version is an element in $H_c^r(E)$. Here is a simple characterisation of which closed forms represent $[S]$.

Controlla che sia spiegato bene prima cosa vuol dire che E è orientato.

Proposition 8.6.5. *A closed r-form $\eta \in \Omega_c^r(E)$ represents $[S]$ $\Longleftrightarrow$*

$$\int_{E_p} \eta = 1$$

for every $p \in S$.

Before discussing the proof, we note that if S is connected the integral $\int_{E_p} \eta$ of any closed $\eta \in \Omega_c^r(E)$ yields a real number that is independent of p, because two distinct fibers are clearly isotopic.

Proof. ($\Longleftarrow$) We must prove that

$$\int_E \omega \wedge \eta = \int_S \omega \tag{24}$$

for every closed $\omega \in \Omega^k(E)$. The map $i \circ \pi \colon E \to E$ is homotopic to the identity, hence in cohomology we get $[\omega] = (i \circ \pi)^*[\omega]$ and therefore

$$\int_E \omega \wedge \eta = \int_E \pi^* i^* \omega \wedge \eta.$$

On a trivialising chart the bundle is like $\mathbb{R}^k \times \mathbb{R}^r \to \mathbb{R}^k$. We use the variables x^i and y^j for $\mathbb{R}^k$ and $\mathbb{R}^r$. The restriction of ω to $\mathbb{R}^k \times \mathbb{R}^r$ may be written as

$$\omega = \sum_{I,J} f^{I,J}(x,y) dx^I \wedge dy^J$$

where the sum is on all pairs I, J of multi-indices with $|I| + |J| = k$. Since $i^*(dy^j) = 0$ we get

$$i^*\omega = f(x) dx^1 \wedge \cdots \wedge dx^k$$

with $f(x) = f^{\{1,\ldots,k\},\varnothing}(x,0)$, and $\pi^* i^* \omega$ has the same expression as $i^*\omega$. On the other hand, we have

$$\eta = \sum_{I,J} g^{I,J}(x,y) dx^I \wedge dy^J$$

where the sum ranges over muti-indices with $|I| + |J| = r$. By hypothesis

$$1 = \int_{\{x\}\times\mathbb{R}^r} \eta = \int_{\mathbb{R}^r} g(x,y)dy$$

for all x, where $g = g^{\varnothing,\{1,\ldots,r\}}$. We deduce that

$$\begin{aligned}\int_{\mathbb{R}^k\times\mathbb{R}^r} \pi^* i^*\omega \wedge \eta &= \int_{\mathbb{R}^k\times\mathbb{R}^r} f(x)dx^1\wedge\cdots\wedge dx^k \wedge \sum_{I,J} g^{I,J}(x,y)dx^I \wedge dy^J \\ &= \int_{\mathbb{R}^k\times\mathbb{R}^r} f(x)g(x,y)dx^1\wedge\cdots\wedge dx^k\wedge dy^1\wedge\cdots\wedge dy^r \\ &= \int_{\mathbb{R}^k} f(x)\left(\int_{\mathbb{R}^r} g(x,y)dy\right)dx = \int_{\mathbb{R}^k} f(x) = \int_{\mathbb{R}^k}\omega.\end{aligned}$$

($\Longrightarrow$) Suppose that S is connected for simplicity. We have $\int_{E_p}\eta = k$ independently on p. Pick a volume form ω_0 on S and set $\omega = \pi^*\omega_0$. We have

$$\int_S \omega = \int_E \omega\wedge\eta = k\int_S\omega.$$

The second equality is proved on trivialising charts as above. Hence $k = 1$. □

8.6.5. Transverse intersection. Let N be an oriented n-manifold without boundary, and let $M, W \subset N$ be two oriented compact transverse submanifolds without boundary of dimension $m, w < n$. Recall that

$$X = M \pitchfork W$$

is also a compact submanifold of dimension $x = m + w - n$. We now show that X inherits an orientation from those of M, W, N.

If $m+w = n$, then X consists of finitely many points, and we assign to each point $p \in X$ the sign $+1 \iff$ two positive basis $u_1,\ldots,u_m$ and $u_1',\ldots,u_w'$ of T_pM and T_pW combine to a positive basis $u_1,\ldots,u_m,u_1',\ldots,u_w'$ of T_pN.

If $m + w > n$ we proceed similarly. At $p \in X$, a basis $v_1,\ldots,v_x$ of T_pM is positive if it may be completed to two positive basis $v_1,\ldots,v_x,u_{x+1},\ldots,u_m$ and $v_1,\ldots,v_x u_{x+1}',\ldots,u_w'$ of T_pU and T_pW such that the resulting basis $v_1,\ldots,v_x,u_{x+1},\ldots,u_m,u_{x+1}',\ldots,u_w'$ of T_pN is positive.

Notazione coerente $-M$?

Exercise 8.6.6. We have

$$M \pitchfork W = (-1)^{mw} W \pitchfork M.$$

The same type of equality $\omega\wedge\eta = (-1)^{kh}\eta\wedge\omega$ holds for any k- and h-forms ω and η on a manifold, and this is not a coincidence! The following theorem, which is the heart of intersection theory, shows that, via Poincaré duality, the transverse intersection of oriented submanifolds corresponds to the wedge products of cohomology classes.

Theorem 8.6.7. *We have*

$$[M \pitchfork W] = [M]\wedge[W].$$

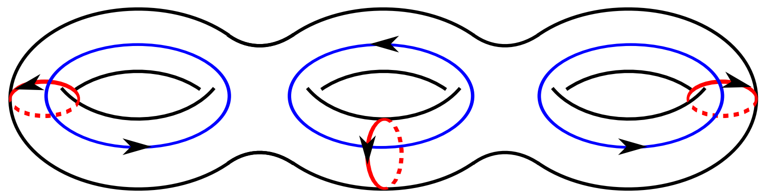

Figure 8.1. A symplectic basis for $H^1(S_3) \cong \mathbb{R}^6$ consists of the Poincaré duals of the oriented curves $\alpha_1, \alpha_2, \alpha_3$ (red) and $\beta_1, \beta_2, \beta_3$ (blue).

Proof. We have

$$\nu X = \nu M \oplus \nu W.$$

If ω_M, ω_W represent the Poincaré duals $[M], [W]$ in the tubular neirhbourhoods $\nu M, \nu W$, the wedge product $\omega_M \wedge \omega_W$ represents $[X]$ in $\nu X = \nu M \oplus \nu W$. $\square$

Example 8.6.8. Each hyperplane $H \subset \mathbb{CP}^n$ is naturally oriented. Since hyperplanes are all isotopic (exercise), the Poincaré dual $[H] \in H^2(\mathbb{CP}^n)$ is independent of the choice of H. Every codimension-k projective subspace $S \subset \mathbb{CP}^n$ is the transverse intersection of k hyperplanes, and hence $[S] = [H] \wedge \cdots \wedge [H] = [H]^k$. In particular $[H]^n \in H^{2n}(\mathbb{CP}^n)$ is the Poincaré dual of a point, which is a generator of $H^{2n}(\mathbb{CP}^n)$.

We deduce that the algebra $H^*(\mathbb{CP}^n)$ is isomorphic to

$$H^*(\mathbb{CP}^n) \cong \mathbb{R}[x]/(x^{n+1})$$

where $x = [H] \in H^2(\mathbb{CP}^n)$.

8.6.6. Algebraic intersection. We now consider the case where M and W have complementary dimension. In this situation $X = M \pitchfork W$ is a finite collection of oriented points p, each equipped with a sign ± 1 depending on whether the orientation of $T_pM \oplus T_pW$ matches with that of T_pN. We let the *algebraic intersection* $i(M, W)$ of M and W be the sum of these values ± 1.

Let us now suppose for simplicity that N is connected. Then

$$[M] \wedge [W] \in H^n(N) = \mathbb{R}$$

is a canonical real number, which coincides by Theorem 8.6.7 with the algebraic intersection $i(M, W)$. Here is an immediate consequence.

Corollary 8.6.9. *The algebraic intersection $i(M, W)$ depends only on the isotopy classes of M and W.*

When N has even dimension $2k$ and $\dim M = \dim W = k$, this furnishes a concrete way to represent and calculate the intersection form in $H^k(N)$.

Example 8.6.10. We examine the genus-g surface S_g. The intersection form on $H^1(S_g) \cong \mathbb{R}^{2g}$ is non-degenerate and antisymmetric. Consider the $2g$

oriented curves $\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g$ depicted in Figure 8.1. Their algebraic intersections are apparent from the figure:

$$i(\alpha_i, \alpha_j) = i(\beta_i, \beta_j) = 0, \qquad i(\alpha_i, \beta_j) = \delta_{ij}, \qquad i(\beta_i, \alpha_j) = -\delta_{ij}.$$

The intersection form of their respective Thom classes is the antisymmetric matrix $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$. Since J is invertible, we can deduce by elementary linear algebra that these $2g$ classes form a basis of $H^1(S_g)$. A basis with such an intersection matrix J is called a *symplectic basis*.

Example 8.6.11. We calculate the intersection form of some compact oriented 4-manifolds M by determining some surfaces in M whose Poincaré dual form a basis of $H^2(M)$.

If we pick two distinct projective lines $l, l' \subset \mathbb{CP}^2$, we have $[l] = [l']$ and l, l' intersect transversely and positively in a single point. The intersection form on $H^2(\mathbb{CP}^2)$ with respect to the basis $[l]$ is hence just the matrix

$$(1).$$

In $S^2 \times S^2$, we pick the spheres $S = S^2 \times \{q\}$ and $S' = \{p\} \times S^2$. They intersect transversely and positively in one point (p, q), hence $[S] \wedge [S'] = 1$. Moreover each sphere $S^2 \times \{q\}$ is disjoint from the isotopic copy $S^2 \times \{q'\}$, hence $[S] \wedge [S] = 0$ and analogously $[S'] \wedge [S'] = 0$. Summing up, the classes $[S], [S']$ form a basis for $H^2(S^2 \times S^2)$ and the intersection form with respect to this basis is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We deduce that $S^2 \times S^2$ has signature zero (which is coherent since $S^2 \times S^2$ is mirrorable).

8.7. Exercises

Exercise 8.7.1. Calculate the Betti numbers of the manifold M obtained from $\mathbb{R}^3$ by removing the x and y axis.

Exercise 8.7.2. Prove that the surface $\mathbb{C} \setminus \mathbb{Z}$ has $b^1 = \infty$.

Exercise 8.7.3. Let $K \subset S^3$ be a knot. Prove that $H^1(S^3 \setminus K) \cong \mathbb{R}$.

Exercise 8.7.4. Let M and N be compact manifolds with boundary and $\varphi\colon \partial M \to \partial N$ a diffeomorphism. Let W be obtained by glueing M and N via φ. Show that

$$\chi(W) = \chi(M) + \chi(N) - \chi(\partial M).$$

Exercise 8.7.5. Let $T = S^1 \times S^1$ be a torus and $p \in T$ a point. Consider the 4-manifold $M = T \times T$ and the submanifolds $N_1 = T \times \{p\}$ and $N_2 = \{p\} \times T$. Calcolate the cohomology groups of the 4-manifold

$$X = M \setminus (N_1 \cup N_2).$$

Exercise 8.7.6. Let L, L' be two affine subspaces of $\mathbb{R}^n$.

(1) Show that the manifolds $\mathbb{R}^n \setminus L$ and $\mathbb{R}^n \setminus L'$ are homotopically equivalent if and only if $\dim L = \dim L'$.
(2) Show that if $\dim L > \dim L'$ every continuous map $f : (\mathbb{R}^n \setminus L) \to (\mathbb{R}^n \setminus L')$ is homotopic to a constant.

Exercise 8.7.7. Let r_1, r_2, r_3 be three lines in $\mathbb{CP}^2$ with empty intersection $r_1 \cap r_2 \cap r_3 = \varnothing$.
(1) Calculate the cohomology groups of the smooth manifold $X = \mathbb{CP}^2 \setminus (r_1 \cup r_2 \cup r_3)$.
(2) Show that there is a map $f : X \to X$ such that $f^* : H^*(X) \to H^*(X)$ is neither the identity nor trivial.

Exercise 8.7.8. Let $\pi : E^{n+k} \to M^n$ be a fibre bundle with fibre F^k. Suppose that M, E, F are all compact, orientable, and without boundary. Show that if $s : M \to E$ is a section, then the Poincaré dual of $s(M)$ is a non-trivial element in $H^k(E)$. Deduce that the Hopf fibration $S^3 \to S^2$ has no sections.

Hint. Use the relation between $\wedge$ and transverse intersection. $\square$

Part 3

Differential geometry

CHAPTER 9

Pseudo-Riemannian manifolds

We have warned the reader multiple times that a smooth manifold M lacks many natural geometric notions, such as distance between points, length of curves, volumes, angles, geodesics. It is now due time to introduce all these concepts, by enriching M with an additional structure g, called *metric tensor*.

A metric tensor g on M is just a smoothly varying scalar product on all tangent spaces. If g is positive definite the pair (M, g) is called a *Riemannian manifold*. If positive definiteness is not assumed, the pair is called more generally a *pseudo-Riemannian manifold*. Riemannian manifolds are fundamental concepts in mathematics, while the theory of the more general pseudo-Riemannian manifolds plays a key role in general relativity.

9.1. The metric tensor

It is a quite remarkable fact that all the various natural geometric notions that we are longing for can be introduced by equipping a smooth manifold with a single additional structure, that of a *metric tensor*.

9.1.1. Pseudo-Riemannian manifolds. Let M be a smooth manifold, possibly with boundary. Recall from Section 7.5.1 that a metric tensor g is a section of the symmetric bundle that defines a scalar product $g(p)$ on T_pM, for every $p \in M$.

Definition 9.1.1. A *pseudo-Riemannian manifold* is a pair (M, g) where M is a smooth manifold and g is a metric tensor on M.

If M is connected the scalar product g has the same signature (p, q) at every point of M and we simply call it the *signature* of g. Of course $p + q = n = \dim M$. Two types of signatures are particularly important in mathematics and in physics: if g is positive definite, that is it has signature $(n, 0)$, we say that (M, g) is a *Riemannian manifold*; if the signature is $(n - 1, 1)$, we say that (M, g) is a *Lorentzian manifold*.

The reader may wonder why we are allowing non positive definite scalar products. The reason is twofold. First, pseudo-Riemannian manifolds play a fundamental role in general relativity: as we will see, the universe is modeled as a Lorentzian manifold with signature $(3, 1)$. Second, perhaps quite surprisingly, the positive definite hypothesis is not really needed to introduce most of the powerful and beautiful tools in Riemannian geometry, like *connections, geodesics, covariant derivatives*, and *curvature*.

Example 9.1.2. A fundamental example of positive definite metric tensor is the Euclidean metric tensor g_E on $\mathbb{R}^n$. The pair $(\mathbb{R}^n, g_E)$ is a Riemannian manifold called the *Euclidean space*.

Example 9.1.3. The Minkowski space $(\mathbb{R}^4, \eta)$ introduced in Section 7.6.1 is a Lorentzian manifold, denoted as $\mathbb{R}^{3,1}$. More generally, for every $p+q=n$ we may define a pseudo-Riemannian structure on $\mathbb{R}^n$ by assigning at every $x \in \mathbb{R}^n$ the metric tensor

$$\eta_{p,q} = \begin{pmatrix} -I_q & 0 \\ 0 & I_p \end{pmatrix}$$

of signature (p, q). We indicate this pseudo-Riemannian manifold as $\mathbb{R}^{p,q}$.

Remark 9.1.4. We have shown in Section 4.5 that every bundle carries a Riemannian metric. Therefore every smooth manifold M has a positive definite metric tensor, that is a structure of Riemannian manifold. The metric tensor is however not unique in any reasonable sense.

Note that the proof of Proposition 4.5.2 does not apply to metrics with any signature (p, q), since these do not form a convex cone! A convex combination of some matrices with signature (p, q) may not have signature (p, q). We cannot guarantee the existence of Lorentzian structures on any M. In fact, as we will see, there are manifolds that do not admit any Lorentzian structure. TBD

If (M, g) is a pseudo-Riemannian manifold, every open subset $U \subset M$ inherits a structure of pseudo-Riemannian manifold, just by restricting g.

9.1.2. In coordinates. Let (M, g) be a Riemannian manifold and $\varphi\colon U \to V$ a chart. The tensor g on U may be transported along φ into a metric tensor $\varphi_* g$ on V, whose coordinates are denoted by

$$g_{ij}(x).$$

Here $g_{ij}(x)$ is a non-degenerate symmetric matrix that depends smoothly on x. For instance, the coordinates of the Euclidean metric tensor are $g_{ij} = \delta_{ij}$.

9.1.3. Vector types, vector lengths, and angles. Let (M, g) denote a pseudo-Riemannian manifold. The tangent space T_pM is equipped with a scalar product at every point $p \in M$. Given two tangent vectors $v, w \in T_pM$, we often write their scalar product $g(p)(v, w)$ simply as $\langle v, w\rangle$, omitting p.

As in Section 7.6, a vector $v \in T_pM$ is called *spacelike*, *timelike*, or *lightlike* if $\langle v, v\rangle$ is (respectively) positive, negative, or null. In all cases, its *length* is

$$\|v\| = \sqrt{|\langle v, v\rangle|}.$$

In particular the length of v is zero $\iff$ v is lightlike. If the scalar product is positive definite, we get a norm $\|\cdot\|$ on T_pM, and we can also define the *angle* θ between two non-zero vectors $v, w \in T_pM$ with the usual formula

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

Angles are not defined if the scalar product is not positive definite.

9.1.4. Conformal modifications. Let (M, g) be a pseudo-Riemannian manifold. There are some very simple ways to modify the metric tensor g.

The simplest possible modification one can make consists of fixing a non-zero scalar $\lambda \in \mathbb{R}$ and multiplying the tensor $g(p)$ by λ at every $p \in M$, thus getting a new metric tensor $g'(p) = \lambda g(p)$. This modification is called a *metric rescaling*. If $\lambda > 0$, this corresponds intuitively to inflating or deflating our manifold depending on whether $\lambda > 1$ or $\lambda < 1$. This modification changes the geometry of the manifold only very mildly. If $\lambda < 0$, the signature of the metric tensor changes from (p, q) to (q, p).

More generally, we may allow the scalar λ to vary smoothly from point to point. If we pick a positive smooth function $\lambda\colon M \to (0, +\infty)$, we may replace g with a new metric tensor $g' = \lambda g$. At every point we have $g'(p) = \lambda(p)g(p)$. This modification does not alter the signature of the metric and is called a *conformal modification*. Two metrics g and g' related by a conformal modification are called *conformally equivalent*: this is an equivalence relation on the set of metrics on M with any fixed signature.

Unlike rescalings, conformal modifications alter much of the geometry of the manifold, as we will see. They are characterised, in the positive definite case, by the fact that they preserve angles:

Proposition 9.1.5. *Two positive definite metric tensors g and g' on M are conformally equivalent $\iff$ they measure the same angles. That is, for every $p \in M$ and $v, w \in T_pM$, the angle between v and w is the same for g and g'.*

Proof. If $g'(p) = \lambda(p)g(p)$, the two scalar products on T_pM differ only by a rescaling and hence measure the same angles. Conversely, it is a linear algebra exercise to show that if $g(p)$ and $g'(p)$ measure the same angles then there is a $\lambda(p) \neq 0$ such that $g(p) = \lambda(p)g'(p)$. $\square$

The lengths of any $v \in T_pM$ with respect to g and $g' = \lambda g$ are related as

$$\|v\|_{g'} = \sqrt{\lambda(p)}\|v\|_g.$$

Example 9.1.6. We can pick an open subset $U \subset \mathbb{R}^n$, a positive function $\lambda\colon U \to (0, +\infty)$, and define a new Riemannian manifold $(U, \lambda g_E)$, that is conformally equivalent to the original Euclidean (U, g_E). This conformal modification rescales the tangent vectors by $\sqrt{\lambda}$, preserving the angles between them: this quite useful feature sometimes helps to visualize part of the geometry of $(U, \lambda g_E)$.

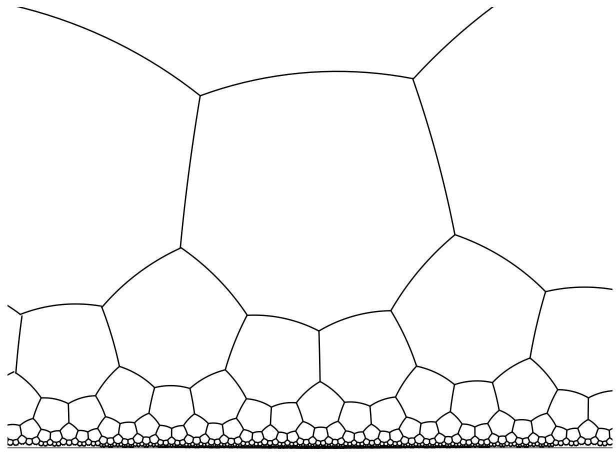

Figure 9.1. A tessellation of the hyperbolic plane H^2 into heptagons. All the heptagons shown here are isometric. We will soon see that their sides are "straight", that is geodesic, although they do not look so at our Euclidean eyes.

9.1.5. Hyperbolic space. We introduce an important Riemannian manifold. The *hyperbolic space*, described via the *half-space model*, is the manifold

$$H^n = \{x \in \mathbb{R}^n \mid x^n > 0\}$$

equipped with the metric tensor

$$g = \frac{1}{(x^n)^2} g_E.$$

This metric tensor is obtained from the Euclidean g_E by conformal modification. Angles are not changed, but all vectors v based at a point x are stretched by a factor $1/x^n$. Note that $1/x^n \to \infty$ as $x^n \to 0$.

The effect of this stretching can be seen in Figure 9.1, which shows a tessellation of the hyperbolic plane into isometric heptagons. All the heptagons actually have the same size, although those closer to the horizontal line $x^2 = 0$ look smaller at our Euclidean eyes. On the other hand, the interior angles that we see at the vertices are correct: they are $2\pi/3$ everywhere.

The *ball model* for the hyperbolic space is the manifold

$$B^n = \left\{x \in \mathbb{R}^n \mid \|x\| < 1\right\}$$

equipped with the metric tensor

$$g = \left(\frac{2}{1 - \|x\|^2}\right)^2 g_E.$$

Also here we perform a conformal modification of the Euclidean tensor and thus angles are not changed, but all vectors v based at $x \in B^n$ are stretched by a factor $2/(1 - \|x\|^2)$. Again note that the factor tends to infinity as $\|x\| \to 1$.

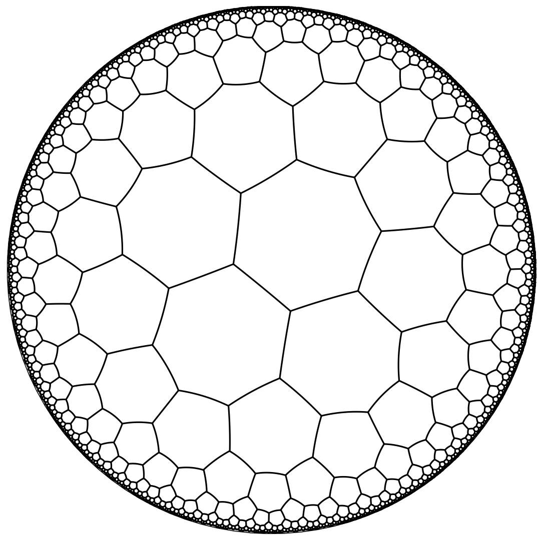

Figure 9.2. A tessellation of the hyperbolic plane B^2 into heptagons. All the heptagons shown here are isometric.

Figure 9.2 shows a tessellation into infinitely many heptagons. Although those closed to $\partial B^2 = S^1$ look smaller, they are actually all isometric.

We have in fact described twice the same space: the two Riemannian manifolds H^n and B^n are *isometric*, a fundamental notion that we now introduce.

9.1.6. Isometries. Every category has its own morphisms; in the presence of pseudo-Riemannian metrics, one typically introduces only isomorphisms.

Definition 9.1.7. A diffeomorphism $\varphi\colon M \to N$ between two pseudo-Riemannian manifolds (M, g) and (N, h) is an *isometry* if

$$\langle v, w \rangle = \big\langle d\varphi_p(v), d\varphi_p(w) \big\rangle$$

for every $p \in M$ and $v, w \in T_pM$.

The reader should be aware that the same symbol $\langle , \rangle$ may denote scalar products on different spaces: in the definition these are $g(p)$ and $h(\varphi(p))$.

Two pseudo-Riemannian manifolds M and N are *isometric* if there is an isometry relating them. Inverses and compositions of isometries are isometries. The isometries $M \to M$ of a pseudo-Riemannian manifold M form a group denoted with $\mathrm{Isom}(M)$ and called the *isometry group* of M.

Exercise 9.1.8. For any matrix $A \in \mathrm{O}(n)$ and any vector $b \in \mathbb{R}^n$, the affine transformation $f(x) = Ax + b$ is an isometry of the Euclidean space $\mathbb{R}^n$.

We will soon prove that, conversely, every isometry of the Euclidean space is of the kind described in the exercise.

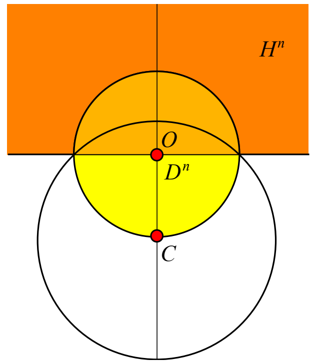

Figure 9.3. The inversion along the sphere with center $C = -e_n = (0, \dots, 0, -1)$ and radius $\sqrt{2}$ is an isometry between the ball and the half-space models of hyperbolic space.

A smooth map $f \colon M \to N$ is a *local isometry at* $p \in M$ if there are open neighbourhoods U and V of p and $f(p)$ such that $f(U) = V$ and $f|_U \colon U \to V$ is an isometry. The map f is a *local isometry* if it is so $\forall p \in M$.

9.1.7. Sphere inversions. We have defined two models for the hyperbolic space and we now prove that they are indeed isometric. We need the following.

Definition 9.1.9. Given $x_0 \in \mathbb{R}^n$ and $r > 0$, consider the sphere

$$S = \left\{ x \in \mathbb{R}^n \;\middle|\; \|x - x_0\| = r \right\}$$

centered at x_0 and with radius r. The *inversion* along S is the map $\varphi \colon \mathbb{R}^n \setminus \{x_0\} \to \mathbb{R}^n \setminus \{x_0\}$ defined as

$$\varphi(x) = x_0 + r^2 \frac{x - x_0}{\|x - x_0\|^2}.$$

Inversions along spheres have many nice properties and should be interpreted as the analogue of reflections along hyperplanes. In the following B^n and H^n denote the ball and half-disc models of hyperbolic space, each equipped with its metric tensor. We need inversions here to get the following.

Exercise 9.1.10. The inversion φ along the sphere with center $-e_n$ and radius $\sqrt{2}$ sends B^n diffeomorphically onto H^n. See Figure 9.3. It is an isometry between the ball and the half-space models of the hyperbolic space.

We denote the n-dimensional hyperbolic space by $\mathbb{H}^n$. This important Riemannian manifold may be represented by one of its isometric models B^n or H^n. Actually, we will discover a third model soon...

Exercise 9.1.11. The following diffeomorphisms are isometries of H^n, B^n:

- The map $\varphi \colon B^n \to B^n$, $\varphi(x) = Ax$ for any $A \in \mathrm{O}(n)$.
- The map $\varphi \colon H^n \to H^n$, $\varphi(x) = \lambda x$ for any $\lambda > 0$.
- The map $\varphi \colon H^n \to H^n$, $\varphi(x) = x + b$, for any $b \in \mathbb{R}^n$ with $b^n = 0$.

9.1.8. Submanifolds. Let (M, g) be a Riemannian manifold. Here is a simple albeit crucial observation: every submanifold $N \subset M$, of any dimension, inherits a positive definite metric tensor $g|_N$ simply by restricting g to the subspace $T_pN \subset T_pM$ at every $p \in N$. Therefore every smooth submanifold of a Riemannian manifold is itself naturally a Riemannian manifold.

In particular, every submanifold $S \subset \mathbb{R}^n$ inherits a Riemannian manifold structure by restricting g_E to S. Using Whitney's Embedding Theorem, we find here another proof that every manifold M carries a Riemannian structure.

A fundamental example is of course the sphere $S^{n-1} \subset \mathbb{R}^n$.

Exercise 9.1.12. For every $A \in \mathrm{O}(n)$, the map $\varphi(x) = Ax$ restricts to an isometry $\varphi\colon S^{n-1} \to S^{n-1}$.

If M is a more general pseudo-Riemannian manifold, it is *not* true that any submanifold $N \subset M$ inherits a pseudo-Riemannian structure! To get this, we need the restriction of g to T_pN to be non-degenerate $\forall p \in N$. If this is the case, we say that N is a *pseudo-Riemannian submanifold* of M.

Exercise 9.1.13. Consider the Minkowski space $\mathbb{R}^{n,1}$ with its constant metric tensor $\langle x, y\rangle = -x^1y^1 + x^2y^2 + \cdots x^ny^n$. The upper sheet of the hyperboloid

$$I^n = \big\{\langle x, x\rangle = -1,\ x^1 > 0\big\}$$

is a smooth submanifold. The tangent space at $p \in I^n$ is

$$T_pI^n = p^\perp = \{x \in \mathbb{R}^{n,1} \mid \langle x, p\rangle = 0\}.$$

(This is completely analogous to S^n, see Exercise 3.7.4.) In particular all the tangent vectors in T_pI^n are spacelike: hence the restriction of $\langle , \rangle$ to T_pI^n is positive definite, and I^n inherits a structure of Riemannian submanifold, inside the Lorentzian manifold $\mathbb{R}^{n,1}$.

The Riemannian manifold I^n is yet another model for the n-dimensional hyperbolic space $\mathbb{H}^n$!

Exercise 9.1.14. Consider the ball model $B^n \subset \mathbb{R}^n$ of hyperbolic space, embedded in $\mathbb{R}^{n,1}$ by sending $(x^1, \ldots, x^n)$ to $(0, x^1, \ldots, x^n)$. Construct a diffeomorphism $\varphi\colon I^n \to B^n$ by projecting along lines passing through $P = (-1, 0, \ldots, 0)$ as in Figure 9.4. Show that φ is an isometry.

The Riemannian manifold I^n is called the *hyperboloid model* for the hyperbolic space. We have discovered as many as three models B^n, H^n, and I^n for the hyperbolic space $\mathbb{H}^n$. None of them is prevalent: one can use the model that she prefers according to the problem she has to solve from case to case. The first two models are easier to visualize, the third one is harder to see but has better algebraic properties.

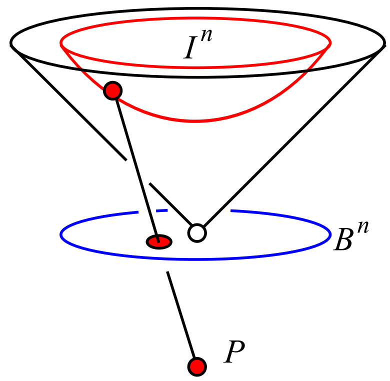

Figure 9.4. By projecting along $P = (-1, 0, \ldots, 0)$ we get an isometry between the two models I^n and B^n of the hyperbolic space. The metric tensor of I^n is the restriction of the Minkowksi constant metric tensor. The metric tensor of B^n is the conformal tensor $(2/(1-\|x\|^2))^2 g_E$.

As in Section 7.6.2, we denote by $\mathrm{O}(n,1)$ the group of matrices that preserve the Minkowski scalar product, and by $\mathrm{O}^+(n,1)$ the index-two subgroup consisting of those that preserve the time orientation. The following is analogous to Exercise 9.1.12.

Exercise 9.1.15. For every $A \in \mathrm{O}^+(n,1)$, the map $\varphi(x) = Ax$ restricts to an isometry $\varphi \colon I^n \to I^n$.

We have discovered that the hyperbolic space $\mathbb{H}^n$ has plenty of isometries, much as the Euclidean space $\mathbb{R}^n$ and the sphere S^n. We will see in the next pages that the Riemannian manifolds $\mathbb{H}^n$, $\mathbb{R}^n$ and S^n are the most symmetric and (for many reasons) important Riemannian manifolds in dimension n.

9.1.9. Products. The product $M \times N$ of two pseudo-Riemannian manifolds (M,g) and (N,h) carries a natural pseudo-Riemannian structure $g \times h$. Recall that $T_{(p,q)} M \times N = T_p M \times T_q N$ and define

$$\big\langle (v_1, w_1), (v_2, w_2) \big\rangle = \langle v_1, v_2 \rangle + \langle w_1, w_2 \rangle$$

for every $v_1, v_2 \in T_p M$ and $w_1, w_2 \in T_q N$. The signature of the product is the sum of the signatures of the factors, so if both M and N are Riemannian then $M \times N$ also is.

Example 9.1.16. The torus $T = S^1 \times S^1$ with the product metric is the *flat torus*. It is important to note that the flat torus is *not* isometric to the torus of Figure 3.4. The first is flat, but the second is not: we will introduce the notion of *curvature* to explain that.

9.1.10. Length of curves. As we promised, we now start to show how the metric tensor alone generates a wealth of fundamental geometric concepts. We start by defining the lengths of smooth curves.

Let $\gamma\colon I \to M$ be a smooth curve in a pseudo-Riemannian manifold M. We define its *length* as

$$L(\gamma) = \int_I \|\gamma'(t)\| dt.$$

Recall that the norm of a vector $v \in T_pM$ is $\|v\| = \sqrt{|\langle v, v\rangle|}$. A *reparametrisation* of the curve γ is obtained by picking an interval diffeomorphism $\varphi\colon J \to I$ and setting $\eta = \gamma \circ \varphi$.

Proposition 9.1.17. *The length of γ is independent of the parametrisation.*

Proof. We have

$$L(\gamma) = \int_I \|\gamma'(t)\| dt = \int_J \big\|\gamma'\big(\varphi(u)\big)\big\| |\varphi'(u)| du = \int_J \|\eta'(u)\| du = L(\eta).$$

The proof is complete. $\square$

More generally, the length $L(\gamma)$ is also invariant if we pre-compose γ with a smooth surjective monotone map $\varphi\colon J \to I$, that is with $\varphi'(t) \geq 0$ everywhere (or $\varphi'(t) \leq 0$ everywhere). With some abuse of language we also call this change of variables a *reparametrisation*.

A curve γ is *spacelike*, *timelike*, or *lightlike* if $\gamma'(t)$ is spacelike, timelike, or lightlike for every $t \in I$. We note that γ is lightlike precisely when $L(\gamma) = 0$.

On a Riemannian manifold M we call $\|\gamma'(t)\|$ the *speed* of the curve γ at the time t. In this context, a curve γ is immersed $\Longleftrightarrow$ it has positive speed at every time t.

9.1.11. Metric space. A connected Riemannian manifold (M, g) is also a metric space, with the following distance: for every $p, q \in M$ we define $d(p, q)$ as the infimum of the lengths of all the paths connecting p to q, that is

$$d(p, q) = \inf\big\{L(\gamma) \bigm| \gamma\colon [a, b] \to M,\ \gamma(a) = p,\ \gamma(b) = q\big\}.$$

Proposition 9.1.18. *This is a distance compatible with the topology of M.*

Proof. We clearly have $d(p, p) = 0$. We now prove that $p \neq q \Rightarrow d(p, q) > 0$. Pick a small open chart $\varphi\colon U \to V$ with $p \in U$, $\varphi(p) = 0$, and $q \notin U$. Choose a disc $D \subset V$ of some small radius r centred at the origin. The transported metric tensor on D is some g_{ij} depending smoothly on $x \in D$.

For every $x \in D$ and $v \in T_x\mathbb{R}^n$, we indicate with $\|v\|_E$ and $\|v\|_g$ the Euclidean and g-norm of v. Since D is compact, there are $M > m > 0$ with

$$m\|v\|_E < \|v\|_g < M\|v\|_E$$

for every $x \in D$ and every $v \in T_x\mathbb{R}^n$. Let α be a curve in V that goes from 0 to some point in ∂D. We know that the Euclidean length of α is $\geq r$, and we deduce that the g-length of α is $> rm$. Since every curve γ connecting p and q must cross $\varphi^{-1}(\partial D)$, we deduce that $L(\gamma) \geq rm$ and hence $d(p, q) \geq rm$.

We clearly have $d(p, q) = d(q, p)$. To show transitivity, we note that if γ is a curve from p to q and η is a curve from q to r, we can concatenate γ

and η to a *smooth* curve from p to r: to get smoothness it suffices to priorly reparametrise γ and η using transition functions.

In our discussion we have also shown that for every neighbourhood U of p there is an $\varepsilon > 0$ such that the d-ball of radius ε is entirely contained in U. Conversely, it is also clear that an open d-ball is open in the topology of M. Therefore d is compatible with the topology of M. $\square$

Remark 9.1.19. The infimum defining $d(p, q)$ may not be a minimum! On $M = \mathbb{R}^2 \setminus \{0\}$ with the Euclidean metric tensor, we have $d\big((1, 0), (-1, 0)\big) = 2$ but there is no curve in M joining $(1, 0)$ and $(-1, 0)$ having length precisely 2.

If g is not positive definite, one may still define d as above on M, but it usually fails to be a distance: if two distinct points $p, q \in M$ are connected by a lightlike curve, we get $d(p, q) = 0$.

9.1.12. Volume form. Recall from Section 7.5.1 that a metric tensor on an oriented manifold induces a volume form. Therefore every oriented pseudo-Riemannian manifold (M, g) has a canonical volume form ω. In coordinates,

$$\omega = \sqrt{|\det g_{ij}|}\,dx^1 \wedge \ldots \wedge dx^n.$$

Note for instance that all the pseudo-Riemannian manifolds $\mathbb{R}^{p,q}$ share the same volume form $\omega = dx^1 \wedge \ldots \wedge dx^n$. If the metric tensor g is altered by a conformal modification by multiplication with a positive function $\lambda\colon M \to (0, +\infty)$, the volume form ω changes accordingly to $\lambda^{\frac{n}{2}}\omega$.

Example 9.1.20. The volume form of the half-space model H^n is

$$\omega = \frac{1}{(x^n)^n} dx^1 \wedge \ldots \wedge dx^n.$$

9.1.13. Lorentzian manifolds. While Riemannian manifolds form the language of modern geometry, Lorentzian manifolds are of fundamental importance in general relativity. The prototypical Lorentzian manifold is the already encountered $(n+1)$-dimensional Minkowski space $\mathbb{R}^{n,1}$, for which it is natural to use the coordinates $x^0 = t, x^1, \ldots, x^n$.

Most of the discussion of Section 7.6.2 extends obviously from $\mathbb{R}^{3,1}$ to $\mathbb{R}^{n,1}$. The group $\mathrm{O}(n, 1)$ of all linear transformations of $\mathbb{R}^{n,1}$ that preserve the scalar product has two homomorphisms onto $\{\pm 1\}$ telling whether a given isomorphism preserves the orientation of $\mathbb{R}^{n,1}$ and of time. The kernels of these homomorphisms are denoted by

$$\mathrm{SO}(n, 1), \qquad \mathrm{O}^+(n, 1)$$

and their intersection

$$\mathrm{SO}^+(n, 1) = \mathrm{SO}(n, 1) \cap \mathrm{O}^+(n, 1)$$

consists of all isomorphisms that preserve both the orientations of $\mathbb{R}^{n,1}$ and of time. This group is one of the four connected components of $\mathrm{O}(n,1)$, see Proposition 7.6.2.

More generally, let V be a vector space equipped with a scalar product $\langle , \rangle$ with signature $(n-1,1)$. The timelike vectors in V form two open cones. A *time orientation* for V is the choice of an open cone, called *future*, while the other open cone is then called (not surprisingly) *past*.

Definition 9.1.21. A Lorentzian manifold M is *time orientable* if there is a locally coherent time orientation on all tangent spaces.

Here *locally coherent* means that the time orientation on T_pM should not jump discontinuously when we move $p \in M$. We express this by requiring that for every $p \in M$ there is a non vanishing vector field X on a neighbourhood U of p such that for every $q \in U$ the vector $X(q)$ is future timelike.

Proposition 9.1.22. *A Lorentzian manifold M is time orientable $\Longleftrightarrow$ there is a global timelike vector field X.*

Proof. If there is such a X, we can use it to define an orientation: at every $p \in M$ the future cone is the one containing $X(p)$. Conversely, given a time orientation we can find a future timelike vector field on an open neighbourhood $U(p)$ of every $p \in M$, and using a partition of unity we can patch all these to a single future timelike vector field on M. Everything works since future timelike vectors form a convex subset of T_pM. □

The Minkowski space $\mathbb{R}^{n,1}$ is naturally oriented and time oriented. Note that being orientable and time orientable are two independent properties:

Exercise 9.1.23. Construct a Lorentzian time orientable and not time orientable structure on both the annulus $S^1 \times \mathbb{R}$ and the Möbius strip.

We can always obtain orientability after passing to a double cover, and the same holds (with a similar proof) for time orientability:

Exercise 9.1.24. If a connected Lorentzian M is not time orientable, it has a double cover $\tilde{M} \to M$ whose induced Lorentzian structure is orientable.

Riemannian and Lorentzian manifolds share many features, but are also quite different in some aspects: for instance, every manifold M has a Riemannian structure, but not necessarily a Lorentzian one, as we now see.

Proposition 9.1.25. *Let M be a manifold. The following are equivalent:*

1. *There exists a Lorentzian structure on M.*
2. *There exists a time orientable Lorentzian structure on M.*
3. *There is a nowhere vanishing vector field on M.*
4. *Either M is non compact, or with boundary, or $\chi(M) = 0$.*

Proof. (2) $\Rightarrow$ (1) is obvious and (4) $\Leftrightarrow$ (3) is well-known. TBD!

(3) $\Rightarrow$ (2). Pick any Riemannian metric g on M. If X is a non-vanishing vector field on M, up to normalising we may suppose that $\|X(p)\| = 1$ for all $p \in M$ and then define a new tensor field

$$g'(v, w) = g(v, w) - 2g(v, X)g(w, X).$$

In arbitrary coordinates, we have

$$g'_{ij} = g_{ij} - 2g_{ik}X^k g_{jl}X^l.$$

We extend $X(p)$ to an orthonormal basis, so that $g_{ij} = \delta_{ij}$ and $g'_{ij} = \left(\begin{smallmatrix} -1 & 0 \\ 0 & I_n \end{smallmatrix}\right)$ with respect to this basis. Therefore g' is a metric tensor of signature $(n, 1)$.

(1) $\Rightarrow$ (3). If M is time orientable, there is a global non-vanishing vector field by Proposition 9.1.22. If M is not time orientable, its double cover $\tilde{M}$ is, hence it has a non-vanishing vector field, hence we get (4) for $\tilde{M}$, which in turn implies (4) also for M, that is equivalent to (3). $\square$

Here we use $\chi(\tilde{M}) = d\chi(N)$.

9.2. Connections

We now want to define geodesics. On a Riemannian manifold, it would be natural to define them as curves that minimise locally the distance; however, differential geometers usually prefer to take a different perspective: they introduce geodesics as curves that go as "straight" as possible.

To formalise this notion of "straight curve" we need somehow to compare tangent vectors at nearby points. This comparison may be formalised via a powerful additional structure called a *connection*. This structure has many interesting features that go beyond the definition of geodesics: it is also a way to derive vector fields along tangent vectors, and for that reason it is also called with another appropriate name: *covariant derivative*. The two notions – connection and covariant derivative – are in fact the same thing, a powerful structure that can be employed for different purposes, which applies to any pseudo-Riemannian manifold, and more generally to any smooth manifold. In fact, we do not need a metric tensor to define a connection. However, we may use the metric tensor to get a *preferred* one, called the *Levi-Civita connection*.

9.2.1. Definition. As we said in the previous chapters, one of the recurring themes in differential topology is the quest for a correct notion of derivation of vector (more generally, tensor) fields on a smooth manifold M. Without equipping M with an additional structure, the best thing that we can do is to derive a vector field Y with respect to another vector field X via the *Lie derivative* $\mathcal{L}_X(Y) = [X, Y]$.

As we have already noted, the definition of $\mathcal{L}_X(Y)$ is *local*, in the sense that its value at $p \in M$ depends only on the values of X and Y in any neighbourhood of p, but *it is not a pointwise definition*, in the sense that it is not determined by the vector $v = X(p)$ alone, as it happens in the usual directional derivative of smooth functions in $\mathbb{R}^n$. We are then urged to introduce a somehow stronger notion of derivation that depends only on the tangent vector $v = X(p)$.

Let M be a smooth manifold.

Definition 9.2.1. A *connection* ∇ is an operation that assigns to every $v \in T_pM$ at every $p \in M$, and to every vector field X defined on a neighbourhood of p, another tangent vector

$$\nabla_v X \in T_pM$$

called the *covariant derivative* of X along v, such that the following holds:

(1) if X and Y agree on a neighbourhood of p, then $\nabla_v X = \nabla_v Y$;

(2) we have linearity in both terms:

$$\begin{aligned} \nabla_v(\lambda X + \mu Y) &= \lambda \nabla_v(X) + \mu \nabla_v(Y), \\ \nabla_{\lambda v + \mu w} X &= \lambda \nabla_v(X) + \mu \nabla_w(X), \end{aligned}$$

where $\lambda, \mu \in \mathbb{R}$ are arbitrary scalars;

(3) the Leibniz rule holds:

$$\nabla_v(fX) = v(f)X(p) + f(p)\nabla_v X$$

for every function f defined in a neighbourhood of p;

(4) ∇ depends smoothly on p.

We must explain the last condition. For every two vector fields X, Y defined in a common open subset $U \subset M$, we require

$$p \mapsto \nabla_{Y(p)} X$$

to be another vector field in U, that we denote simply by $\nabla_Y X$. That is, we require $\nabla_{Y(p)} X$ to vary smoothly with respect to the point $p \in U$.

9.2.2. Christoffel symbols. We now examine a given connection ∇ on a chart. If we consider the coordinate constant vector fields $e_i = \frac{\partial}{\partial x^i}$, we get

$$\nabla_{e_i} e_j = \Gamma_{ij}^k e_k$$

where we have used the Einstein summation convention, for some real numbers Γ_{ij}^k that depend smoothly on p because of the smoothness assumption (4).

The smooth functions Γ_{ij}^k are called the *Christoffel symbols* of the connection. On a chart, these determine the connection completely: indeed, for every vector field $X = X^j e_j$ and tangent vector $v = v^i e_i$ at some point we get

$$\begin{aligned} \nabla_v X = v^i \nabla_{e_i}(X^j e_j) &= v^i \frac{\partial X^j}{\partial x^i} e_j + v^i X^j \nabla_{e_i} e_j \\ &= v^i \frac{\partial X^j}{\partial x^i} e_j + v^i X^j \Gamma_{ij}^k e_k \end{aligned}$$

by applying linearity and the Leibniz rule. We may rewrite this equality as

$$\nabla_v X = \left(v^i \frac{\partial X^k}{\partial x^i} + v^i X^j \Gamma_{ij}^k \right) e_k. \tag{25}$$

The formula says that the covariant derivative ∇_v is the usual directional derivative along v plus an additional correction term that is determined by the Christoffel symbols Γ_{ij}^k. In particular

$$\nabla_{e_i} X = \frac{\partial X}{\partial x^i} + X^j \Gamma_{ij}^k e_k.$$

Note that the directional derivative is not a chart-independent operation! If it were, we would use it as a preferred connection – but we cannot. In fact, on a manifold M there is usually no preferred connection.

You may think at Γ_{ij}^k as some additional correction term that transforms the directional derivative into a chart-independent operation. It is worth noting that the additional correction term $v^i X^j \Gamma_{ij}^k e_k$ is not as ugly as it might look at a first glance – in fact, it is actually very nice: it only depends linearly on v and $X(p)$, and not on the derivatives of X at p. The directional derivative is also linear in v, but it depends on the behaviour of X on a neighbourhood of p and not only on $X(p)$. So the additional correction term is of a simpler nature than the directional derivative.

Conversely, on any open subset U in $\mathbb{R}^n$ or $\mathbb{R}^n_+$, for every choice of smooth maps $\Gamma_{ij}^k : U \to \mathbb{R}$ there is a connection ∇ whose Christoffel symbols are Γ_{ij}^k. The connection ∇ is defined via (25), and one verifies that the axioms (1)-(4) are satisfied.

When the connection is read on another chart the Christoffel symbols modify in some appropriate way. Now we must admit that their transformation formula is not very nice. Luckily, we will never need it, so we can forget it.

Exercise 9.2.2. If the coordinates change as

$$\frac{\partial}{\partial \hat{x}^i} = \frac{\partial x^k}{\partial \hat{x}^i} \frac{\partial}{\partial x^k}$$

the Christoffel symbols modify accordingly as follows:

$$\hat{\Gamma}_{ij}^k = \frac{\partial x^p}{\partial \hat{x}^i} \frac{\partial x^q}{\partial \hat{x}^j} \Gamma_{pq}^r \frac{\partial \hat{x}^k}{\partial x^r} + \frac{\partial \hat{x}^k}{\partial x^m} \frac{\partial^2 x^m}{\partial \hat{x}^i \partial \hat{x}^j}.$$

If only the first term were present, the Christoffel symbols Γ_{ij}^k would vary like a tensor field of type $(1, 2)$. Unfortunately, the second term with its second derivatives forbids us to interpret the Christoffel symbols Γ_{ij}^k as coordinates of some tensor field. In fact, this is not surprising: if ∇ were a tensor field, the value of $\nabla_v X$ in p would depend only on $X(p)$, and this would contradict any reasonable idea of derivative, because the derivative of an object like a vector field X measures (in some sense) how the object X varies in the direction v, and it cannot be determined only by the value $X(p)$ that the object has in p.

Remark 9.2.3. Like tensor fields and many other objects, connections can be transported along diffeomorphisms. If $\varphi : M \to N$ is a diffeomorphism and ∇ is a connection on M, we define the connection $\varphi_* \nabla$ on N in the obvious way as $(\varphi_* \nabla)_{d\varphi_p(v)} Y = d\varphi_p(\nabla_v \varphi^* Y)$ for any $Y \in \mathfrak{X}(N)$, $p \in M$, and $v \in T_p M$.

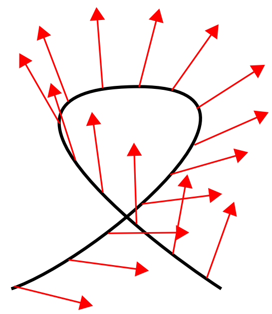

Figure 9.5. This vector field on an immersed curve is not induced by a vector field on M, since it gives distinct vectors at the same point of M.

9.2.3. Curves suffice. We know that $\nabla_v X \in T_pM$ depends only on the behaviour of X on any neighbourhood of p. In fact, its restriction to a smaller subset suffices to determine $\nabla_v X$.

Proposition 9.2.4. *The covariant derivative $\nabla_v X \in T_pM$ depends only on v and on the restriction of X to any curve tangent to v.*

Proof. On a chart, the equation (25) shows that $\nabla_v X$ depends only on $v, X(p)$, and the directional derivative of X along v. These in turn depend only on the restriction of X to any curve tangent to v. □

In particular, two vector fields that coincide on some curve tangent to v have the same covariant derivative along v. This leads us to study vector fields along curves, and to define their covariant derivatives.

9.2.4. Vector fields along curves. We define a notion of vector field along a curve that is valid in wide generality, for any kind of curve. We then use a connection ∇ to derive these vector fields along the curve.

Definition 9.2.5. Let M be a manifold and $\gamma\colon I \to M$ a curve. A *vector field* along γ is a smooth map $X\colon I \to TM$ with $X(t) \in T_{\gamma(t)}M$ for all $t \in I$.

The vector field X is *tangent* to γ if $X(t)$ is a multiple of $\gamma'(t)$ for all t. For instance, the *velocity field* of γ is the vector field $\gamma'(t)$ and is of course tangent to γ. The vector fields along γ form naturally a vector space.

Every vector field X defined in some open neighbourhood $U \subset M$ of the support of γ induces a vector field $X(t) = X(\gamma(t))$ on γ. If γ is an embedding, every vector field X on γ is induced by some vector field X on M by Proposition 4.4.1. This is false if γ is not an embedding, see for instance Figure 9.5.

Let ∇ be a fixed connection on M. For every vector field X along γ, we define another vector field D_tX on γ called its *covariant derivative*, as follows.

If γ is an embedding, the vector field X is induced by a vector field X on M and for every $t \in I$ we define

$$D_tX = \nabla_{\gamma'(t)}X.$$

The vector field D_tX does not depend on the extension of X outside of γ thanks to Proposition 9.2.4. If γ is an immersion, it is locally an embedding and hence the above definition applies locally. On a chart we get

$$D_tX = X'(t) + \gamma'(t)^i X^j(t)\Gamma_{ij}^k(\gamma(t))e_k. \tag{26}$$

The definition of D_t may also be extended to curves γ that are not immersions, using this formula and showing that it is chart-independent, or by pinpointing a complete list of axioms for D_t as in Exercise 9.4.1 below.

9.2.5. Parallel transport. When we introduce some kind of derivative, it is natural to investigate the objects whose derivative is everywhere zero.

Let M be a smooth manifold equipped with a connection ∇. Let $\gamma\colon I \to M$ be any curve. A vector field X along γ is *parallel* if

$$D_tX = 0$$

for all $t \in I$. Here is a very important existence and uniqueness property:

Proposition 9.2.6. *For every $t_0 \in I$ and every $v \in T_{\gamma(t_0)}M$ there is a unique parallel vector field X on γ with $X(t_0) = v$.*

Proof. We first consider the case where $\gamma(I)$ is entirely contained in the domain U of a chart $\varphi\colon U \to V$. Using (26), the problem reduces to solving a system of n linear differential equations in $X^k(t)$ with $k = 1, \dots, n$, that is:

$$\frac{dX^k}{dt} + \gamma'(t)^i X^j(t)\Gamma_{ij}^k(\gamma(t)) = 0. \tag{27}$$

The system has a unique solution satisfying the initial condition $X^k(t_0) = v^k$ for all k. The solution exists for all $t \in I$ because the system is linear.

For every $t \in I$ we can cover the segment $[t_0, t]$ or $[t, t_0]$ with finitely many charts and we conclude. $\square$

For every $t \in I$, we think at the vector $X(t)$ as the one obtained from $v = X(t_0)$ by *parallel transport* along γ. We have just discovered a very nice (and maybe unexpected) feature of connections: they may be used to transport tangent vectors along curves. This is a crucial property.

It is sometimes useful to denote the parallel-transported vector $X(t)$ as

$$X(t) = \Gamma(\gamma)_{t_0}^t(v)$$

to stress the dependence on all the objects involved. We get a map

$$\Gamma(\gamma)_{t_0}^t\colon T_{\gamma(t_0)}M \longrightarrow T_{\gamma(t)}M$$

called the *parallel transport map*.

Proposition 9.2.7. *The parallel transport map is a linear isomorphism.*

Proof. The map is linear because (27) is a linear homogeneous system of differential equations, so solutions form a vector space. It is an isomorphism because its inverse is $\Gamma(\gamma)_t^{t_0}$. $\square$

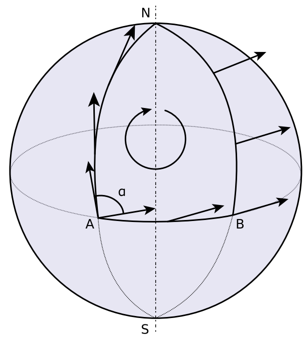

Figure 9.6. By parallel-transporting a vector v along the edges of a spherical triangle in S^2, from A to N to B and back to A, we end up with another vector w that makes some angle α with the original v. In S^2 the angle α is proportional to the area of the triangle ABN, and in general it is connected to the curvature of the manifold. The connection ∇ that we are using here is the Levi - Civita connection naturally associated to the metric, yet to be defined in Section 9.3.

Note that

$$\Gamma(\gamma)_{t_0}^{t_2} = \Gamma(\gamma)_{t_1}^{t_2} \circ \Gamma(\gamma)_{t_0}^{t_1}$$

for every triple $t_0, t_1, t_2 \in I$. The smooth dependence on initial values tells us that $\Gamma(\gamma)_t^{t'}$ depends smoothly on t and t', when read on charts.

We now understand where the name "connection" comes from: the operator ∇ can be used to connect via isomorphisms all the tangent spaces T_pM at the points $p = \gamma(t)$ visited by any curve γ. It is important to stress here that the isomorphisms depend heavily on the chosen curve γ: two distinct curves γ_1 and γ_2, both connecting the same points p and q, produce in general two different isomorphisms between the tangent spaces T_pM and T_qM. This may hold also if γ_1 and γ_2 are homotopic. As we will see, the *curvature* of ∇ measures precisely this discrepancy. See Figure 9.6.

Remark 9.2.8. A continuous map $\gamma\colon I \to M$ is *piecewise smooth* if it is a concatenation of finitely many smooth curves. Parallel transport extends to piecewise smooth curves in the obvious way, see Figure 9.6.

9.2.6. Connections form an affine space. Does every smooth manifold admit some connection ∇? And if it does, how many connections are there? The answer to the first question is positive but we postpone it to the next section. We can easily answer the second one here.

Recall that a tensor field T of type $(1,2)$ on M is a bilinear map

$$T(p)\colon T_pM \times T_pM \longrightarrow T_pM$$

that depends smoothly on p.

Proposition 9.2.9. *If ∇ is a connection on M and $T \in \Gamma\big(\mathcal{T}_1^2(M)\big)$ is a tensor field of type $(1,2)$, then the operator $\nabla' = \nabla + T$, defined as*

$$\nabla'_v X = \nabla_v X + T(p)(v, X(p))$$

for every $p \in M$, $v \in T_pM$, and vector field X near p, is also a connection. Every connection ∇' on M arises in this way.

Proof. To prove that ∇' is a connection, we show that it satisfies the Leibniz rule (the other axioms are obvious). We have:

$$\begin{aligned}\nabla'_v(fX) &= \nabla_v(fX) + T(p)\big(v, f(p)X(p)\big)\\ &= v(f)X + f(p)\nabla_v X + f(p)T(p)\big(v, X(p)\big)\\ &= v(f)X + f(p)\nabla'_v X.\end{aligned}$$

Conversely, if ∇' is another connection, we consider the expressions in coordinates (25) for both $\nabla'_v X$ and $\nabla_v X$ and discover that

$$\nabla'_v X - \nabla_v X = v^i X^j \big((\Gamma')_{ij}^k - \Gamma_{ij}^k\big) e_k.$$

The right-hand expression describes a tangent vector at p that depends (linearly) only on the tangent vectors v and $X(p)$. If we indicate this vector as $T(p)(v, X(p))$, we get a tensor field T of type (1,2). In coordinates, we have

$$T_{ij}^k = (\Gamma')_{ij}^k - \Gamma_{ij}^k.$$

The proof is complete. □

A connection is not a tensor field, but the difference of two connections is. The space of all connections ∇ on M is naturally an affine space on the (infinite-dimensional) space $\Gamma\big(\mathcal{T}_1^2(M)\big)$.

Remark 9.2.10. We can use Exercise 9.2.2 to confirm that $T_{ij}^k = (\Gamma')_{ij}^k - \Gamma_{ij}^k$ are the coordinates of a tensor (the second partial derivatives cancel).

9.2.7. Covariant derivative of tensor fields. A covariant derivative ∇ on M gives rise to parallel transport, and parallel transport in turn generates a more powerful notion of covariant derivative that applies to any tensor field. We explain this phenomenon here.

The short description is that parallel transport allows us to identify all the tangent spaces along a curve, and with this tool we can differentiate any kind of tensor field that is defined on this curve.

More precisely, let T be a tensor field of type (h,k) on a neighbourhood of $p \in M$. For any $v \in T_pM$, we would like to define

$$\nabla_v T \in \mathcal{T}_h^k(T_pM).$$

We do this using parallel transport along curves as follows. Choose an embedded curve $\gamma\colon I \to M$ with $0 \in I$, $\gamma(0) = p$ and $\gamma'(0) = v$. The parallel transport $\Gamma(\gamma)_{t_0}^{t_1}$ along γ is an isomorphism between tangent spaces, which extends canonically to an isomorphism of tensor spaces:

$$\Gamma(\gamma)_{t_0}^{t_1}\colon \mathcal{T}_h^k(T_{\gamma(t_0)}M) \longrightarrow \mathcal{T}_h^k(T_{\gamma(t_1)}M).$$

We define

$$\nabla_v T = \frac{d}{dt}\Gamma(\gamma)_t^0\big(T(\gamma(t))\big)\Big|_{t=0}.$$

Proposition 9.2.11. *The definition is independent of the choice of γ. In coordinates we get*

$$\begin{aligned}(\nabla_v T)^{i_1,\dots,i_h}_{j_1,\dots,j_k} ={}& v^i\frac{\partial}{\partial x^i}(T^{i_1,\dots,i_h}_{j_1,\dots,j_k}) + v^i T^{j,i_2\dots,i_h}_{j_1,\dots,j_k}\Gamma^{i_1}_{ij} + \cdots + v^i T^{i_1,\dots,i_{h-1},j}_{j_1,\dots,j_k}\Gamma^{i_h}_{ij} \\ & - v^i T^{i_1\dots,i_h}_{l,j_2,\dots,j_k}\Gamma^l_{ij_1} - \cdots - v^i T^{i_1\dots,i_h}_{j_1,\dots,j_{k-1},l}\Gamma^l_{ij_k}.\end{aligned}$$

Proof. We pick a chart and write everything in coordinates in $\mathbb{R}^n$ or $\mathbb{R}^n_+$. We would like to study how the canonical basis of $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$ are transported along γ. We set $p = 0$ and $w_i(t) = \Gamma(\gamma)_0^t(e_i)$. From (27) we deduce that

$$\dot w_i(t) + \gamma'(t)^j w_i^k(t)\Gamma^l_{jk}(\gamma(t))e_l = 0.$$

In particular at $t = 0$ we get

$$\dot w_i(0) + v^j\Gamma^l_{ji}(0)e_l = 0,$$

from which we deduce (exercise) that the derivative of the dual basis satisfies

$$\dot w^i(0) - v^j\Gamma^i_{jl}(0)e^l = 0.$$

At any time $t \in I$ we can write

$$T = T^{i_1,\dots,i_h}_{j_1,\dots,j_k} w_{i_1}\otimes\cdots\otimes w_{i_h}\otimes w^{j_1}\otimes\cdots\otimes w^{j_k}. \tag{28}$$

By definition the covariant derivative of T is

$$\nabla_v T = \dot T^{i_1,\dots,i_h}_{j_1,\dots,j_k}(0)e_{i_1}\otimes\cdots\otimes e_{i_h}\otimes e^{j_1}\otimes\cdots\otimes e^{j_k}.$$

All the terms in (28) depend on time, so we now derive it with respect to t. We omit the symbol $\otimes$ for simplicity and get

$$\begin{aligned} v^i\frac{\partial T}{\partial x^i}\Big|_{x=0} = \frac{dT}{dt}\Big|_{t=0} &= \dot T^{i_1,\dots,i_h}_{j_1,\dots,j_k}(0)e_{i_1}\cdots e_{i_h}e^{j_1}\cdots e^{j_k} \\ &\quad + T^{i_1,\dots,i_h}_{j_1,\dots,j_k}(0)\dot w_{i_1}(0)e_{i_2}\cdots e_{i_h}e^{j_1}\cdots e^{j_k} + \cdots \\ &\quad + T^{i_1,\dots,i_h}_{j_1,\dots,j_k}(0)e_{i_1}\cdots e_{i_h}\dot w^{j_1}(0)e^{j_2}\cdots e^{j_k} + \cdots \\ &= \nabla_v T - T^{i_1,\dots,i_h}_{j_1,\dots,j_k}(0)v^j\Gamma^l_{ji_1}(0)e_l e_{i_2}\cdots e_{i_h}e^{j_1}\cdots e^{j_k} - \cdots \\ &\quad + T^{i_1,\dots,i_h}_{j_1,\dots,j_k}(0)v^j\Gamma^{j_1}_{jl}(0)e_{i_1}\cdots e_{i_h}e^l e^{j_2}\cdots e^{j_k} + \cdots \end{aligned}$$

The conclusion follows by renaming indices. $\square$

The general principle is always the same: in coordinates the covariant derivative is the directional derivative plus some linear correction terms governed by the Christoffel symbols. Despite its slightly awe-inspiring appearance due to the presence of many indices, we could not have hoped for a simpler formula.

In particular, when $T = f$ is a function we get $\nabla_v f = v(f)$ and when $T = X$ is a vector field we recover the original definition of ∇_X (and this is quite reassuring). If ω is a 1-form, then

$$\nabla_v \omega = v^i \frac{\partial \omega}{\partial x^i} - v^i \omega_l \Gamma_{ij}^l e^j = v^i \left(\frac{\partial \omega_j}{\partial x^i} - \omega_l \Gamma_{ij}^l \right) e^j.$$

If g is a metric tensor, then

$$(\nabla_v g)_{jk} = v^i \left(\frac{\partial g_{jk}}{\partial x^i} - g_{lk} \Gamma_{ij}^l - g_{jl} \Gamma_{ik}^l \right). \tag{29}$$

Corollary 9.2.12. *The following hold:*

1. *If T and U agree on a neighbourhood of p, then $\nabla_v T = \nabla_v U$.*
2. *$\nabla_v T$ is linear both in v and T.*
3. *The Leibniz rule is satisfied for any pair of tensor fields T, U near p:*
$$\nabla_v (T \otimes U) = (\nabla_v T) \otimes U(p) + T(p) \otimes \nabla_v U.$$
4. *∇ depends smoothly on p, in the sense that $\nabla_X T$ is a tensor field for every vector field X.*
5. *∇_v commutes with contractions.*

We may interpret the connection ∇ as a particular linear map

$$\nabla \colon \Gamma(\mathcal{T}_h^k(M)) \longrightarrow \Gamma(\mathcal{T}_h^{k+1}(M))$$

determined by requiring that $\nabla(T)$ sends a vector field X to $\nabla_X T$. The coordinates of ∇T are

$$\begin{aligned}\nabla_i T_{j_1,\ldots,j_k}^{i_1,\ldots,i_h} = \frac{\partial}{\partial x^i}(T_{j_1,\ldots,j_k}^{i_1,\ldots,i_h}) + T_{j_1,\ldots,j_k}^{j,i_2\ldots,i_h} \Gamma_{ij}^{i_1} + \cdots + T_{j_1,\ldots,j_k}^{i_1,\ldots,i_{h-1},j} \Gamma_{ij}^{i_h} \\ - T_{l,j_2,\ldots,j_k}^{i_1\ldots,i_h} \Gamma_{ij_1}^l - \cdots - T_{j_1,\ldots,j_{k-1},l}^{i_1\ldots,i_h} \Gamma_{ij_k}^l.\end{aligned}$$

It is often convenient, as in the formula, to write the i-th coordinate of ∇T as a pedix of ∇ to stress its peculiar role and avoid potential ambiguities.

Remark 9.2.13. Using the covariant derivative we have defined parallel transport along curves; conversely, we have just seen that parallel transport along curves determines the covariant derivative. So covariant derivative and parallel transport are essentially the same thing.

9.3. The Levi-Civita connection

On a Riemannian manifold we can talk about distances between points and length of curves. On a more general pseudo-Riemannian manifold M we can talk about volumes. We now show that a pseudo-Riemannian manifold M

also has a preferred connection, called the *Levi-Civita connection*. We will use it to define geodesics in the next section.

9.3.1. Introduction. A smooth manifold M carries many different connections, and we are now looking at some reasonable way to discriminate between them. The main motivation is the following ambitious question: if M has a metric tensor g, is there a connection ∇ that is more suited to g?

An elegant and useful way to understand a connection ∇ consists of examining some tensor fields that are assigned canonically to ∇. We now introduce one of these.

9.3.2. Torsion. Let ∇ be a connection on a smooth manifold M. The *torsion* T of ∇ is a tensor field of type $(1, 2)$ defined as follows. For every $p \in M$ and $v, w \in T_pM$ we set

$$T(p)(v, w) = \nabla_v Y - \nabla_w X - [X, Y](p)$$

where X and Y are any vector fields defined in a neighbourhood of p extending the tangent vectors v and w. Of course we need to prove that this definition is well-posed, a fact that is not evident at all at first sight.

Proposition 9.3.1. *The tangent vector $T(p)(v, w)$ is independent of the extensions X and Y.*

Proof. In coordinates we have

$$\begin{aligned} T(p)(v, w) &= \left(v^i \frac{\partial Y^k}{\partial x^i} + v^i Y^j \Gamma^k_{ij} - w^i \frac{\partial X^k}{\partial x^i} - w^i X^j \Gamma^k_{ij} - v^i \frac{\partial Y^k}{\partial x^i} + w^i \frac{\partial X^k}{\partial x^i} \right) e_k \\ &= \left(v^i w^j \Gamma^k_{ij} - w^i v^j \Gamma^k_{ij} \right) e_k = v^i w^j (\Gamma^k_{ij} - \Gamma^k_{ji}) e_k. \end{aligned}$$

The proof is complete. $\square$

Along the proof we have also shown that in coordinates we have

$$T^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}.$$

A connection ∇ is *symmetric* if its torsion vanishes, that is if $\Gamma^k_{ij} = \Gamma^k_{ji}$ on any coordinate chart. The torsion is clearly an antisymmetric tensor, that is $T(p)(v, w) = -T(p)(w, v)$ for all v, w. Finally, if we contract the torsion T with two vector fields X and Y we get the elegant equality of vector fields:

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

Remark 9.3.2. If the torsion vanishes, we get

$$[X, Y] = \nabla_X Y - \nabla_Y X.$$

This equality may be interpreted by saying that the Lie bracket $[X, Y]$ can be defined in a coordinate-independent way using the covariant derivative ∇ in place of the (coordinate-dependent) directional derivative (see Exercise 5.4.4).

Remark 9.3.3. The torsion is natural, that is it commutes with diffeomorphisms. We mean that if $\varphi\colon M \to N$ is a diffeomorphism and ∇ is a connection for M with torsion T, the transported tensor field $\varphi_* T$ is the torsion of $\varphi_*\nabla$. In particular ∇ is symmetric $\Longleftrightarrow$ $\varphi_*\nabla$ is.

9.3.3. Bilinear operators on vector fields. We have already encountered in this book three bilinear operators

$$\mathfrak{X}(M) \times \mathfrak{X}(M) \longrightarrow \mathfrak{X}(M)$$

that are quite dissimilar in nature, and it is important to recognise their mutual differences. These are the Lie bracket $[,]$ (that is intrinsic of M), the connection ∇ (that is not intrinsic and has to be added to M), and the torsion T (that depends on ∇). In all three cases, given two vector fields X and Y, we can define a third one

$$[X, Y], \qquad \nabla_X Y, \qquad T(X, Y).$$

The main difference between these three operators is the following:

- $[X, Y]$ at p depends on X and Y;
- $\nabla_X Y$ at p depends on $X(p)$ and Y;
- $T(X, Y)$ at p depends on $X(p)$ and $Y(p)$.

When we write that "$[X, Y]$ at p depends on X and Y", we mean that the datum of $X(p)$ and $Y(p)$ is not enough to determine $[X, Y](p)$. We need to know the behaviour of both X and Y in a neighbourhood of p. These differences express the fact that the operator T is the only one among the three that is in fact a tensor field.

Remark 9.3.4. Some authors describe these differences by saying that the operator T is $C^\infty(M)$-bilinear, that is $T(fX, gY) = fgT(X, Y)$ for every $f, g \in C^\infty(M)$. Analogously, ∇ is left $C^\infty(M)$-linear, that is $\nabla_{fX} Y = f\nabla_X Y$, but is not right $C^\infty(M)$-linear. The Lie bracket is neither left nor right $C^\infty(M)$-linear.

9.3.4. Compatible connections. We now consider a pseudo-Riemannian manifold (M, g). As we said above, we would like to assign an appropriate conection ∇ to g. We start by defining a reasonable compatibility condition.

We say that a connection ∇ is *compatible* with g if every parallel transport isomorphism

$$\Gamma(\gamma)_{t_0}^{t_1}\colon T_{\gamma(t_0)}M \longrightarrow T_{\gamma(t_1)}M$$

is actually an isometry, for every curve $\gamma\colon I \to M$ and every $t_0, t_1 \in I$. This condition is quite strong, since we are imposing it on every curve γ.

As the following proposition shows, this is a robust definition,[1] because it may be expressed in various different and simple ways: by requiring that

[1] In mathematics, we may say that a definition is *robust* if it can be expressed in various different ways. This is of course not a rigorous concept, but it is an evident fact that definitions that are both robust and simple to state are by far those preferred by mathematicians: the main reason is that our brain storage for definitions is very limited, and it is important

$\nabla g = 0$, which translates into a concrete equation (30) relating g and Γ_{ij}^k, or by asking that a natural version of the Leibniz rule applies. Note that if $X, Y \in \mathfrak{X}(U)$ are vector fields, their scalar product $\langle X, Y\rangle \in C^\infty(U)$ is a smooth function.

Proposition 9.3.5. *The following conditions are equivalent:*

(1) *the connection ∇ is compatible with g;*
(2) *$\nabla g = 0$, that is on every chart we have*

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ki}^l g_{lj} + \Gamma_{kj}^l g_{li}; \tag{30}$$

(3) *for every $p \in M$, $v \in T_pM$, and vector fields X, Y near p we get*

$$v\langle X, Y\rangle = \langle \nabla_v X, Y(p)\rangle + \langle X(p), \nabla_v Y\rangle;$$

(4) *for every curve $\gamma\colon I \to M$ and vector fields X, Y on it we get*

$$\frac{d}{dt}\langle X, Y\rangle = \langle D_t X, Y\rangle + \langle X, D_t Y\rangle.$$

Proof. (1)⇒(2). Since every parallel transport $\Gamma(\gamma)_t^0$ is an isometry, its action on $(0, 2)$ tensors sends $g(\gamma(t))$ to $g(\gamma(0))$, independently of t. Therefore by definition we get $\nabla g = 0$. Apply (29) to get (30).

(2)⇒(3). Using Corollary 9.2.12-(3) we get

$$\nabla_v(g \otimes X \otimes Y) = g \otimes \nabla_v X \otimes Y + g \otimes X \otimes \nabla_v Y$$

and we conclude by contracting (allowed by Corollary 9.2.12-(5)).

(3)⇒(4). Pick $v = \gamma'(t)$.

(4)⇒(1). If X and Y are parallel, the function $\langle X, Y\rangle$ is constant, hence parallel transport is an isometry. □

We note that if (30) holds on all the charts of an atlas, then it also does at any compatible chart, since it is equivalent to $\nabla g = 0$.

9.3.5. The Levi-Civita connection. As promised, we now assign to any pseudo-Riemannian manifold (M, g) a canonical connection ∇ called the *Levi-Civita connection*.

Theorem 9.3.6. *Every pseudo-Riemannian manifold (M, g) has a unique symmetric compatible connection ∇. On any chart, its Christoffel symbols are*

$$\Gamma_{ij}^l = \frac{1}{2} g^{kl} \left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} \right). \tag{31}$$

---

to choose a small number of good ones that are powerful enough to apply to a big number of different complex situations.

Proof. We start by proving uniqueness. Let ∇ be a symmetric compatible connection. On a chart, we write (30) three times

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{li}, \quad \frac{\partial g_{jk}}{\partial x^i} = \Gamma^l_{ij} g_{lk} + \Gamma^l_{ik} g_{lj} \quad \frac{\partial g_{ki}}{\partial x^j} = \Gamma^l_{jk} g_{li} + \Gamma^l_{ji} g_{lk}.$$

with i, j, k permuted cyclically, and using symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$ we get

$$\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} = 2\Gamma^l_{ij} g_{lk}.$$

By multiplying both members with the inverse matrix g^{km} we find

$$\Gamma^m_{ij} = \frac{1}{2} g^{km} \left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} \right).$$

This shows that Γ^l_{ij} and hence ∇ are uniquely determined.

Concerning existence, we now use (31) to define ∇ locally on a chart. The connection is clearly symmetric and it is also compatible because

$$\Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{li} = \frac{1}{2} \left(\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ki}}{\partial x^j} \right) + \frac{1}{2} \left(\frac{\partial g_{ji}}{\partial x^k} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{kj}}{\partial x^i} \right) = \frac{\partial g_{ij}}{\partial x^k}.$$

The resulting ∇ is chart-independent: if not, we would get two different symmetric and compatible connections on some open set, which is impossible. Therefore all the ∇ constructed along charts glue to a global ∇ on M. □

The unique symmetric compatible connection ∇ is called the *Levi-Civita connection* of (M, g).

Example 9.3.7. If $\mathbb{R}^n$ is equipped with the Riemannian metric g, the Christoffel symbols $\Gamma^k_{ij} = 0$ vanish everywhere and the Levi-Civita connection coincides with the usual directional derivative. More generally, this holds for any pseudo-Riemannian manifold $\mathbb{R}^{p,q}$ since g_{ij} is constant.

We will since now equip every pseudo-Riemannian manifold (M, g) with its Levi-Civita connection ∇.

Remark 9.3.8. The Levi-Civita connection is *natural*, that is it commutes with isometries. We mean that every isometry $\varphi\colon (M, g) \to (N, h)$ between pseudo-Riemannian manifolds sends the Levi-Civita connection ∇ of g to the Levi-Civita connection $\varphi_*\nabla$ of h. This holds because $\varphi_*\nabla$ is symmetric and compatible with h, see Remark 9.3.3.

Remark 9.3.9. While the compatibility assumption looks natural, the reasons for preferring a symmetric connection may look obscure at this point. We can express three arguments in its favour: (i) this seems the only (or at least the simplest) way to get a canonical and natural connection; (ii) symmetry has some nice consequences at various points, for instance we get that the Levi-Civita connection behaves well with submanifolds (see the next section); (iii) symmetry is assumed in general relativity based on physical grounds.

Remark 9.3.10. If we rescale the metric g by some constant $\lambda \neq 0$, we get a new metric $g' = \lambda g$ with the same Levi-Civita connection $\nabla' = \nabla$. We can verify this by looking at the formula for Γ_{ij}^k in coordinates, or by noticing that ∇ is still symmetric and compatible with g'.

If g is modified by a more complicated conformal transformation, the connection ∇ may be altered dramatically, as we will see in some important examples like the conformal models for the hyperbolic space.

9.3.6. Submanifolds. Let M be a pseudo-Riemannian manifold without boundary and $N \subset M$ a pseudo-Riemannian submanifold. Both M and N have their Levi-Civita connections ∇^M and ∇^N. We now show that ∇^N is easily determined by ∇^M. This is particularly useful when the ambient space is $M = \mathbb{R}^{p,q}$ since here ∇^M is just the usual directional derivative.

Let $p \in N$ be a point and $v \in T_pN$ a tangent vector. Let X be a vector field tangent to N defined on a neighbourhood of p in N. Extend X arbitrarily to a vector field on a neighbourhood of p in M. Let $\pi \colon T_pM \to T_pN$ be the orthogonal projection.

Proposition 9.3.11. *The following holds:*

$$\nabla_v^N X = \pi\big(\nabla_v^M X\big).$$

Proof. At every $p \in N$ we may choose coordinates such that $p = 0$, $M = \mathbb{R}^n \times \mathbb{R}^k$, $N = \mathbb{R}^n \times \{0\}$, and after a linear transformation we may also require that $g_{ij}(0)$ is diagonal, and hence the inverse $g^{ij}(0)$ also is. In these coordinates we can easily check, by looking at (31), that the Christoffel symbols $\Gamma_{ij}^l(0)$ of N are precisely those of M with $1 \leq i, j, l \leq n$. Note that this holds only at the point $p = 0$, not in the nearby.

Using the formula (25) we easily deduce that $\nabla_v^N X = \pi(\nabla_v^M X)$. $\square$

Let $\gamma \colon I \to N$ be a curve and X be a vector field on γ. We denote by $D_t^M X$ and $D_t^N X$ the covariant derivatives of X along γ with respect to the two connections ∇^M and ∇^N.

Corollary 9.3.12. *The following holds:*

$$D_t^N X = \pi\left(D_t^M X\right)$$

where $\pi \colon T_{\gamma(t)}M \to T_{\gamma(t)}N$ is the orthogonal projection.

In particular the vector field X is parallel on N if and only if its covariant derivative on M is everywhere orthogonal to N. The case where M is the Euclidean $\mathbb{R}^m$ or more generally $\mathbb{R}^{p,q}$ is particularly simple to describe since the covariant derivative D_t^M is just the ordinary directional derivative.

Corollary 9.3.13. *Consider a pseudo-Riemannian submanifold $N \subset \mathbb{R}^{p,q}$ and a curve $\gamma \colon I \to N$. A vector field X on γ is parallel (on N) if and only if its derivative $X'(t)$ in $\mathbb{R}^{p,q}$ is orthogonal to $T_{\gamma(t)}N$ for every $t \in I$.*

Of course, orthogonality is to be intended with respect to the metric tensor $\eta_{p,q}$ of the pseudo-Riemannian manifold $\mathbb{R}^{p,q}$.

9.4. Exercises

Exercise 9.4.1. [Definition of $D_t X$ via axioms] Prove that there is a unique way to assign to any vector field X on a curve $\gamma \in I \to M$ another vector field $D_t X$ on γ such that

(1) If X and X' agree on a subinterval $J \subset I$, then $D_t X$ and $D_t X'$ do.
(2) The map D_t is linear on vector fields on γ.
(3) $D_t(fX) = f'X + f D_t X$ for any function $f\colon I \to \mathbb{R}$.
(4) If the restriction of X to a subinterval $J \subset I$ is induced by a vector field Y on an open subset of M, then $D_t X = \nabla_{\gamma'(t)} Y$ for all $t \in J$.

Exercise 9.4.2. Prove Proposition 9.3.11 by *defining* $\nabla_v X = \pi\big(\nabla^M_v X\big)$ and showing that the so obtained ∇ is a symmetric connection on N compatible with the metric g. By uniqueness of the Levi-Civita connection, $\nabla = \nabla^N$.

Exercise 9.4.3. Calculate the area of the following domain

$$[-a, a] \times [b, \infty)$$

in the half-plane model H^2 of hyperbolic space.

Exercise 9.4.4. Write the Euclidean metric g on $\mathbb{R}^2 \setminus \{0\}$ in polar coordinates (ρ, θ). Determine the Christoffel symbols of the Levi-Civita connection with respect to the variables (ρ, θ).

Exercise 9.4.5. Let (M, g) be a pseudo-Riemannian manifold. Prove that in any coordinates the following hold:

$$\begin{aligned}
\Gamma^j_{ji} &= \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^i} \sqrt{\det g}, \\
\operatorname{div}(X) &= \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^i} \left(X^i \sqrt{\det g} \right), \\
\Delta f &= \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^i} \left(\frac{\partial f}{\partial x^j} g^{ij} \sqrt{\det g} \right).
\end{aligned}$$

Here on Γ^j_{ji} we use the Einstein summation convention.

CHAPTER 10

Geodesics

We know that every pseudo-Riemannian manifold (M, g) has a preferred connection ∇, and now we use ∇ to define geodesics. In fact, a connection ∇ is enough to define the geodesics, the background metric g plays no role.

On a Riemannian manifold we finally respond to one of our primary motivations, by showing that geodesics are precisely the curves that minimise the path length, at least locally (not necessarily globally). Although geodesics are defined quite indirectly through ∇, their relation with g is very tight.

10.1. Geodesics

10.1.1. Definition. Let M be a manifold equipped with a connection ∇.

Definition 10.1.1. A smooth curve $\gamma\colon I \to M$ is a *geodesic* if the velocity field $\gamma'(t)$ is parallel along γ.

Recall that this means that $D_t\gamma' = 0$ for every $t \in I$. A quite simple (and not much exciting) example of geodesic is the constant map $\gamma(t) = p$, that has $\gamma'(t) = 0$ for al t. Such a geodesic is called *trivial* or *constant*.

Proposition 10.1.2. *Every non-trivial geodesic is an immersion.*

Proof. Since the field $\gamma'(t)$ is parallel, it is null at some $t \in I \iff$ it is null everywhere $\iff \gamma$ is trivial. $\square$

A geodesic is *maximal* if it is not the restriction of a longer geodesic $\eta\colon J \to M$ with $I \subsetneq J$. Geodesics have many nice properties; the first important one is that they exist, uniquely once a starting point and a direction are fixed:

Proposition 10.1.3. *For every $p \in M$ and $v \in T_pM$ there is a unique maximal geodesic $\gamma\colon I \to M$ with $0 \in I$, $\gamma(0) = p$, and $\gamma'(0) = v$.*

Proof. In coordinates, a curve $\gamma(t) = x(t)$ is a geodesic if and only if the following holds for all k, see (26):

$$\frac{d^2x^k}{dt^2} + \frac{dx^i}{dt}\frac{dx^j}{dt}\Gamma^k_{ij} = 0. \tag{32}$$

This is a second-order system of ordinary differential equations. The Cauchy–Lipschitz Theorem 1.3.5 ensures that the system has locally a unique solution with prescribed initial data $x(0) = p$ and $\frac{dx}{dt}(0) = v$. $\square$

We write the unique maximal geodesic γ tangent to $v \in T_pM$ at $t = 0$ as

$$\gamma_v \colon I_v \longrightarrow M.$$

Note that the interval domain $I_v \subset \mathbb{R}$ also depends on v. When $v = 0$ we get the trivial constant geodesic $\gamma_0 \colon \mathbb{R} \to M$, $\gamma_0(t) = p$.

This is a quite remarkable fact: a connection ∇ furnishes at every point p a canonical family of curves exiting from p at every possible direction like a firework starbust. The second-order system of differential equations (32) make sense in any coordinate system. These may be written as simply as

$$\ddot{x}^k + \dot{x}^i \dot{x}^j \Gamma_{ij}^k = 0. \tag{33}$$

To define geodesics we only need a connection ∇, not a pseudo-Riemannian metric. If ∇ is the Levi-Civita connection of a Riemannian metric g, the number $\langle \gamma'(t), \gamma'(t) \rangle$ is clearly independent of t all along the geodesic γ. If this number is positive, null, or negative, the geodesic is correspondingly a spacelike, timelike, or lightlike curve. The norm $\|\gamma'(t)\| = \sqrt{|\langle \gamma'(t), \gamma'(t) \rangle|}$ is constant.

On a Riemannian manifold every geodesic travels at constant speed. One may wonder if the same geodesic run at a different constant speed is still a geodesic. This is true thanks to the following more general fact, that holds for all connections ∇, without the need of a background metric.

Proposition 10.1.4. *If γ is a geodesic, then $\eta(t) = \gamma(ct)$ is also a geodesic, for every non-zero $c \in \mathbb{R}$.*

Proof. If $\nabla_v X = 0$, then also

$$\nabla_{cv} cX = c^2 \nabla_v X = 0.$$

This concludes easily the proof. □

In particular, we have $\gamma_{cv}(t) = \gamma_v(ct)$.

10.1.2. Examples. We study the geodesics in some pseudo-Riemannian manifolds encountered in the previous pages.

Example 10.1.5. On $\mathbb{R}^n$ with the Euclidean metric we have $\Gamma_{ij} = 0$ and hence the geodesics are precisely the straight lines $\gamma(t) = p + tv$. More generally, this holds also for $\mathbb{R}^{p,q}$, where the geodesic is timelike, lightlike, or spacelike according to the type of v.

Example 10.1.6. Let $N \subset \mathbb{R}^{p,q}$ be a Riemannian submanifold. By Corollary 9.3.13, a curve $\gamma \colon I \to N$ is a geodesic if and only if $\gamma''(t)$ is orthogonal to $T_{\gamma(t)}N$ for all $t \in I$.

Example 10.1.7. By the previous example, every maximal circle on S^n run at constant speed is a geodesic. In other words, for every $p \in S^n$, every unitary vector $v \in T_pS^n = p^\perp$, and every $c > 0$, the curve $\gamma \colon \mathbb{R} \to S^n$ defined as

$$\gamma(t) = \cos(ct) \cdot p + \sin(ct) \cdot v$$

is the maximal geodesic that starts from p in the direction v at speed c. To prove this it suffices to check that $\gamma(t) \in S^n$ and $\gamma''(t)$ is parallel to $\gamma(t)$, hence orthogonal to $T_{\gamma(t)}S^n$. By Proposition 10.1.3 these are precisely all the maximal geodesics in the sphere S^n.

Example 10.1.8. We have already remarked some analogies between S^n and the hyperboloid model I^n for the hyperbolic space. Using exactly the same argument as in the previous example (with the Lorentzian scalar product replacing the Euclidean one) we see easily that for every $p \in I^n$, every unitary vector $v \in T_pI^n = p^\perp$, and every $c > 0$, the curve $\gamma\colon \mathbb{R} \to I^n$,

$$\gamma(t) = \cosh(ct) \cdot p + \sinh(ct) \cdot v$$

is the maximal geodesic that starts from p in the direction v at speed c.

In both the previous examples the support of the geodesic γ is the intersection of S^n or I^n with the plane generated by p and v. We get a circle in S^n and a hyperbola in I^n.

Example 10.1.9. If we calculate the Christoffel symbols for the half-plane model H^2 of the hyperbolic space, with coordinates (x, y), we find (exercise)

$$\Gamma^1_{11} = \Gamma^2_{12} = \Gamma^2_{21} = \Gamma^1_{22} = 0, \quad \Gamma^2_{11} = \frac{1}{y}, \quad \Gamma^1_{12} = \Gamma^1_{21} = \Gamma^2_{22} = -\frac{1}{y}.$$

The geodesic equations are then

$$\ddot{x} - \frac{2}{y}\dot{x}\dot{y} = 0, \qquad \ddot{y} + \frac{1}{y}\big((\dot{x})^2 - (\dot{y})^2\big) = 0.$$

A family of solutions is

$$x = c, \qquad y = e^{dt}.$$

The supports are vertical lines and their speed is $|d|$. Another family is

$$x = \lambda \tanh dt + c, \qquad y = \lambda \frac{1}{\cosh dt}.$$

The supports are half-circles of equation $(x - c)^2 + y^2 = \lambda^2, y > 0$ and the speed is $|d|$. These two family of geodesics describe all the maximal geodesics of H^2 since any tangent vector at any point is tangent to one of these.

Summing up, the (supports of the) maximal geodesics of H^2 are vertical lines and half-circles orthogonal to the horizontal axis as in Figure 10.1.

Remark 10.1.10. Like the Levi-Civita connection, geodesics are natural, that is they are preserved by isometries. If $\varphi\colon M \to N$ is an isometry between pseudo-Riemannian manifolds, a curve $\gamma\colon I \to M$ is a geodesic $\Longleftrightarrow$ $\varphi \circ \gamma$ is.

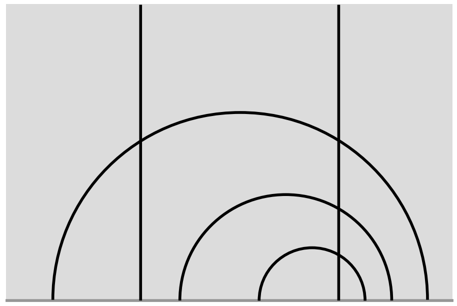

Figure 10.1. The supports of the maximal geodesics in the half-plane model H^2 of the hyperbolic plane are lines and half-circles orthogonal to the horizontal axis.

10.1.3. Geodesic flow. Let M be a smooth manifold equipped with a connection ∇. It would be nice if we could represent all the geodesics in M as the integral curves of some fixed vector field on M. However, this is clearly impossible! On a vector field, there is only one integral curve crossing each point p, but there are infinitely many geodesics through p, one for each direction $v \in T_pM$. To make this strategy work we only need to replace M with its tangent bundle TM.

We first note that using the derivative we may lift canonically every curve $\eta\colon I \to M$ to a curve $\eta'\colon I \to TM$, and that it makes sense to consider the second derivative $\eta''(t) \in T_{\eta'(t)}TM$.

Let M be equipped with a connection ∇. We define the *geodesic vector field* X in TM as follows: for every $v \in TM$, let $\gamma_v\colon I_v \to M$ be the unique maximal geodesic with $\gamma_v'(0) = v$. We define $X(v) = \gamma_v''(0)$.

The geodesic vector field X is smooth because the geodesic γ_v depends smoothly on the initial data. By construction its maximal integral curves are precisely all the lifts of all the maximal geodesics in M. The vector field X generates a flow Φ on TM called the *geodesic flow*. The flow Φ moves the points in TM along the lifted geodesics.

The geodesic flow Φ is defined on some maximal open subset U of $TM \times \mathbb{R}$ containing $TM \times \{0\}$. We have $U \cap \big(\{v\} \times \mathbb{R}\big) = \{v\} \times I_v$. With moderate effort, mostly relying on theorems proved in the previous chapters, we have defined a quite general and fascinating geometric flow on (the tangent bundle of) every manifold M equipped with a connection ∇.

10.1.4. Exponential map. We now define a useful map that is tightly connected with the geodesic flow, called the *exponential map*. We start by defining the following subset of the tangent bundle:

$$V = \big\{ v \in TM \bigm| 1 \in I_v \big\} \subset TM.$$

Recall that $I_v \subset \mathbb{R}$ is the domain of γ_v. In words, the set V consists of all the tangent vectors v such that the geodesic γ_v exists for at least 1 unit of time, so that $\gamma_v(1)$ makes sense. The *exponential map* is

$$\begin{aligned} \exp \colon V &\longrightarrow M \\ v &\longmapsto \gamma_v(1). \end{aligned}$$

For every $p \in M$ we define

$$V_p = V \cap T_pM, \qquad \exp_p = \exp|_{V_p}.$$

We see as usual M embedded in TM as the zero-section.

Proposition 10.1.11. *The domain V is an open neighbourhood of M and* $\exp$ *is smooth. Each V_p is open and star-shaped with respect to* 0. *We have*

$$\gamma_v(t) = \exp_p(tv)$$

for every $v \in TM$ and $t \in \mathbb{R}$ such that both members are defined.

Proof. Let U be the open domain of the geodesic flow Φ. We have $V = \big\{v \in TM \mid v \times \{1\} \in U\big\}$ and hence V is open. The map $\exp(v) = \pi(\Phi(v, 1))$ is smooth. Proposition 10.1.4 gives

$$\exp(tv) = \gamma_{tv}(1) = \gamma_v(t)$$

hence V_p is star-shaped and $\gamma_v(t) = \exp(tv)$. □

Here is one important fact about the exponential map:

Proposition 10.1.12. *The map* $\exp_p$ *is a local diffeomorphism at* $0 \in V_p$.

Proof. We have $d(\exp_p)_0 \colon T_0(T_pM) \to T_pM$, but since T_pM is a vector space we get $T_0(T_pM) = T_pM$ and hence $d(\exp_p)_0$ is an endomorphism. For every $v \in T_pM$ we have $\exp_p(tv) = \gamma_v(t)$ for all sufficiently small t. Therefore $d(\exp_p)_0(v) = \gamma_v'(0) = v$. We have proved that $d(\exp_p)_0 = \mathrm{id}$. In particular, it is invertible and hence $\exp_p$ is a local diffeomorphism at 0. □

The proposition says that the exponential map $\exp_p$ may be used as a parametrisation of a sufficiently small open neighbourhood of p. After many pages, we recover here a very intuitive idea: the tangent space T_pM should approximate the manifold near the point p. This idea may be realised concretely, via the exponential map, only after fixing a connection on M.

Example 10.1.13. In the space $\mathbb{R}^{p,q}$ the geodesics are just the Euclidean lines run at constant speed. Therefore $V = T\mathbb{R}^{p,q} = \mathbb{R}^{p,q} \times \mathbb{R}^{p,q}$ and

$$\exp \colon \mathbb{R}^{p,q} \times \mathbb{R}^{p,q} \longrightarrow \mathbb{R}^{p,q}, \qquad \exp(p, v) = p + v.$$

Example 10.1.14. Consider the sphere S^n. Example 10.1.7 shows that for this Riemannian manifold we have $V = TS^n$ and

$$\exp(v) = \cos\|v\| \cdot p + \sin\|v\| \cdot \frac{v}{\|v\|}$$

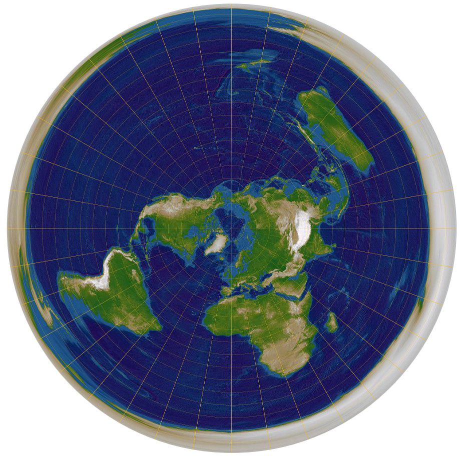

Figure 10.2. If we model the Earth as S^2 and look at the exponential map from the north pole N, the disc D of radius π in $T_N S^2$ is mapped to S^2 as shown here. The points in ∂D are all sent to the south pole.

for every $p \in S^n$ and $v \in T_p S^n$. Note that when $\|v\| = \pi$ we get $\exp(v) = -p$.

The map $\exp_p$ sends the open disc $B(0, \pi) \subset T_p M$ of radius π diffeomorphically onto $S^n \setminus \{-p\}$, while its boundary sphere $\partial B(0, \pi)$ goes entirely to the antipodal point $-p$. See Figure 10.2. Note in particular that $\exp_p$ is not a local diffeomorphism at the points in $\partial B(0, \pi)$. In general, it is guaranteed to be a local diffeomorphism only at the origin.

Example 10.1.15. On the hyperboloid model I^n of the hyperbolic space, Exercise 10.1.8 shows that $V = TI^n$ and

$$\exp(v) = \cosh \|v\| \cdot p + \sinh \|v\| \cdot \frac{v}{\|v\|}$$

for every $p \in I^n$ and $v \in T_p I^n$. The map $\exp_p \colon T_p I^n \to I^n$ is a diffeomorphism, with inverse

$$\exp_p^{-1}(q) = \frac{q + \langle p, q\rangle p}{\|q + \langle p, q\rangle p\|} \operatorname{arccosh}\big(-\langle p, q\rangle\big).$$

10.2. Normal coordinates

The exponential map furnishes some nice local parametrisations called *normal coordinates*, that we now investigate. These are extremely useful in many computations and play an important role in general relativity.

10.2.1. Definition. Consider a pseudo-Riemannian manifold (M, g) and pick a point $p \in M$. Recall that the exponential map is a local diffeomorphism at the origin of $T_p M$ and hence furnishes a diffeomorphism $\exp_p \colon U \to \exp_p(U)$ on some sufficiently small open neighbourhood $U \subset T_p M$ of 0.

If we fix an orthonormal basis at $T_p M$ we get an isometric identification of $T_p M$ with $\mathbb{R}^{p,q}$, the vector space $\mathbb{R}^n$ equipped with the diagonal metric tensor

$\eta_{p,q}$, where (p, q) is the signature of M. With this identification, the map $\exp_p$ is a parametrisation from an open neighbourhood $U \subset \mathbb{R}^{p,q}$ of 0 to an open neighbourhood $\exp_p(U)$ of p.

The neighbourhood $\exp_p(U)$ of p inherits from this parametrisation some coordinates $x^1, \ldots, x^n$ called *normal coordinates*, with $p = 0$. The normal coordinates are the best kind of coordinates that we can hope for near a fixed point p: they are uniquely determined up to a linear isometry of $\mathbb{R}^{p,q}$, since they only depend on the choice of an orthonormal basis in T_pM, and they have many nice properties that we now investigate.

10.2.2. Properties. The first notable feature of normal coordinates is the following, which follows from Proposition 10.1.11.

Proposition 10.2.1. *In normal coordinates, the geodesic emanated from the origin in the direction v is the Euclidean ray $\gamma_v(t) = tv$.*

The geodesics emanated from the origin are Euclidean rays run at constant speed. This is only valid at the origin! The geodesics that do not cross the origin are not necessarily Euclidean lines. One consequence is the following.

Proposition 10.2.2. *In normal coordinates we have*

$$g_{ij}(0) = \eta_{ij}, \qquad \frac{\partial g_{ij}}{\partial x^k}(0) = 0, \tag{34}$$

$$\Gamma^k_{ij}(0) = 0, \qquad \frac{\partial \Gamma^k_{ij}}{\partial x^l}(0) + \frac{\partial \Gamma^k_{jl}}{\partial x^i}(0) + \frac{\partial \Gamma^k_{li}}{\partial x^j}(0) = 0. \tag{35}$$

Proof. The geodesic equation (33) is satisfied by the curves $x(t) = tv$, $\forall v \in \mathbb{R}^n$. Plugging $x(t)$ in the equation we get

$$v^i v^j \Gamma^k_{ij}(0) = 0$$

for every $v \in \mathbb{R}^n$, and hence $\Gamma^k_{ij}(0) = 0$ since $\Gamma^k_{ij} = \Gamma^k_{ji}$ by linear algebra. By deriving the geodesic equation we get the third order equations

$$\dddot{x}^k + \ddot{x}^i \dot{x}^j \Gamma^k_{ij} + \dot{x}^i \ddot{x}^j \Gamma^k_{ij} + \dot{x}^i \dot{x}^j \frac{\partial \Gamma^k_{ij}}{\partial x^l} \dot{x}^l = 0.$$

If we substitute $x(t) = tv$ again we get

$$v^i v^j v^l \frac{\partial \Gamma^k_{ij}}{\partial x^l}(0) = 0$$

for every $v \in \mathbb{R}^n$. Exercise 2.7.3 now yields the last equality. Concerning g_{ij}, the first equality is a consequence of $d(\exp_p)_0 = \mathrm{id}$ and the second follows from $\Gamma^k_{ij}(0) = 0$ and (30). $\square$

Of course the Christoffel symbols Γ^k_{ij} and the derivatives of g_{ij} are guaranteed to vanish only at the origin, and not at the other nearby points. Indeed the second derivatives

$$\frac{\partial^2 g_{ij}}{\partial x^k \partial x^l}(0)$$

do not vanish in general. We will see that these quantities, that depend on four indices i, j, k, l, are linearly related with the components of a very important tensor field of type $(1,3)$ called the *Riemann tensor*, which is often non-zero.

We can see the normal coordinates as the result of a doomed attempt to find some coordinates near p where $g_{ij} = \eta_{p,q}$ everywhere. This is not possible, as we will see, and the best that we can get is that $g_{ij} = \eta_{p,q}$ up to the first order at a preferred point.

10.2.3. The Riemannian case: geodesic balls. On a Riemannian manifold it is possible, at every given point, to find some small neighbourhoods that look roughly (but not exactly!) like small Euclidean balls. These are called *geodesic balls*.

On any metric space X, for every point $p \in X$ and radius $r > 0$ we define as usual the *metric ball*

$$B(p,r) = \{x \in X \mid d(x,p) < r\}.$$

Let (M, g) be a Riemannian manifold, and $p \in M$ a point. The positive-definite $g(p)$ gives T_pM the structure of a metric space. Let $r > 0$ be sufficiently small such that the exponential map

$$\exp_p \colon B(0,r) \to M$$

is defined and is an embedding. The image $\exp_p\big(B(0,r)\big)$ in M is called the *geodesic ball of radius r centred at p*.

A geodesic ball is indeed diffeomorphic to a ball, because it is the diffeomorphic image of $B(0,r)$. If we fix an orthonormal basis for T_pM, we get an identification $T_pM = \mathbb{R}^n$ and some normal coordinates on the geodesic ball. In normal coordinates, the geodesic ball is like the Euclidean ball $B(0,r) \subset \mathbb{R}^n$ with a metric tensor g_{ij} that varies smoothly on $x \in B(0,r)$, and that is equal to the Euclidean δ_{ij} only at first order at $p = 0$.

Fare esempio sfera?

10.2.4. The Gauss Lemma. We now prove a fundamental lemma on geodesic balls that has many important geometric consequences. Its proof is non-trivial and very instructive since it uses many of the subtle properties of the Levi-Civita connection that were studied in the previous pages.

Let (M, g) be a Riemannian manifold and $p \in M$ a point. Consider a geodesic ball $B = \exp_p(B(0,r))$. We recall that $r > 0$ is implicitly small enough so that $\exp_p$ is well-defined and an embedding on $B(0,r)$.

For every $0 < r' < r$, we let the *geodesic sphere of radius r'* be the image

$$S(p,r') = \exp_p(\partial B(0,r')).$$

Like geodesic balls, geodesic spheres exist only for small $r' > 0$, and they are indeed diffeomorphic (but often not isometric) to Euclidean spheres. Recall that the geodesics emanated from the origin are Euclidean rays.

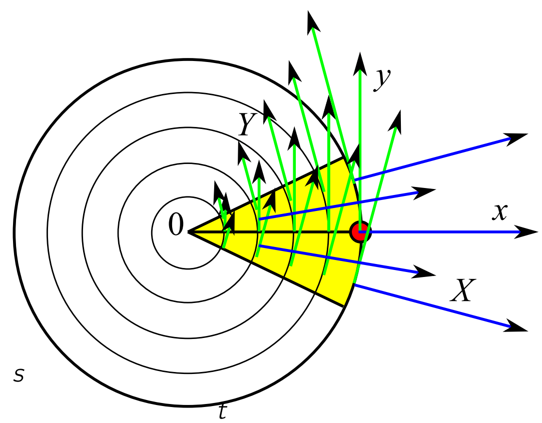

Figure 10.3. The Gauss Lemma says that the vectors x and y are orthogonal. To prove this, we extend x and y to two commuting vector fields X (blue) and Y (green). Then we show that $\langle X, Y\rangle$ is constant along the rays, and hence vanishes everywhere.

Lemma 10.2.3 (Gauss Lemma). *The Euclidean rays emanated from the origin and the geodesic spheres are everywhere orthogonal.*

It is worth saying that normal coordinates are not conformal in general, that is the angles induced by g are usually not equal to the Euclidean angles: this is only guaranteed to be true for the angles between rays and spheres.

Proof. We use normal coordinates and identify the geodesic ball B with $B(0,r) \subset \mathbb{R}^n$. Consider a ray s generated by some $x \in S^{n-1}$ and a geodesic sphere $S = \partial B(0, r')$ with $r' < r$, intersecting at the point $r'x$. Of course

$$T_{r'x}s \oplus T_{r'x}S = T_{r'x}\mathbb{R}^n = \mathbb{R}^n.$$

The line $T_{r'x}s$ is generated by x itself, and let $y \in T_{r'x}S$ be any vector. We need to prove that $g(r'x)(x,y) = 0$. Let $W \subset \mathbb{R}^n$ be the linear subspace generated by the vectors x and y. By restricting everything to $B(0,r) \cap W$ we may suppose for simplicity that we are in dimension $n = 2$.

Write $B^* = B \setminus \{0\}$. We may suppose that the Euclidean norm of y is r'. Up to rotating everything we get $x = (1,0)$ and $y = (0, r')$, both considered as tangent vectors at $r'x = (r', 0)$. We can extend x and y to the vector fields

$$X = \frac{\partial}{\partial \rho}, \qquad Y = \frac{\partial}{\partial \theta}$$

in B^*, see Figure 10.3. These are the vector fields induced by the polar coordinates ρ, θ. We denote the scalar product $g(p)$ by $\langle , \rangle$, omitting p.

Our aim is to prove that

$$\frac{\partial}{\partial \rho}\langle X, Y\rangle = 0$$

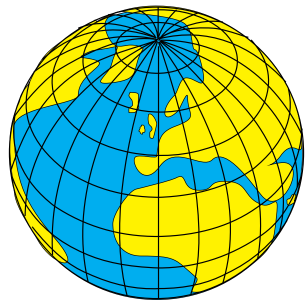

Figure 10.4. The Geometric Gauss Lemma says that geodesic spheres centered at a point p are orthogonal to the geodesics exiting from p. On a sphere, when p is the north pole it says that parallels are orthogonal to meridians, as we all know.

on every ray. When $X, Y \to 0$, we have that $\|X\|$ stays bounded while $Y \to 0$, so $\langle X, Y\rangle \to 0$. These two facts altogether imply that $\langle X, Y\rangle = 0$ everywhere in B^*, so in particular $g(r'x)(x, y) = \langle X, Y\rangle(r'x) = 0$ as required.

Since $\frac{\partial}{\partial\rho}$ and $\frac{\partial}{\partial\theta}$ are induced by some coordinates change, they commute, and hence $[X, Y] = 0$. Using Proposition 9.3.5 we find

$$\frac{\partial}{\partial\rho}\langle X, Y\rangle = \langle \nabla_X X, Y\rangle + \langle X, \nabla_X Y\rangle.$$

We have $\nabla_X X = 0$ everywhere because the radii are geodesics. Since $[X, Y] = 0$ and the torsion vanishes, we get $\nabla_X Y = \nabla_Y X$ and therefore

$$\frac{\partial}{\partial\rho}\langle X, Y\rangle = \langle X, \nabla_Y X\rangle = \frac{1}{2}\frac{\partial}{\partial\theta}\langle X, X\rangle = 0$$

because $\langle X, X\rangle = 1$ everywhere (the radii are geodesics, hence the radial norms are equal to the Euclidean norm). The proof is complete. $\square$

The Gauss Lemma says that the metric tensor at every point x in the geodesic ball decomposes orthogonally into a radial part that coincides with the Euclidean metric, and a tangential part, tangent to the geodesic sphere, that may be quite arbitrary.

Figures 10.4 and 10.5 show the Gauss Lemma in action on the sphere and on the hyperbolic plane. On S^2, the geodesic spheres centered at the north pole are clearly the parallels. On H^2, we have the following.

Exercise 10.2.4. The geodesic sphere at $(x, y) \in H^2$ of radius $r > 0$ is the Euclidean circle with centre $(x, y\cosh r)$ and Euclidean radius $y \sinh r$.

Figure 10.5 shows some geodesic rays and spheres emanating from a fixed point. We can appreciate visually that they are all orthogonal. Remember that H^2 is a conformally equivalent to the Euclidean plane, so angles are represented correctly (whereas lengths are not).

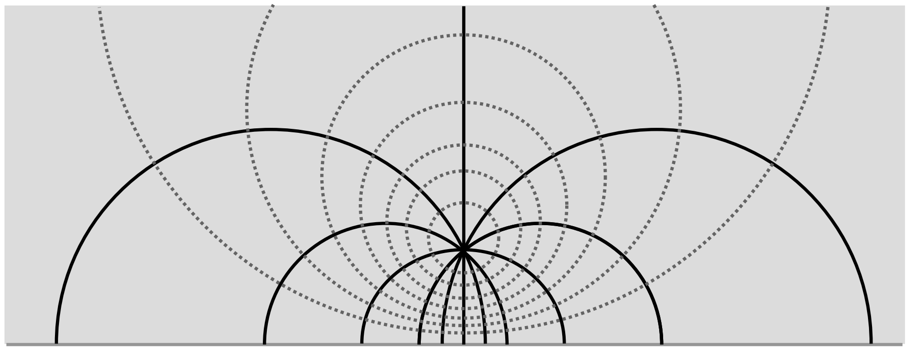

Figure 10.5. The Gauss Lemma on the half-plane model H^2 of the hyperbolic plane. Thick curves are geodesics, dotted curves are geodesic spheres. These two families of curves meet at right angles.

10.2.5. Minimising curves. We now start to study the tight connection between geodesics and distance between points.

Let M be a Riemannian manifold and $p, q \in M$ two points. We are interested in the smooth curves that connect p to q, that is the $\gamma\colon [a,b] \to M$ with $\gamma(a) = p$ and $\gamma(b) = q$. Recall that the length $L(\gamma)$ of γ is independent of its parametrisation. Recall also that $d(p,q)$ is the infimum of all the lengths of all the smooth curves connecting p and q. This infimum may not be realised in some cases; if it does, that is if there is a curve γ with $L(\gamma) = d(p,q)$, then the curve γ is called *minimising*.

Exercise 10.2.5. If γ is minimising, the restriction of γ to any closed subinterval $[c,d] \subset [a,b]$ is also minimising, that is $L(\gamma|_{[c,d]}) = d(\gamma(c), \gamma(d))$.

Let $p \in M$ a point. Let $B = \exp_p(B(0,r)) \subset M$ be a geodesic ball centred at p with radius r, and $q \in B$ be any other point. We know that B contains a unique radial geodesic $\gamma_{p,q}\colon [0,1] \to B$ connecting p to q.

Proposition 10.2.6. *The geodesic $\gamma_{p,q}$ is a minimising curve. Every other minimising curve in M connecting p to q is obtained by reparametrising $\gamma_{p,q}$.*

Proof. The point q belongs to the geodesic sphere $S = \exp_p(\partial B(0,r'))$ with $r' = L(\gamma_{p,q})$. Every curve γ in M connecting p to q contains an initial subcurve η supported in the closure of the geodesic ball $\exp_p(B(0,r'))$, that connects 0 to some point $q' \in S$ that may be different from q.

By the Gauss Lemma, the velocity $\eta'(t)$ decomposes g-orthogonally into a radial component $\eta'(t)_r$ (parallel to the rays exiting from p) and a tangential component (tangent to the geodesic spheres) $\eta'(t)_t$. We get

$$L(\gamma) \geq L(\eta) = \int \|\eta'(t)_r + \eta'(t)_t\| \geq \int \|\eta'(t)_r\| \geq r' = L(\gamma_{p,q}).$$

The third (in-)equality holds because the two vectors are g-orthogonal.

Therefore $L(\gamma) \geq L(\gamma_{p,q})$, and the equality holds if and only if $\eta = \gamma$ and $\eta'(t)_t = 0$, $\eta'(t)_r > 0$ $\forall t$, that is if and only if $\gamma(t)$ is obtained by reparametrising $\gamma_{p,q}$. □

Here are two geometric corollaries.

Corollary 10.2.7. *The geodesic sphere of radius r centred at p consists precisely of all the points in M at distance r from p.*

Corollary 10.2.8. *The geodesic ball of radius r centred at p consists precisely of all the points in M at distance $< r$ from p.*

Recall that geodesic spheres and balls exist only for sufficiently small r. In particular for every $p \in M$ and sufficiently small r we have

$$\exp_p\big(B(0,r)\big) = B(p,r).$$

For large r we only have the inequality

$$\exp_p\big(B(0,r)\big) \subset B(p,r).$$

10.2.6. Totally normal neighbourhoods. Let M be a Riemannian manifold. We have discovered that every point $p \in M$ has a neighbourhood U that is nice with respect to p, and now we want to be more democratic and show that we may pick a U that is also nice with respect to every point $q \in U$.

We say that an open subset $U \subset M$ is *totally normal* if for every $q \in U$ there is a geodesic ball centred at q containing U.

Example 10.2.9. On the Euclidean $\mathbb{R}^n$ every bounded open set is a totally normal neighbourhood. On the sphere S^n, every open set that does not contain any pair of antipodal points is a totally normal neighbourhood.

Proposition 10.2.10. *Every $p \in M$ has a totally normal neighbourhood U.*

Proof. Recall that $\exp\colon V \longrightarrow M$ is defined on some open neighbourhood $V \subset TM$ of M. We consider the map

$$\begin{aligned} F\colon \quad V &\longrightarrow M \times M \\ (p,v) &\longmapsto (p, \exp_p(v)). \end{aligned}$$

For every $p \in M$ we have $F(p,0) = (p,p)$ and a natural isomorphism $T_{(p,0)}V = T_pM \times T_pM$, so that $dF_{(p,0)}$ is an endomorphism. The differential

$$dF_{(p,0)} = \begin{pmatrix} \mathrm{id} & 0 \\ * & d(\exp_p)_0 \end{pmatrix} = \begin{pmatrix} \mathrm{id} & 0 \\ * & \mathrm{id} \end{pmatrix}$$

is invertible and hence F is a local diffeomorphism at $(p,0)$. Therefore there are a neighbourhood W of p and a $\delta > 0$ such that the restriction of F to

$$W' = \big\{(p,v) \bigm| p \in W, \|v\| < \delta\big\}$$

is a diffeomorphism onto its image $F(W')$. In particular the metric ball $B(q,\delta)$ is a geodesic ball for all $q \in W$.

Pick a neighbourhood $U \subset W$ of p such that $U \times U \subset F(W')$. By construction $U \subset B(q,\delta)$ for all $q \in U$. $\square$

10.2.7. Locally minimising curves. We have defined the geodesics as the solutions of certain differential equations, and we can finally characterise them using only the distance between points.

Let M be a Riemannian manifold. We say that a curve $\gamma\colon I \to M$ is *locally minimising* if every $t \in I$ has a compact neighbourhood $[t_0, t_1] \subset I$ such that the restriction $\gamma|_{[t_0,t_1]}$ is minimising. By Exercise 10.2.5, a minimising curve is also locally minimising, but the converse does not hold in general.

Being locally minimising and being a geodesic are both local properties of a curve. In fact, they are the same property, up to reparametrising.

Theorem 10.2.11. *A curve $\gamma\colon I \to M$ is locally minimising $\Longleftrightarrow$ it is obtained by reparametrising a geodesic.*

Proof. For every $t \in I$, pick a totally normal neighbourhood U containing $\gamma(t)$ and let $[t_0, t_1] \subset I$ be a neighbourhood of t such that $\gamma([t_0, t_1]) \subset U$. There is a geodesic ball $B(\gamma(t_0), r)$ containing U and hence $\gamma([t_0, t_1])$. We apply Proposition 10.2.6 with $p = \gamma(t_0)$ and $q = \gamma(t_1)$, and get that $\gamma|_{[t_0,t_1]}$ is minimising if and only if it is a reparametrised geodesic. $\square$

The theorem is also true for piecewise smooth curves (see Remark 9.2.8), since by means of transition functions we can reparametrise them as smooth curves that have velocity zero at the angles. Geodesics are precisely the locally minimising curves, in a very robust manner.

10.2.8. Convex subsets. We can extend the usual notion of convexity to all Riemannian manifolds.

Definition 10.2.12. A subset $S \subset M$ of a Riemannian manifold M is *convex* if any two points p, q in S are joined by a unique (up to reparametrisation) minimising geodesic γ in M. It is *strictly convex* if any two points p, q in the closure $\bar{S}$ are joined by a unique (up to reparametrisation) minimising geodesic whose interior is contained in the interior of S.

Example 10.2.13. On the Euclidean space $\mathbb{R}^n$, these definitions match with the usual notions of (strict) convexity. On the sphere S^n, the geodesic ball $B(p, r)$ is convex when $r \leq \pi/2$ and strictly convex when $r < \pi/2$.

We will prove that geodesic balls of sufficiently small radius are strictly convex. To this purpose, we need the following.

Lemma 10.2.14. *For every point $p \in M$ there is a $r_0 > 0$ such that $B(p, r_0)$ is a geodesic ball, and every geodesic tangent to the geodesic sphere $S(p, r)$ stays locally outside $B(p, r)$, for every $0 < r < r_0$.*

As an example, on S^n the maximum r_0 with this property is $r_0 = \pi/2$.

Proof. Use normal coordinates, that is represent a geodesic ball $B(p, \bar{r})$ as $B(0, \bar{r}) \subset \mathbb{R}^n$ for some small $\bar{r} > 0$. For every $(x, v) \in B(0, \bar{r}) \times S^{n-1}$ pick the geodesic $\gamma_{x,v} \colon J_{x,v} \to B(0, \bar{r})$ with $0 \in J_{x,v}$, $\gamma_{x,v}(0) = x$, and $\gamma'_{x,v}(0) = v$. Consider the map

$$F(x, v) = \frac{\partial^2}{\partial t^2}\big(\|\gamma_{x,v}(t)\|^2\big)\Big|_{t=0}.$$

Here $\|\cdot\|$ is the Eucldean norm. When $x = 0$, the geodesic is radial $\gamma_{0,v}(t) = tv$ and hence $F(0, v) = 2$. By continuity there is a $0 < r_0 < \bar{r}$ such that $F(x, v) > 0$ for all $\|x\| \leq r_0$, and hence $\|\gamma_{x,v}(t)\|^2$ has a strict local minimum at $t = 0$, whenever $\|x\| \leq r_0$. □

We can now prove that small geodesic balls are strictly convex.

Proposition 10.2.15. *For every point $p \in M$ there is a $r_0 > 0$ such that $B(p, r)$ is a strictly convex geodesic ball, for every $0 < r < r_0$.*

Proof. We know that there is a $r_1 > 0$ such that $B(p, r_1)$ is a geodesic ball and every geodesic tangent to the geodesic sphere $S(p, r)$ stays locally outside $B(p, r)$, for every $0 < r \leq r_1$.

Using totally normal neighbourhoods we can find a $0 < r_0 < r_1/2$ such that every pair of points $q, q' \in \overline{B(p, r_0)}$ has a unique minimising geodesic $\gamma_{q,q'}$ of length at most $r_1/2$. (We can do this because on a totally normal neighbourhood the minimising geodesic, and hence its length, varies smoothly on the points.)

In particular $\gamma_{q,q'}$ is contained in $B(p, r_1)$. If we represent $B(p, r_1)$ in normal coordinates, we see that the maximum of $\|\gamma_{q,q'}(t)\|^2$ must be at one of its endpoints, otherwise $\gamma_{q,q'}(t)$ would be tangent to a geodesic sphere locally from inside. Therefore $B(p, r)$ is strictly convex for every $r \leq r_0$. □

Convex subsets have two nice properties: they are closed under intersection, and contractible (exercise). These imply the following:

Proposition 10.2.16. *Every smooth manifold M has a locally finite open covering $\{U_i\}$ where every non-empty intersection of U_i's is contractible.*

Proof. Put an arbitrary metric on M and use convex neighbourhoods. □

10.2.9. Injectivity radius. Let (M, g) be a Riemannian manifold. The *injectivity radius* $\mathrm{inj}_p M$ at a point $p \in M$ is the supremum of all $r > 0$ such that the restriction of $\exp_p$ to $B(0, r)$ is defined and is an embedding.

Proposition 10.2.17. *The supremum is actually a maximum.*

Proof. By Exercise 3.12.10 "embedding" is equivalent to "injective immersion". If the restriction of $\exp_p$ to $B(0, r_i)$ is an injective immersion for $r_i \to r$, then the restriction to $B(0, r)$ also is. □

We get a function $\text{inj}: M \to (0, +\infty]$ that sends p to $\text{inj}_p M$. The *injectivity radius* $\text{inj}M$ of M is the infimum of this function.

Example 10.2.18. The injectivity radius of S^n is π, since inj is constantly π at every point $p \in S^n$. The injectivity radius of $\mathbb{R}^n$ and hyperbolic space is $+\infty$, since it is so at every point. See Examples 10.1.13, 10.1.14, and 10.1.15.

Proposition 10.2.19. *The function* inj *is lower semi-continuous, that is*

$$\liminf_{p_i \to p} \text{inj}_{p_i} M \geq \text{inj}_p M.$$

Proof. If the restriction of $\exp_p$ to the compact set $\overline{B(0, r)}$ is defined and an embedding, then $\exp_{p_i}$ also is for every p_i sufficiently close to p by Proposition 5.7.9, which applies (with the same proof) also to manifolds with boundary. □

Corollary 10.2.20. *The function* inj *has a positive minimum on every compact subset $K \subset M$. In particular, if M is compact then* $\text{inj}M > 0$.

Compactness is necessary here. On $M = \mathbb{R}^n \setminus \{0\}$ with the Euclidean metric, we have $\text{inj}_x M = \|x\|$ and hence $\text{inj}M = 0$.

10.2.10. The pseudo-Riemannian case. We now extend some of the previous results to the more general pseudo-Riemannian case. We will furnish also an enhanced version of the Gauss Lemma that is valid on the whole domain of the exponential map.

Despite the absence of distances and geodesic balls, we may still define a reasonable notion of totally normal neighbourhood on pseudo-Riemannian manifolds.

Definition 10.2.21. Let M be a pseudo-Riemannian manifold. An open subset $Z \subset M$ is a *normal neighbourhood* at $p \in Z$ if there is an open star-shaped neighbourhood $U \subset T_pM$ of 0 where $\exp_p$ is defined and an embedding, and with $Z = \exp_p(U)$.

If M is not Riemannian, an open subset $Z \subset M$ is *totally normal* if it is a normal neighbourhood at every point $q \in Z$.

On a Riemannian manifold this notion of totally normal neighbourhood is slightly different from the one given in the previous pages, so to avoid ambiguity we restricted it to pseudo-Riemannian manifolds that are not Riemannian.

Theorem 10.2.22. *Every $p \in M$ has a totally normal neighbourhood.*

Proof. We already know this for a Riemannian M, so we stick to the non-Riemannian case.

Pick some normal coordinates $x^1, \ldots, x^n$ on a normal neighbourhood of p. Consider for small $r > 0$ the open ball $B(r) = \{(x^1)^2 + \cdots + (x^n)^2 < r\}$. We claim that if r if sufficiently small the ball $B(r)$ is totally normal.

To prove this we recall that $\exp\colon V \to M$ is defined on some open neighbourhood $V \subset TM$ of M. We consider the map

$$\begin{aligned} F\colon \quad V &\longrightarrow M \times M \\ (p, v) &\longmapsto (p, \exp_p(v)). \end{aligned}$$

Using normal coordinates and the identification $T_x B(r) = \mathbb{R}^n$ for all $x \in B(r)$ we may write F locally near $(p, 0)$ as

$$F\colon B(r') \times B(r') \longrightarrow B(r) \times B(r)$$

for a sufficiently small $r' > 0$. We have already proved that $dF_{(p,0)}$ is invertible and hence F is a local diffeomorphism at $(p, 0)$. Therefore there are a neighbourhood $W \subset B(r') \times B(r')$ of $(0, 0)$ and $r'' < r$ such that F sends diffeomorphically W to $B(r'') \times B(r'')$.

Since $\Gamma_{ij}^k(0) = 0$, after taking an even smaller $r'' > 0$ if necessary we may suppose that the symmetric matrix

$$\delta_{ij} - \sum_{k=1}^{n} \Gamma_{ij}^k x^k$$

is positive definite at every $x \in B(r'')$. We prove that $B(r'')$ is totally normal.

For every $q \in B(r'')$ we set $W_q = W \cap T_q M$. Since $F(W_q) = \{q\} \times B(r'')$, we have $\exp_q(W_q) = B(r'')$. We conclude by showing W_q is star-shaped. Pick $v \in W_q \subset B(r')$. The geodesic $x(t) = \gamma_v(t) = F(p, tv)$ lies in $B(r)$ for all $t \in [0, 1]$, and we have $x(0) = q$, $x(1) = q' \in B(r'')$. We find

$$\begin{aligned} \frac{d^2\big(\|x(t)\|^2\big)}{dt^2} &= 2\big(\langle \dot{x}, \dot{x}\rangle + \langle \ddot{x}, x\rangle\big) = 2\big(\langle \dot{x}, \dot{x}\rangle - \langle \dot{x}^i \dot{x}^j \Gamma_{ij}^k e_k, x\rangle\big) \\ &= 2\big(\delta_{ij} - \Gamma_{ij}^k x^k\big)\dot{x}^i \dot{x}^j > 0. \end{aligned}$$

Therefore $\|x(t)\|^2$ reaches its maximum at one of its endpoints, and hence $x(t) \in B(r'')$ for all $t \in [0, 1]$. We deduce that $tv \in W_q$ for all $t \in [0, 1]$. □

Totally normal subsets Z are useful to study geodesics. First of all, they furnish a safety zone where geodesics cannot be killed nor trapped: when extended, these must hit the boundary ∂Z in both directions in finite time.

Proposition 10.2.23. *Inside a totally normal set Z, every geodesic can be extended either for all times, or until it reaches a point in the topological boundary ∂Z.*

Proof. Given a non-trivial geodesic γ in Z, consider a point p lying in γ. Now $Z = \exp_p(U)$ for a star-shaped $U \subset T_p M$, and we can reparametrise $\gamma(t) = \exp_p(tv)$ as a radial geodesic with respect to p, hence we conclude. □

Corollary 10.2.24. *A geodesic $\gamma\colon (a, b) \to M$ is extendible at $b \iff$ it is continuously extendible at b.*

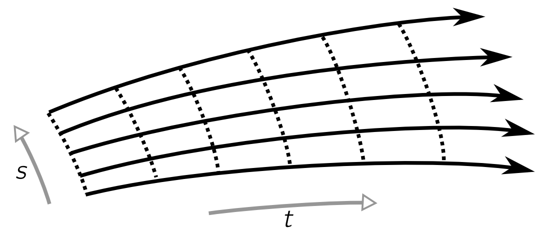

Figure 10.6. A (particularly nice) example of smooth family of curves $\gamma_s(t)$.

Proof. ($\Rightarrow$) is obvious, so we turn to ($\Leftarrow$). If γ extends to a continuous map $\gamma\colon (a,b] \to M$, pick a totally normal neighbourhood Z of $p = \gamma(b)$. We have $\gamma([c,b]) \subset Z$ for some $a < c < b$. By Proposition 10.2.23 the geodesic γ can be prolonged after b since $\gamma(b) \notin \partial Z$. □

Remark 10.2.25. Let $\gamma\colon (a,b) \to M$ be a maximal geodesic. If $b < +\infty$, then $\gamma'(t) \in TM$ must diverge (exit from any compact subset) as $t \to b$, see Section 1.3.7. On a Riemannian manifold, we know that $\|\gamma'(t)\|$ is constant, so we can easily deduce that $\gamma(t)$ must diverge, and from this fact we obtain Corollary 10.2.24 without using totally normal neighbourhoods. On a more general pseudo-Riemannian manifold this argument is fallacious: there are compact pseudo-Riemannian manifolds with geodesics γ that do not extend to $\mathbb{R}$, see Exercise 10.6.1. There $\gamma'(t)$ diverges while $\gamma(t)$ does not.

Proposition 10.2.26. *Let $Z \subset M$ be a totally normal set. For any $p, q \in Z$, there is a unique geodesic $\gamma_{p,q}\colon [0,1] \to Z$ with $\gamma_{p,q}(0) = p, \gamma_{p,q}(1) = q$.*

Proof. The set Z is a normal neighbourhood for p, hence every point $q \in Z$ is connected radially to p by a unique geodesic. □

10.2.11. Family of curves.

To prove the enhanced version of the Gauss Lemma we will need to study some smooth families of curves and vector fields along them.

Definition 10.2.27. A *family of curves* is a smooth map $f\colon (-\varepsilon,\varepsilon) \times I \to M$ where $I \subset \mathbb{R}$ is some interval.

We write $\gamma_s(t) = f(s,t)$ and think of it as a family of curves γ_s depending on s. We only require f to be smooth, so both the curves and the way they vary can be pretty complicated in general. If f is an embedding, its image is a surface $S \subset M$ that we can visualise as a nice disjoint family of embedded curves as in Figure 10.6. If $df_{(s,t)}$ is injective at some point (s,t), then f is locally an embedding and we get this picture at least for the points near (s,t).

A *vector field* along f is a smooth map $X\colon (-\varepsilon,\varepsilon) \times I \to TM$ such that $X(s,t) \in T_{f(s,t)}M$ for every (s,t). This is like having a vector field on each curve γ_s, that varies smoothly with s. Two important examples are

$$S(s,t) = df_{(s,t)}\left(\frac{\partial}{\partial s}\right), \qquad T(s,t) = df_{(s,t)}\left(\frac{\partial}{\partial t}\right).$$

These are the tangent vector fields of the curves $f(\cdot, t)$ and $f(s, \cdot) = \gamma_s$. We call S and T the *coordinate vector fields* of f. If f is an embedding, its image is a surface $S \subset M$ and a vector field along f may be interpreted simply as a tangent vector field on S. This interpretation works locally near every point (s,t) such that $df_{(s,t)}$ is injective.

Let M be equipped with a symmetric connection ∇. Let X be a vector field along f. As we did for curves, we can now define the covariant derivatives of X along the variables s and t. We let $D_tX(s,t)$ be the covariant derivative of $X(s,\cdot)$ along the curve $f(s,\cdot)$, and $D_sX(s,t)$ be the covariant derivative of $X(\cdot,t)$ along $f(\cdot,t)$. Both D_sX and D_tX are new vector fields along f.

Lemma 10.2.28. *If ∇ is symmetric, we get*

$$D_tS = D_sT.$$

Proof. If f is an embedding, then S and T are vector fields on the image surface S. Since these are f-related with the commuting coordinate fields $\frac{\partial}{\partial s}$ and $\frac{\partial}{\partial t}$, we get $[S,T] = 0$, and hence $\nabla_T S = \nabla_S T$ by the symmetry of the connection. The proof is complete.

On a more general f, we work in coordinates. Now f has image in $\mathbb{R}^n$ and

$$S(s,t) = \frac{\partial f}{\partial s}, \qquad T(s,t) = \frac{\partial f}{\partial t}.$$

Therefore

$$D_tS = D_t\left(\frac{\partial f}{\partial s}\right) = \frac{\partial^2 f}{\partial t \partial s} + \frac{\partial f^i}{\partial t}\frac{\partial f^j}{\partial s}\Gamma_{ij}^k e_k.$$

By symmetry $\Gamma_{ij}^k = \Gamma_{ji}^k$ and hence we get the same expression for D_sT. □

10.2.12. The enhanced Gauss Lemma. Let now (M, g) be a pseudo-Riemannian manifold. At every p we have the exponential map $\exp_p\colon V_p \to M$, defined on some open star-shaped subset $V_p \subset T_pM$. For every $v \in T_pM$ we identify $T_v(T_pM) = T_pM$ canonically (since T_pM is a vector space). By assigning the same scalar product $g(p) = \langle , \rangle$ to each tangent space we get a pseudo-Riemannian structure on T_pM and hence on V_p.

Both V_p and M are pseudo-Riemannian manifolds, and $\exp_p\colon V_p \to M$ is not an isometry in general: as we will see, the curvature of M is responsible for that. In some sense, the exponential map is an isometry only *radially*. This is precisely the content of the following important result, see Figure 10.7.

Lemma 10.2.29 (Enhanced Gauss Lemma). *For every $v \in V_p$ we have*

$$\langle (d\exp_p)_v(v), (d\exp_p)_v(w) \rangle = \langle v, w \rangle \qquad \forall w \in T_v(T_pM) = T_pM$$

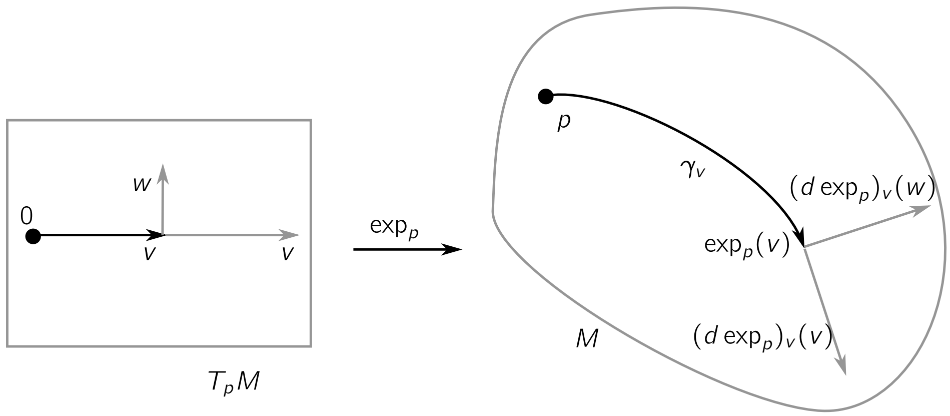

Figure 10.7. The Gauss Lemma says that $\exp_p$ is a kind of radial isometry: the scalar products with the radial vectors v are preserved, but the map may distort in the directions w orthogonal to the radial vector v. Both $\langle v, v\rangle$ and $\langle v, w\rangle$ are preserved, but $\langle w, w\rangle$ may not be.

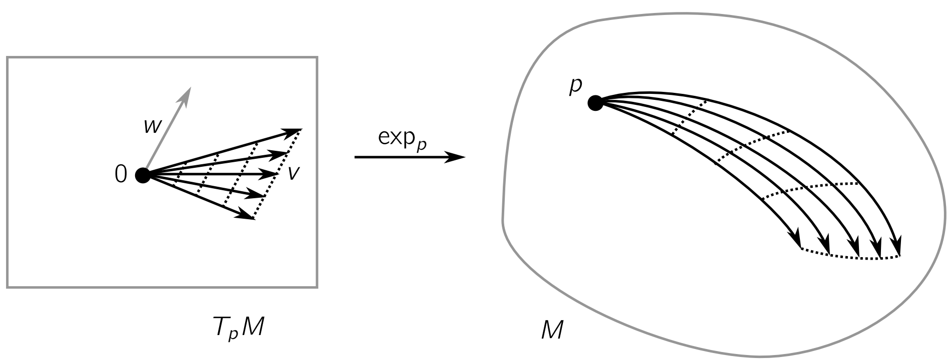

Figure 10.8. The family of curves used to prove the Gauss Lemma. Each γ_s is a geodesic exiting from p.

Proof. Consider the family of curves $f\colon (-\varepsilon, \varepsilon) \times [0, 1] \to M$,

$$f(s, t) = \exp_p\big(t(v + sw)\big).$$

See Figure 10.8. Here $\epsilon > 0$ is small enough so that $tv_s \in V_p \, \forall s, t$. Let S and T be the coordinate vector fields along f. The curve $\gamma_s = f(s, \cdot)$ is the geodesic with initial velocity $v + sw$. Therefore $D_t T = 0$ and $\langle T, T\rangle = \langle v + sw, v + sw\rangle$. By Lemma 10.2.28, we get

$$\frac{\partial}{\partial t}\langle S, T\rangle = \big\langle D_t S, T\big\rangle + \big\langle S, D_t T\big\rangle = \langle D_s T, T\rangle = \frac{1}{2}\frac{\partial}{\partial s}\langle T, T\rangle = \langle v, w\rangle.$$

Since $\langle S, T\rangle(0, 0) = \langle 0, v\rangle = 0$, we deduce that $\langle S, T\rangle(t, 0) = t\langle v, w\rangle$, and therefore $\langle S, T\rangle(1, 0) = \langle v, w\rangle$. This is in fact the thesis. $\square$

10.3. Completeness

A Riemannian manifold M is also a metric space, so it makes perfectly sense to consider whether it is *complete* or not – a notion that is meaningless for unstructured smooth manifolds. We prove here the *Hopf – Rinow Theorem*, that shows that completeness may actually be stated in multiple equivalent ways, one of which involves only geodesics.

10.3.1. Geodesically complete manifolds. Let M be a manifold. A connection ∇ on M is *geodesically complete* if every maximal geodesic $\gamma(t)$ in M is defined for all times $t \in \mathbb{R}$. A pseudo-Riemannian manifold is geodesically complete if its Levi-Civita connection is.

Example 10.3.1. The pseudo-Riemannian manifolds $\mathbb{R}^{p,q}$, S^n, and $\mathbb{H}^n$ are geodesically complete. See Section 10.1.2.

Let M be a Riemannian manifold. We say that M is *complete* if its underlying metric space is. This notion is not present on more general pseudo-Riemannian manifolds, since no reasonable distance is defined on them.

Recall that the distance $d(p, q)$ of two points $p, q \in M$ is the infimum of the lengths of all the curves γ joining p and q; if such an infimum is realised by γ, then γ is called *minimising* and we have discovered in the last section that a minimising curve γ must be a geodesic (up to a reparametrisation). Here is one nice consequence of geodesical completeness:

Proposition 10.3.2. *If a Riemannian manifold M is connected and geodesically complete, every two points $p, q \in M$ are joined by a minimising geodesic.*

Proof. Pick a geodesic ball $B(p, r)$ at p, with geodesic sphere $S(p, r)$. If $q \in B(p, r)$ we are done. Otherwise, let $p_0 \in S(p, r)$ be a point at minimum distance from q. Let $v \in T_pM$ be the unique unit vector such that $\gamma_v(r) = p_0$.

By hypothesis, the geodesic $\gamma_v(t) = \exp_p(tv)$ exists for all $t \in \mathbb{R}$. Set $d = d(p, q)$. We now show that $\gamma_v(d) = q$. To do so, let $I \subset [0, d]$ be the subset of all times t such that $d(\gamma_v(t), q) = d - t$. This set is non-empty and closed, and using Theorem 10.2.11 we deduce that it is also open (exercise). Therefore $I = [0, d]$ and we are done. $\square$

Corollary 10.3.3. *If a Riemannian manifold M is connected and geodesically complete, the exponential map $\exp_p \colon T_pM \to M$ is surjective at every $p \in M$. Moreover for every $r > 0$ we have*

$$\exp_p(B(0, r)) = B(p, r). \tag{36}$$

Here $B(p, r)$ is the metric ball, and recall that it is a geodesic ball only for sufficiently small r. When r is big, the metric ball is of course not necessarily homeomorphic to a ball. When M is not complete, the equality (36) is guaranteed only for small $r > 0$, and only the inclusion $\subset$ is valid in general.

10.3.2. The Hopf – Rinow Theorem. The following important theorem says that different notions of completeness are actually equivalent.

Theorem 10.3.4 (Hopf – Rinow). *Let M be a connected Riemannian manifold. The following are equivalent:*

(1) *M is geodesically complete.*
(2) *A subset $K \subset M$ is compact $\Longleftrightarrow$ it is closed and bounded.*
(3) *M is complete.*

Proof. (1)$\Rightarrow$(2). Let $K \subset M$ be a subset. Compact always implies closed and bounded, so we prove the converse. Take a point $p \in M$. If K is bounded, there is a $r > 0$ such that $K \subset B(p, r) = \exp_p(B(0, r))$, where in the last equality we use that M is geodesically complete and Corollary 10.3.3. Hence K is contained in the compact set $\exp_p(\overline{B(0, r)})$. If K is closed, it is also compact.

(2)$\Rightarrow$(3). Every Cauchy sequence is bounded, so it has compact closure. Therefore it contains a converging subsequence, and hence it converges.

(3)$\Rightarrow$(1). Let $\gamma \colon I \to M$ be a maximal geodesic. We know that I is open, and since M is complete we prove that it is also closed. If $t_i \in I$ converges to some $t^* \in \mathbb{R}$, then $\gamma(t_i)$ is a Cauchy sequence (because γ is a Lipschitz map!) and hence converges to some $p \in M$. We pick a chart near p and note that both $\gamma(t)$ and $\gamma'(t)$ stay both bounded as $t \to t^*$, so the solution extends past t^* and we get $t^* \in I$, see Section 1.3.7. $\square$

Corollary 10.3.5. *Every compact Riemannian manifold is geodesically complete.*

Quite surprisingly, this fact is no longer true for general pseudo-Riemannian manifolds: see Exercise 10.6.1. Here is another non-trivial corollary.

Corollary 10.3.6. *Every closed submanifold N of a geodesically complete Riemannian manifold M is also geodesically complete.*

Proof. The inclusion map $N \hookrightarrow M$ is always 1-Lipschitz. Therefore, if N is closed and M is complete, we easily deduce that N is complete. $\square$

Corollary 10.3.7. *Every smooth manifold has a geodesically complete Riemannian metric.*

Proof. By Whitney's Embedding Theorem, every smooth manifold is diffeomorphic to a closed submanifold of $\mathbb{R}^n$. $\square$

The following simple criterion is natural and useful.

Proposition 10.3.8. *Let M be a connected Riemannian manifold and $p \in M$ a point. If $\exp_p$ is defined on the whole of T_pM then M is complete.*

Proof. The proof of Proposition 10.3.2 applies *as is*, to show that every $q \in M$ is joined to p by a minimising geodesic. With this in hand, we can follow the (1)$\Rightarrow$(2) proof of the Hopf – Rinow Theorem and deduce that a subset $K \subset M$ is compact $\Longleftrightarrow$ it is closed and bounded. Hence by the Hopf – Rinow Theorem again M is complete. $\square$

10.4. Isometries

A smooth manifold M has plenty of self-diffeomorphisms – many can be constructed for instance by taking the flows of arbitrary complete vector fields. On the contrary, a pseudo-Riemannian manifold has typically few non-trivial self-isometries – very often none.

We prove here that isometries are "rigid," in the sense that they are determined by their first order behaviour at any point. Using this we classify all the isometries of the spaces $\mathbb{R}^{p,q}$, S^n, and H^n.

By combining smooth coverings and local isometries we get the powerful notion of pseudo-Riemannian covering. We introduce some notable examples.

10.4.1. Rigidity. The following theorem says that every isometry is determined by its first-order behaviour at any point of the (connected) domain.

Proposition 10.4.1. *Let $f, g\colon M \to N$ be two isometries between pseudo-Riemannian manifolds. If M is connected, and there is a $p \in M$ such that*

$$f(p) = g(p), \qquad df_p = dg_p,$$

then $f = g$.

Proof. Let $U \subset M$ consist of all points $q \in M$ such that $f(q) = g(q)$ and $df_q = dg_q$. The set U is clearly closed, and we now prove that it is also open. Since $p \in U$ and M is connected, we deduce that $U = M$ and we are done.

We prove that U is open. Pick $q \in U$. Isometries send geodesics to geodesics, so for every $v \in T_qM$ we get

$$f \circ \gamma_v = \gamma_{df_q(v)}.$$

Since $df_q = dg_q$, we deduce that $f = g$ on the support of γ_v. By varying v and using that $\exp_q$ is a local diffeomorphism at the origin we deduce that $f = g$ on an open neighbourhood Z of q. Therefore $Z \subset U$ and we are done. $\square$

10.4.2. Action on frames. We now want to study the manifolds that have a particularly high degree of symmetries.

Let M be a connected pseudo-Riemannian manifold. A *frame* on M is the datum of a point $q \in M$ and an orthonormal basis $v_1, \ldots, v_n$ of T_qM, ordered such that $\langle v_i, v_i \rangle$ is 1 if $i \leq p$ and -1 if $i > p$. (The signature is $(p, n-p)$.)

The isometry group $\mathrm{Isom}(M)$ of M acts naturally on its frames. The action is free by Proposition 10.4.1. In some natural sense, the manifolds with "the highest degree of symmetries" are those where this action is transitive. On such manifolds, for any pair of frames there is a unique isometry sending the first to the second. Here are some important examples.

Proposition 10.4.2. *The isometry groups of $\mathbb{R}^{p,q}$, S^n, $\mathbb{H}^n$ are:*

$$\begin{aligned}
\mathrm{Isom}(\mathbb{R}^{p,q}) &= \{x \mapsto Ax + b \mid A \in \mathrm{O}(p,q), b \in \mathbb{R}^{p,q}\} \\
\mathrm{Isom}(S^n) &= \mathrm{O}(n+1), \\
\mathrm{Isom}(\mathbb{H}^n) &= \mathrm{O}^+(n,1).
\end{aligned}$$

For $\mathbb{H}^n$ we use the hyperboloid model I^n. In all these cases the isometry group acts transitively on the frames.

Proof. Using linear algebra we see that the proposed groups are indeed isometries and act transitively on frames (exercise). Since they act transitively on frames, they form the whole isometry group (because its action is free). $\square$

10.4.3. Homogeneous and isotropic manifolds. We have seen that the most symmetric pseudo-Riemannian manifolds are those whose isometry groups act transitively on frames. We now introduce some weaker symmetry requirements that are also very interesting.

Let M be a connected pseudo-Riemannian n-manifold. We say that M is *homogeneous* if for any pair of points $p, q \in M$ there is an isometry of M sending p to q. We say that M is *isotropic* at some point p if for every pair of vectors $v, w \in T_pM$ with $\langle v, v\rangle = \langle w, w\rangle$ there is an isometry fixing p whose differential at p sends v to w. The manifold M is *isotropic* if it is so at every point $p \in M$.

Of course if $\mathrm{Isom}(M)$ acts transitively on frames then M is both homogeneous and isotropic. We propose a few instructing exercises. Let M be a Riemannian manifold. Pensare al caso pseudo-Riemanniano

Exercise 10.4.3. If M is homogeneous, it is complete.

Exercise 10.4.4. If M is isotropic, it is homogeneous.

Exercise 10.4.5. If M is isotropic at a single point $p \in M$ it is not necessarily homogeneous (construct a counterexample).

A pseudo-Riemannian manifold M is *locally homogeneous* if for every two points $p, q \in M$ there is an isometry $\varphi \colon U(p) \to V(q)$ of some of their neighbourhoods $U(p)$ and $V(q)$ sending p to q. Similarly M is *locally isotropic* at $p \in M$ if there is an open neighbourhood $U(p)$ of p such that $g|_{U(p)}$ is isotropic at p. The manifold M is locally isotropic if it is so at every $p \in M$.

We will see that manifolds with constant sectional curvature are locally homogeneous and locally isotropic. These manifolds belong to a wider class of objects called *locally symmetric spaces* that we will study in the next pages.

10.4.4. Pseudo-Riemannian coverings. Every time we introduce some structure on manifolds, we get a corresponding notion of covering. Pseudo-Riemannian structures make no exception.

A *(pseudo-)Riemannian covering* is a smooth map $\pi\colon M \to N$ between (pseudo-)Riemannian manifolds that is both a smooth covering and a local isometry. Much of the machinery introduced in Section 3.5 for smooth coverings adapt to pseudo-Riemannian coverings with the same (omitted) proofs.

Structures can be lifted: every time we have a topological covering $\pi\colon \tilde{M} \to M$ and M has a (pseudo-)Riemannian structure, this structure lifts from M to $\tilde{M}$ so that π is promoted to a (pseudo-)Riemannian covering.

Quite conversely, if M is a (pseudo-)Riemannian manifold and $\Gamma < \mathrm{Isom}(M)$ acts freely and properly discontinuously, the quotient $M/_\Gamma$ has a unique structure of a (pseudo-)Riemannian manifold such that the projection $\pi\colon M \to M/_\Gamma$ is a (pseudo)-Riemannian covering.

Example 10.4.6. The group $\mathbb{Z}^n < \mathrm{Isom}(\mathbb{R}^n)$ of translations acts freely and properly discontinuously, so the quotient torus $T^n = \mathbb{R}^n/_{\mathbb{Z}^n}$ inherits the structure of a Riemannian manifold.

Analogously the lens spaces $L(p,q) = S^3/_\Gamma$ introduced in Section 3.5.6 inherit a Riemannian structure from S^3 since $\Gamma < \mathrm{O}(4) = \mathrm{Isom}(S^4)$. The resulting map $S^3 \to L(p,q)$ is a Riemannian covering.

We leave a couple of exercises. Let M and N be connected pseudo-Riemannian manifolds of the same dimension.

Exercise 10.4.7. Let $f\colon M \to N$ be a local isometry. If M is geodesically complete, then f is a pseudo-Riemannian covering.

Hint. Prove that normal open subsets of N are well covered by lifting geodesics from N to M. □

Exercise 10.4.8. Let $f\colon M \to N$ be a pseudo-Riemannian covering. Show that M is geodesically complete $\Longleftrightarrow$ N is geodesically complete.

10.4.5. Killing vector fields. Let (M,g) be a pseudo-Riemannian manifold. We defined the Lie derivative $\mathcal{L}$ of tensor fields in Section 5.4.8.

Definition 10.4.9. A *Killing vector field* on M is a vector field X such that

$$\mathcal{L}_X g = 0.$$

Remember that a vector field X gives rise to a flow $\Phi\colon U \to M$ defined on a maximal domain $U \subset M \times \mathbb{R}$. We set $U_t = \{p \in M \mid (p,t) \in U\}$ and get a diffeomorphism $\Phi_t\colon U_t \to \Phi_t(U)$ by setting $\Phi_t(p) = \Phi(p,t)$.

Proposition 10.4.10. *A vector field X is Killing $\Longleftrightarrow$ Φ_t is an isometry $\forall t$.*

Proof. We have $\mathcal{L}_X g = 0 \Longleftrightarrow g$ is invariant under the flow, that is each Φ_t is an isometry. □

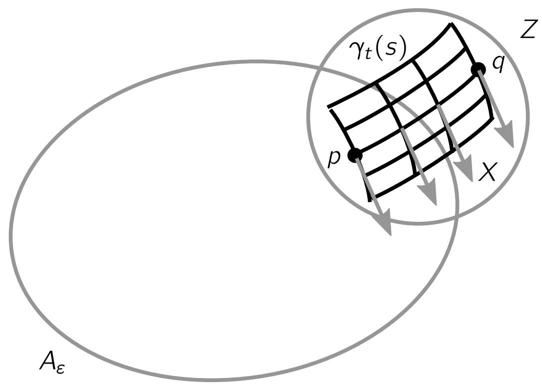

Figure 10.9. If the flow is defined for $(-\epsilon, \epsilon)$ on A_ε, it is so also on Z.

Example 10.4.11. On $\mathbb{R}^{p,q}$ every constant vector field is a Killing field; the flow consists of translations. On $\mathbb{R}^2$, another Killing vector field is $X(x, y) = (-y, x)$; its flow consists of rotations around the origin. On S^2 the vector field $X(x, y, z) = (-y, x, z)$ is Killing; the flow consists of rotations around the z axis. The same X is Killing on the hyperboloid model I^2 of hyperbolic space.

We are interested in Killing vector fields because they gives rise to a one-parameter family of isometries Φ_t, defined on some open set U_t. If the Killing vector field X is complete, we get a one-parameter family $\Phi_t \in \mathrm{Isom}(M)$ of isometries for M. In general, a Killing vector field may not be complete! For instance, pick any non-trivial constant vector field X on a proper open subset $V \subset \mathbb{R}^n$ of Euclidean space: here the isometries Φ_t never extend to V.

Here is a simple criterion on M that guarantees the completeness of every Killing vector field.

Proposition 10.4.12. *If M is geodesically complete, every Killing vector field X on M is complete.*

Proof. We may suppose M connected. Let $A_\varepsilon \subset M$ be the set of points p where the integral curve starting from p exists at least on $(-\varepsilon, \varepsilon)$. We pick $\varepsilon > 0$ with $A_\varepsilon \neq \emptyset$ and show that $A_\varepsilon = M$. This suffices by Lemma 5.2.4.

Let $Z \subset M$ be a totally normal subset. We show that if $A_\varepsilon \cap Z = \emptyset$, then $Z \subset A_\varepsilon$. This easily implies that $A_\varepsilon = M$. Pick $p \in A_\varepsilon \cap Z$ and $q \in Z$. They are joined by a geodesic $\gamma\colon [0, 1] \to M$. Using the flow Φ_t of X we define

$$\gamma_t(s) = \Phi_t(\gamma(s))$$

for every (s, t) lying in the maximal subset $U \subset [0, 1] \times \mathbb{R}$ where this quantity is defined. See Figure 10.9. Since $p = \gamma(0) \in A_\varepsilon$ we have $0 \times (-\varepsilon, \varepsilon) \subset U$. Since $\gamma_0 = \gamma$ and each Φ_t is an isometry, each γ_t is a geodesic where it is defined. Since X is complete, geodesics are actually defined everywhere and hence $[0,1]\times(-\varepsilon,\varepsilon)\subset U$. Therefore $q=\gamma(1)\in A_\varepsilon$. $\square$

Proposition 10.4.13. *The following are equivalent for a vector field X:*

(1) *X is Killing.*
(2) *$X\langle V,W\rangle = \langle [X,V],W\rangle + \langle V,[X,W]\rangle$ for any local vector fields V,W.*
(3) *∇X is a g-skew-adjoint $(1,1)$-tensor field, that is*

$$\langle \nabla_v X, w\rangle + \langle v, \nabla_w X\rangle = 0 \qquad \forall v,w\in T_pM\ \forall p\in M.$$

(4) *$\langle \nabla_v X, v\rangle = 0$ for all $v\in T_pM$, $\forall p\in M$.*

Proof. The field X is Killing $\Leftrightarrow \mathcal{L}_X g = 0 \Leftrightarrow (\mathcal{L}_X g)(V,W)=0$ for any local (*i.e.* defined on some open subset) vector fields V,W. By Exercise 5.4.14

$$\mathcal{L}_X(g(V,W)) = (\mathcal{L}_X g)(V,W) + g(\mathcal{L}_X V,W) + g(V,\mathcal{L}_X W).$$

Therefore, using again Exercise 5.4.14 at various points, we find The proof is complete. $\square$

Here is one nice concrete applications of Killing vector fields.

Proposition 10.4.14. *If X is Killing and $\gamma\colon I\to M$ is a geodesic, then*

$$\langle \gamma', X\rangle = C$$

is constant for all $t\in I$.

Proof. By deriving it we get

$$\langle D_t\gamma', X\rangle + \langle \gamma', D_t X\rangle = \langle \gamma', \nabla_{\gamma'} X\rangle = 0$$

by Proposition 10.4.13-(4). $\square$

In presence of a Killing vector field, one may try to use this simple equation to determine the geodesics directly, without calculating the Christoffel symbols.

Example 10.4.15. The *Schwarzschild half-plane* is the following Lorentzian surface (P,g). Fix $M>0$. Use the coordinates (r,t) for $\mathbb{R}^2$ and consider the half-plane $P=\{r>2M\}$. Set $h(r)=1-2M/r$ and

$$g=\begin{pmatrix} 1/h(r) & 0 \\ 0 & -h(r)\end{pmatrix}.$$

Since g depends only on r, the vector field $\frac{\partial}{\partial t}$ is Killing and $\langle \gamma', \frac{\partial}{\partial t}\rangle = C$ is constant on every geodesic γ. We now classify for the lightlike geodesics $\gamma(s)=(r(s),t(s))$. These must satisfy

$$\frac{(\dot r)^2}{h(r)} - (\dot t)^2 h(r) = 0, \qquad -h(r)\dot t = C.$$

These equations can be solved, and one finds that the lightlike geodesics are

$$\gamma(s) = \big(s+2M, \pm(s+2M\ln s)+c\big)$$

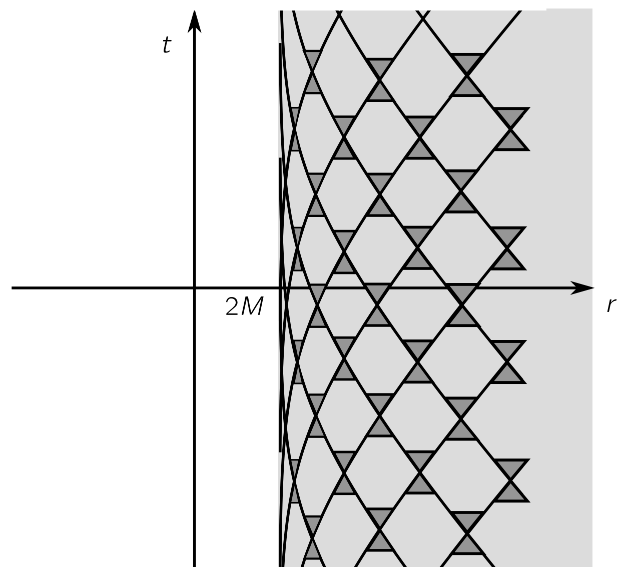

Figure 10.10. The lightlike geodesics on the Scharzschild half-plane, together with the light cones at some points.

where $c \in \mathbb{R}$. By drawing the lightlike geodesics as in Figure 10.10 we get a visual understanding of the Schwarzschild half-plane.

Exercise 10.4.16. Rediscover the geodesics of the half-space model H^2 of hyperbolic space (already determined in Exercise 10.1.9) using Killing vector fields, without computing the Christoffel symbols.

If X and Y are Killing vector fields on (M, g), then $[X, Y]$ also is, because of Exercise 5.9.3. This shows that the Killing vector fields form a subalgebra of the Lie algebra $\mathfrak{X}(M)$ of all vector fields on M.

10.5. Gradient, divergence, Laplacian, and Hessian

The reader who has read the various chapters on smooth manifolds may have felt deprived of some of the analytic concepts that were familiar in the study of functions and vector fields in $\mathbb{R}^n$, like gradient, divergence, Laplacian, and Hessian. We can finally define all of them on a pseudo-Riemannian manifold (M, g), using g and the Levi-Civita connection ∇.

10.5.1. Gradient. Let (M, g) be a pseudo-Riemannian manifold. The differential df of a function f is a tensor field of type $(0, 1)$, and by raising its index we get a vector field $\mathrm{grad} f$ called the *gradient* of f. In coordinates:

$$(\mathrm{grad} f)^i = g^{ij}(df)_j = g^{ij}\frac{\partial f}{\partial x^j}.$$

Of course this is the usual gradient on the Euclidean $\mathbb{R}^n$.

Exercise 10.5.1. Let $f\colon M \to \mathbb{R}$ be a map. If $c \in \mathbb{R}$ is a regular value, the gradient of f is everywhere orthogonal to the level submanifold $f^{-1}(c)$.

10.5.2. Divergence. Let M be a manifold equipped with a connection ∇. Using the connection we can define the *divergence* of various tensor fields, actually without needing any metric tensor, by first taking their covariant derivative, and then contracting the new index with an old one.

In coordinates, the divergence of a tensor field T of type (h, k) is

$$\nabla_i T^{i_1,\ldots,i_{j-1},i,i_{j+1},\ldots,i_h}_{j_1,\ldots,j_k}.$$

The operation is possible only when $h \geq 1$, and in case $h \geq 2$ it depends on the position j of the upper index that is contracted.

As an example, the divergence of a vector field X is the smooth function

$$\mathrm{div}(X) = \nabla_i X^i.$$

In coordinates

$$\mathrm{div}(X) = \nabla_i X^i = \frac{\partial X^i}{\partial x^i} + X^j \Gamma^i_{ij}.$$

On the Euclidean $\mathbb{R}^n$ this is the usual divergence. On a more general pseudo-Riemannian manifold (M, g), the divergence maintains the fundamental property it has in the Euclidean space: it measures at the first order how the volume changes along the flow of X. This is shown in the following proposition. Let M be oriented, and let ω be the volume form derived from g.

Proposition 10.5.2. *We have the following equality*

$$\mathrm{div}(X)\omega = \mathcal{L}_X(\omega)$$

Proof. This equality must be proved for every $p \in M$. We use normal coordinates at p. By the Cartan magic formula the right term equals

$$\begin{aligned} d\iota_X\omega &= d\left(\sum_{i=1}^n (-1)^{i-1} X^i |\det g_{jk}| dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n\right) \\ &= \sum_{i=1}^n \frac{\partial}{\partial x^i}\left(X^i |\det g_{jk}|\right) dx^1 \wedge \cdots \wedge dx^n \\ &= \sum_{i=1}^n \frac{\partial X^i}{\partial x^i} |\det g_{jk}| dx^1 \wedge \cdots \wedge dx^n = \sum_{i=1}^n \frac{\partial X^i}{\partial x^i}\omega = \nabla_i X^i \omega = \mathrm{div} X\omega. \end{aligned}$$

In the third equality we used that $\frac{\partial g}{\partial x^i} = 0$ and hence $\frac{\partial \det g}{\partial x^i} = 0$, in the last that the Christoffel symbols vanish and hence the covariant derivative at $p = 0$ equals the directional derivative. □

In this proof (and in more that will follow) we used normal coordinates to prove the equality of two given tensor fields. In normal coordinates many computations simplify considerably; in particular the Christoffel symbols vanish and so the covariant derivatives at p magically reduce to the directional ones.

Proposition 10.5.3. *If X is a vector field and f a function on M, we get*

$$\mathrm{div}(fX) = f\mathrm{div}(X) + g(\mathrm{grad}f, X).$$

Proof. By the Leibniz rule:

$$\mathrm{div}(fX) = \nabla_i(fX^i) = (df)_i X^i + f\nabla_i(X^i) = g_{ij}(\mathrm{grad}f)^j X^i + f\mathrm{div}(X).$$

The proof is complete. □

During the proof of Proposition 10.5.2 we noticed that $\mathrm{div}(X)\omega = d(\iota_X\omega)$ is an exact n-form. We now would like to apply Stokes' Theorem, and to this purpose we briefly introduce boundaries in the realm of Riemannian geometry.

10.5.3. Pseudo-Riemannian manifolds with boundary. The whole theory of pseudo-Riemannian manifolds and of connections extends to manifolds M with boundary with the appropriate modifications. The few adjustments that are to be made are usually straightforward: the metric tensor g is defined on the whole of M, the theorems (like the existence of normal coordinates) are still valid at the interior points of M, and sometimes also at the boundary points after the appropriate modifications. For instance, given a point $p \in \partial M$ and a vector $v \in T_pM$, there is a unique geodesic γ_v starting from p with direction v only if the vector v points towards the interior of M. To preserve clarity, *a manifold is always intended to be boundaryless except when mentioned explicitly*.

Like any submanifold, the boundary ∂M of a pseudo-Riemannian manifold with boundary may inherit a structure of pseudo-Riemannian manifold if the restriction of g to T_pM is nowhere degenerate (this is always guaranteed if M is Riemannian). If ∂M is a pseudo-Riemannian manifold, it comes equipped with an *outward normal field* ν, a vector field in M with support in ∂M defined by requiring that $\nu(p)$ be the only unit vector (that is, $g(\nu(p), \nu(p)) = \pm 1$) that lies in the outward half-space of T_pM cut by $T_p\partial M$. If M is oriented, then ∂M gets an orientation as well and it is hence also equipped with a volume form $\omega_{\partial M}$, induced by $g|_{\partial M}$. At every $p \in M$ we get

$$\omega(p)(\nu(p), v_1, \ldots, v_{n-1}) = \omega_{\partial M}(p)(v_1, \ldots, v_{n-1})$$

for every $v_1, \ldots, v_{n-1} \in T_pM$.

Theorem 10.5.4 (Divergence theorem). *Let X be a compactly supported vector field on an oriented pseudo-Riemannian manifold M with (possibly empty) boundary. Then*

$$\int_M \mathrm{div}(X)\omega = \int_{\partial M} g(X, \nu)\omega_{\partial M}.$$

Proof. By Cartan's magic formula $\mathrm{div}(X)\omega = \mathcal{L}_X(\omega) = d\iota_X\omega$. By Stokes

$$\int_M \mathrm{div}(X)\omega = \int_M d\iota_X\omega = \int_{\partial M} \iota_X\omega = \int_{\partial M} g(X, \nu)\omega_{\partial M}.$$

The proof is complete. □

10.5.4. Divergence and codifferential of k-forms. We have encountered the codifferential δ of k-forms in Section 7.5.7, and we now show that it coincides in fact with the divergence (up to raising an index). Let (M, g) be an oriented pseudo-Riemannian manifold.

Proposition 10.5.5. *In coordinates, for every $\alpha \in \Omega^k(M)$ we have*

$$(\delta\alpha)_{i_1,\ldots,i_{k-1}} = -\nabla_j \alpha^j{}_{i_1,\ldots,i_{k-1}}$$

Proof. Recall that $\delta\alpha = (-1)^{kn+n+1+m} * d * \alpha$ where g has segnature (p, m). Let us use normal coordinates. By linearity we may suppose that

$$\alpha = f\, dx^1 \wedge \cdots \wedge dx^k.$$

From Exercise 2.5.3 we deduce that

$$*\alpha = f\frac{\sqrt{|\det g|}}{(n-k)!} g^{1j_1}\cdots g^{kj_k}\epsilon_{j_1\cdots j_n} dx^{i_{k+1}} \wedge \cdots \wedge dx^{i_n},$$

$$d * \alpha = \frac{\partial}{\partial x^l}\left(f\frac{\sqrt{|\det g|}}{(n-k)!} g^{1j_1}\cdots g^{kj_k}\right)\epsilon_{j_1\cdots j_n} dx^l \wedge dx^{i_{k+1}} \wedge \cdots \wedge dx^{i_n}.$$

In normal coordinates, when we evaluate everything at 0 we simply get

$$d * \alpha = (-1)^{m'}\frac{\partial f}{\partial x^l} dx^l \wedge dx^{k+1} \wedge \cdots \wedge dx^n$$

where m' is the number of -1's among $g^{11}, \ldots, g^{kk}$. Finally

$$\begin{aligned} *d * \alpha &= \sum_{l=1}^k (-1)^m g^{ll}(-1)^{(n-k)(k-1)+l-1}\frac{\partial f}{\partial x^l} dx^1 \wedge \cdots \wedge \widehat{dx^l} \wedge \cdots \wedge dx^k \\ &= (-1)^{kn+n+1+m}\sum_{l=1}^k (-1)^l g^{ll}\frac{\partial f}{\partial x^l} dx^1 \wedge \cdots \wedge \widehat{dx^l} \wedge \cdots \wedge dx^k, \\ \delta\alpha &= \sum_{l=1}^k (-1)^l g^{ll}\frac{\partial f}{\partial x^l} dx^1 \wedge \cdots \wedge \widehat{dx^l} \wedge \cdots \wedge dx^k. \end{aligned}$$

Therefore

$$(\delta\alpha)_{i_1,\ldots,i_{k-1}} = -\sum_{l=1}^k g^{ll}\frac{\partial f}{\partial x^l}\epsilon_{l,i_1,\ldots,i_{k-1}}$$

where as usual $\epsilon_{j_1,\ldots,j_k}$ is zero, except when $(j_1, \ldots, j_k)$ is a permutation of $(1, \ldots, k)$, and in this case it is the sign of the permutation.

On the other hand, at the origin in normal coordinates we have

$$-\nabla_j \alpha^j{}_{i_1,\ldots,i_{k-1}} = -\sum_{l=1}^n g^{ll}\frac{\partial}{\partial x^l}\alpha_{l,i_1,\ldots,i_{k-1}} = -\sum_{l=1}^k g^{ll}\frac{\partial f}{\partial x^l}\epsilon_{l,i_1,\ldots,i_{k-1}}.$$

The proof is complete. □

For the sake of completeness, we write an analogous formula for the differential d. The formula shows that, although d is defined without using g, it can be recovered from ∇ in a quite reasonable way: the differential d is the antisymmetric part of ∇, times a constant $k+1$.

Exercise 10.5.6. In coordinates, for every $\alpha \in \Omega^k(M)$ we have

$$(d\alpha)_{i_1,\ldots,i_{k+1}} = (k+1)\nabla_{[i_1}\alpha_{i_2,\ldots,i_{k+1}]}.$$

Corollary 10.5.7. *A k-form α is closed $\Longleftrightarrow$ $\nabla\alpha$ is symmetric.*

Recall that a k-form ω is harmonic if $\Delta\omega = 0$, equivalently if $d\omega = 0$ and $\delta\omega = 0$. Every vector field X induces a 1-form $\omega = g(X,\cdot)$, with $\omega_i = g_{ij}X^j$.

Corollary 10.5.8. *The 1-form ω is harmonic $\Longleftrightarrow$ $\mathrm{div}X = 0$ and ∇X is a g-self-adjoint tensor field of type $(1,1)$.*

10.5.5. The Laplacian. The *Laplacian* Δf of a function $f \in C^\infty(M)$ was already defined in Section 7.5.8 as

$$\Delta f = (\delta d + d\delta)f = \delta d f.$$

The same Laplacian may be defined in a more familiar way as a composition of the gradient and the divergence (with a minus sign):

Proposition 10.5.9. *We have $\Delta f = -\mathrm{div}(\mathrm{grad} f)$.*

Proof. In coordinates

$$\delta d f = -\nabla_j (df)^j = -\nabla_j (\mathrm{grad} f)^j.$$

The proof is complete. □

Proposition 10.5.10. *In coordinates we get*

$$\Delta f = -g^{ij}\left(\frac{\partial^2 f}{\partial x^i \partial x^j} - \frac{\partial f}{\partial x^k}\Gamma^k_{ij}\right).$$

Proof. We find

$$\Delta f = -\nabla_i (\mathrm{grad} f)^i = -\nabla_i (g^{ij}(df)_j) = -g^{ij}\nabla_i (df)_j$$

whence the formula. We have used that $\nabla g^{ij} = 0$. □

This is the usual Laplacian on the Euclidean $\mathbb{R}^n$. By applying Proposition 10.5.3 with $X = \mathrm{grad}h$, we find that for any pair of functions $f, h \in C^\infty(M)$:

$$\mathrm{div}(f\,\mathrm{grad}h) = -f\Delta h + g(\mathrm{grad}f, \mathrm{grad}h) \tag{37}$$

We can now integrate by parts like in the familiar Euclidean $\mathbb{R}^n$:

Proposition 10.5.11 (Green's formula). *Let $f, h \in C^\infty(M)$ be functions on an oriented Riemanian manifold M with (possibly empty) boundary. Then*

$$\int_M f\Delta h = \int_M g(\mathrm{grad}f, \mathrm{grad}h) - \int_{\partial M} g(\nu, \mathrm{grad}h) f.$$

Proof. By integrating (37) and applying the Divergence Theorem,

$$-\int_M f\Delta h + \int_M g(\mathrm{grad} f, \mathrm{grad} h) = \int_M \mathrm{div}(f\mathrm{grad} h) = \int_{\partial M} g(\nu, f\mathrm{grad} h).$$

The proof is complete. □

Corollary 10.5.12. *If M is compact and $\partial M = \varnothing$, for any $f, h \in C^\infty(M)$*

$$\int_M f\Delta h = \int_M h\Delta f.$$

We already obtained this result in Exercise 7.5.9.

Corollary 10.5.13. *If M is compact and $\partial M = \varnothing$, for any $f \in C^\infty(M)$*

$$\int_M \Delta f = 0.$$

10.5.6. The Hessian. Let M be equipped with a connection ∇. The *Hessian* of a function $f \in C^\infty(M)$ is the tensor field $\nabla^2 f = \nabla(\nabla f) = \nabla(df)$ of type $(0,2)$. Its coordinates are

$$\nabla_i \nabla_j f = \frac{\partial^2 f}{\partial x^i \partial x^j} - \frac{\partial f}{\partial x^l}\Gamma_{ij}^l.$$

If ∇ is symmetric, the Hessian $\nabla^2 f$ also is. This applies of course to the case where ∇ is the Levi-Civita connection of a pseudo-Riemannian metric g.

By looking at the expressions in coordinates we see immediately that the Laplacian is minus the trace of the Hessian:

$$\Delta f = -g^{ij}\nabla_i \nabla_j f.$$

10.6. Exercises

Exercise 10.6.1 (The Clifton – Pohl torus). Consider the manifold $M = \mathbb{R}^2 \setminus \{0\}$ with the Lorentzian metric

$$g(x,y) = \frac{2}{x^2+y^2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Every map $f(x,y) = (\lambda x, \lambda y)$ is an isometry. In particular we may quotient M by the isometry $f(x,y) = (2x, 2y)$ and get a surface T diffeomorphic to a torus. The metric tensor pushes forward to a Lorentzian structure on T. Prove that the following curves

$$\gamma(t) = \left(\frac{1}{1-t}, 0\right), \qquad \eta(t) = (\tan(t), 1)$$

are both maximal geodesics defined on $(0,\infty)$ and $(-\frac{\pi}{2}, \frac{\pi}{2})$. Therefore T is compact but not geodesically complete.

Exercise 10.6.2. Consider the half-plane model H^2 of hyperbolic space. Let $v_0 = (0,1)$ be a tangent vector at $(0,1) \in H^2$. Let v_t be the parallel transport of v_0 along the curve $\gamma(t) = (t,1)$. Show that v_t makes an angle t with the vertical axis. Deduce that γ is not a geodesic.

Hint. Use the Christoffel symbols from Example 10.1.9. □

Exercise 10.6.3. Consider the connection ∇ on $\mathbb{R}^3$ having Christoffel symbols

$$\Gamma_{12}^3 = \Gamma_{23}^1 = \Gamma_{31}^2 = 1,$$
$$\Gamma_{21}^3 = \Gamma_{32}^1 = \Gamma_{13}^2 = -1,$$

and $\Gamma_{ij}^k = 0$ in all the other cases. Show that the connection is compatible with the Euclidean metric tensor but it is not symmetric. Determine its geodesics.

Exercise 10.6.4. Consider the ball model B^n of hyperbolic space. Pick $v \in S^{n-1}$. Show that the maximal geodesic through the origin with direction v is

$$\gamma_v(t) = \tanh(t/2)v = \frac{e^t - 1}{e^t + 1}v.$$

Deduce that the exponential map $\exp_p : T_p\mathbb{H}^n \to \mathbb{H}^n$ is a diffeomorphism $\forall p \in \mathbb{H}^n$.

Exercise 10.6.5. Identify $\mathbb{R}^2$ with $\mathbb{C}$ and write the half-plane model as $H^2 = \{z \in \mathbb{C} \mid \Im z > 0\}$. Show that the transformations

$$z \mapsto \frac{az + b}{cz + d}$$

with $a, b, c, d \in \mathbb{R}$ and $\det \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) > 0$ are isometries of H^2.

Exercise 10.6.6. Consider the hyperboloid model $I^n \subset \mathbb{R}^{n,1}$. Show that for any $p, q \in I^n$ we get

$$\cosh d(p, q) = -\langle p, q \rangle.$$

Exercise 10.6.7. Prove that two connections ∇, ∇' on M have the same geodesics $\iff$ the difference $D = \nabla - \nabla'$ is an antisymmetric tensor.[1] Deduce the following:

(1) $\nabla = \nabla' \iff$ they have the same geodesics and torsion.
(2) For any ∇ there is a unique ∇' with the same geodesics and without torsion.

Hint. Prove that D is antisymmetric $\iff D(X, X) = 0$ for any vector field X $\iff \nabla'_X X = \nabla_X X$ for any vector field X $\iff$ they share the same geodesics. $\square$

Exercise 10.6.8. Let (M, g) be a Riemannian manifold. For every p and $v \in V_p \subset T_pM$, and every curve η in V_p connecting 0 and v such that $d(\exp_p)_{\eta(t)}$ is non-singular for every t, show that

$$L(\exp_p \circ \eta) \geq \|v\|$$

with an equality if and only if η is a reparametrisation of the radial line $t \mapsto tv$.

Exercise 10.6.9. Let (M, g) be a Lorentzian manifold. Let $p \in M$ be a point and η a curve in V_p starting from 0. If $\exp_p \circ \eta$ is time-like, then η is entirely contained in one of the two timelike cones of T_pM.

Exercise 10.6.10. Let (M, g) be a Lorentzian manifold. For every p and $v \in V_p \subset T_pM$, and every curve η in V_p connecting 0 and v such that $d(\exp_p)_{\eta(t)}$ is non-singular for every t, show that

$$L(\exp_p \circ \eta) \leq \|v\|$$

with an equality if and only if η is a reparametrisation of the radial line $t \mapsto tv$.

[1]Recall from Proposition 9.2.9 that D is a tensor field of type $(1, 2)$ and hence we can interpret $D(p)$ as a bilinear map $T_pM \times T_pM \to T_pM$. By antisymmetry here we mean that $D(p)(v, w) = -D(p)(w, v)$ for any $p \in M$ and $v, w \in T_pM$. In coordinates: $D_{ij}^k = -D_{ji}^k$.

CHAPTER 11

Curvature

How can we distinguish two psuedo-Riemannian manifolds? Globally, they may have different topologies – and this is often detected by invariants like the fundamental group or De Rham cohomology – so we are now interested in constructing some *local* invariants. Can we measure locally how a pseudo-Riemannian manifold differs from being the more familiar $\mathbb{R}^{p,q}$ space?

The answer to all these questions is *curvature*, and the most complete answer is a formidable tensor field called the *Riemann curvature tensor*. This tensor field is pretty complicated and one sometimes wishes to examine some more reasonable tensor fields obtained from it via appropriate contractions: these are the *Ricci tensor* and finally the *scalar curvature*. A more geometric invariant which is in fact equivalent to the Riemann curvature tensor is the *sectional curvature*.

11.1. The Riemann curvature tensor

Let M be a smooth manifold, equipped with a connection ∇. We have already experienced with the torsion tensor T that one of the most efficient and natural ways to encode some information from ∇ is to build an appropriate tensor field. Tensor fields are lovely because they furnish some precise data at every single point $p \in M$. The torsion tensor is useless in the pseudo-Riemannian context, since $T = 0$ by assumption, so we must look for something else.

11.1.1. Definition. Recall that a tensor field of type $(1, n)$ on M is a multilinear map

$$\underbrace{T_pM \times \cdots \times T_pM}_{n} \longrightarrow T_pM$$

that depends smoothly on p.

Definition 11.1.1. The *Riemann curvature tensor* R is a tensor field on M of type $(1, 3)$ defined as follows. For every point $p \in M$ and vectors $u, v, w \in T_pM$ we set

$$R(p)(u, v, w) = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$$

where X, Y, Z are vector fields extending u, v, w on some neighbourhood of p.

Of course it is crucial here to prove that this (quite intimidating, we must admit) definition is well-posed.

Proposition 11.1.2. *The tangent vector $R(p)(u, v, w)$ is independent of the extensions X, Y, and Z.*

Proof. Armed with patience and optimism, we write everything in coordinates and get

$$\begin{aligned}\nabla_X\nabla_Y Z &= \nabla_X\left(Y^i\frac{\partial Z^k}{\partial x^i}e_k + Y^iZ^j\Gamma_{ij}^k e_k\right)\\ &= X^j\frac{\partial Y^i}{\partial x^j}\frac{\partial Z^k}{\partial x^i}e_k + X^jY^i\frac{\partial^2 Z^k}{\partial x^j\partial x^i}e_k + X^jY^i\frac{\partial Z^l}{\partial x^i}\Gamma_{jl}^k e_k\\ &\quad + X^m\frac{\partial Y^i}{\partial x^m}Z^j\Gamma_{ij}^k e_k + X^mY^i\frac{\partial Z^j}{\partial x^m}\Gamma_{ij}^k e_k + X^mY^iZ^j\frac{\partial \Gamma_{ij}^k}{\partial x^m}e_k\\ &\quad + X^mY^iZ^j\Gamma_{ij}^l\Gamma_{lm}^k e_k.\end{aligned}$$

There are 7 terms. If we calculate the difference $\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z$ the terms number 2, 3, and 5 cancel, and the terms 1 and 4 form the expression

$$[X,Y]^i\frac{\partial Z^k}{\partial x^i}e_k + [X,Y]^iZ^j\Gamma_{ij}^k e_k = \nabla_{[X,Y]}Z.$$

From this we deduce that $R(p)(u, v, w)$ consists only of the terms number 6 and 7 that depend (linearly) on u, v, and w and not on their extensions. The proof is complete. □

The tensor field R is therefore well-defined. To check that it is indeed smooth, we work on a chart and note that during the proof we have also found implicitly the coordinates of R in terms of the Christoffel symbols and their derivatives. After renaming indices we get

$$R_{ijk}{}^l = R(e_i, e_j, e_k)^l = \frac{\partial \Gamma_{jk}^l}{\partial x^i} - \frac{\partial \Gamma_{ik}^l}{\partial x^j} + \Gamma_{im}^l\Gamma_{jk}^m - \Gamma_{jm}^l\Gamma_{ik}^m. \tag{38}$$

In particular $R_{ijk}{}^l$ depends smoothly on the point. In particular we have

$$R(u, v, w)^l = R_{ijk}{}^l u^i v^j w^k.$$

The only example that we make for the moment is rather trivial.

Example 11.1.3. On the pseudo-Riemannian manifold $\mathbb{R}^{p,q}$ the Christoffel symbols vanish and therefore $R_{ijk}{}^l = 0$ everywhere.

Like torsion, parallel transport, and geodesics, the Riemann tensor is naturally associated to ∇. Therefore, as usual, if a diffeomorphism $\varphi\colon M \to N$ carries the connection ∇ on M to the connection $\varphi_*\nabla$ on N, it also sends the Riemann tensor R of the first to the Riemann tensor φ_*R of the second.

As every tensor field, the Riemann tensor gives a $C^\infty(M)$-multilinear map

$$R\colon \mathfrak{X}(M)\times\mathfrak{X}(M)\times\mathfrak{X}(M)\longrightarrow\mathfrak{X}(M)$$

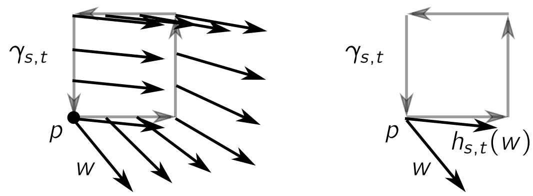

Figure 11.1. Given two commuting vector fields X and Y extending v and w, for every small $s, t > 0$ a quadrilateral loop $\gamma_{s,t}$ based in p is defined as the concatenation of four integral curves of $X, Y, -X$, and $-Y$ that last precisely the time s, t, s, t respectively. By Proposition 5.4.13, on a chart we may write X and Y as two coordinate vector fields, so γ_t is a rectangle of sides $s \times t$ as in the picture. The holonomy along $\gamma_{s,t}$ is the parallel transport along $\gamma_{s,t}$ (left). The Riemann tensor measures the deviation of the holonomy $h_{s,t}$ along $\gamma_{s,t}$ from the identity (right).

that can be written elegantly as

$$R(X, Y, Z) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

It is sometimes useful to consider another version of the Riemann tensor, where all the indices are in their lower position:

$$R_{ijkl} = R_{ijk}{}^m g_{lm}.$$

In this version the Riemann tensor is a tensor of type $(0, 4)$. Of course we can transform it back to the original $(1, 3)$ tensor using g^{lm}, so there is no loss of information in using one version instead of the other.

11.1.2. Holonomy along small quadrilaterals. At this stage, the Riemann tensor may look frustratingly complicated. Why do we need as much as four indices to encode curvature? We answer to this question by describing a simple and intuitive geometric interpretation.

The geometric interpretation is roughly the following. Look at Figure 11.1. If we parallel-transport a vector w along a small quadrilateral, we end up with a different vector. The Riemann tensor furnishes (at the second order) the rate of change of this vector. To diligently produce this output, the Riemann tensor needs three input vectors: two to describe the quadrilateral, plus w. It transforms three vectors into a vector, so it is a tensor field of type $(1, 3)$.

Here is a rigorous description. Let $u, v \in T_pM$ be two tangent vectors at some point $p \in M$. It is always possible (exercise: pick a chart) to extend them locally to two commuting vector fields X and Y. Pick a third vector $w \in T_pM$ and extend it to any vector field Z. Since $[X, Y] = 0$ we get

$$R(p)(u, v, w) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z.$$

For sufficiently small $t > 0$, let $\gamma_{s,t}$ be the closed loop based in p constructed as in Figure 11.1-(left) as the concatenation of four integral curves of X, Y, $-X$, and $-Y$, lasting precisely the time s, t, s, t respectively. Of course the loop $\gamma_{s,t}$ closes up because the two vector fields (and hence their flows) commute. We can now parallel-transport the vector w along the curve $\gamma_{s,t}$ as shown in the figure, to find at the end a new vector $h_{s,t}(w) \in T_pM$, called the *holonomy* of w along $\gamma_{s,t}$.

Theorem 11.1.4. *We have*

$$h_{s,t}(w) = w - R(p)(u,v,w)st + o(s^2+t^2).$$

Proof. On a chart $X = e_1$ and $Y = e_2$. The quadrilateral has vertices $A = (0,0)$, $B = (s,0)$, $C = (s,t)$, $D = (0,t)$. Let $w(s)$ be the vector w parallel-transported along A. Then $\frac{dw^i}{ds} + w^k\Gamma^i_{1k} = 0$ which gives

$$\begin{aligned}\frac{dw^i}{ds} &= -w^k\Gamma^i_{1k},\\ \frac{d^2w^i}{ds^2} &= -\frac{dw^k}{ds}\Gamma^i_{1k} - w^k\frac{\partial\Gamma^i_{1k}}{\partial x^1} = w^j\Gamma^k_{1j}\Gamma^i_{1k} - w^k\frac{\partial\Gamma^i_{1k}}{\partial x^1}\end{aligned}$$

and therefore the Taylor expansion for $w(s)$ is

$$w^i(s) = w^i - w^k\Gamma^i_{1k}s + \left(w^j\Gamma^k_{1j}\Gamma^i_{1k} - w^k\frac{\partial\Gamma^i_{1k}}{\partial x^1}\right)s^2 + o(s^2)$$

where the Christoffel symbols and their derivatives are calculated in $(0,0)$. We now let $w(s,t)$ be the vector $w(s)$ parallel-transported along B. Analogously,

$$w^i(s,t) = w^i(s) - w^k(s)\Gamma^i_{2k}t + \left(w^j\Gamma^k_{2j}\Gamma^i_{2k} - w^k(s)\frac{\partial\Gamma^i_{2k}}{\partial x^2}\right)t^2 + o(t^2)$$

where the Christoffel symbols and their derivatives are now calculated at $(s,0)$. By carefully combining the two formulas, together with

$$\Gamma^i_{2k}(s,0) = \Gamma^i_{2k} + \frac{\partial\Gamma^i_{2k}}{\partial x^1}s + o(s)$$

we get

$$\begin{aligned}w^i(s,t) = {}& w^i - w^k\Gamma^i_{1k}s + \left(w^j\Gamma^k_{1j}\Gamma^i_{1k} - w^k\frac{\partial\Gamma^i_{1k}}{\partial x^1}\right)s^2\\ & - (w^k - w^l\Gamma^k_{1l}s)\left(\Gamma^i_{2k} + \frac{\partial\Gamma^i_{2k}}{\partial x^1}s + o(s)\right)t\\ & + \left(w^j\Gamma^k_{2j}\Gamma^i_{2k} - w^k\frac{\partial\Gamma^i_{2k}}{\partial x^2}\right)t^2 + o(s^2+t^2).\end{aligned}$$

By reordering terms we finally find

$$\begin{aligned}w^i(s,t) = {}& w^i - w^k\Gamma^i_{1k}s - w^k\Gamma^i_{2k}t + \left(w^j\Gamma^k_{1j}\Gamma^i_{1k} - w^k\frac{\partial\Gamma^i_{1k}}{\partial x^1}\right)s^2\\ & + \left(w^l\Gamma^k_{1l}\Gamma^i_{2k} - w^k\frac{\partial\Gamma^i_{2k}}{\partial x^1}\right)st + \left(w^j\Gamma^k_{2j}\Gamma^i_{2k} - w^k\frac{\partial\Gamma^i_{2k}}{\partial x^2}\right)t^2 + o(s^2+t^2).\end{aligned}$$

All the Christoffel symbols and their derivatives are now calculated at $(0,0)$. If $w_*^i(s,t)$ is the vector transported from A to C passing through B we get an analogous formula, and their difference is

$$\begin{aligned} w^i(s,t) - w_*^i(s,t) &= \left(w^l \Gamma_{1l}^k \Gamma_{2k}^i - w^l \Gamma_{2l}^k \Gamma_{1k}^i - w^k \frac{\partial \Gamma_{2k}^i}{\partial x^1} + w^k \frac{\partial \Gamma_{1k}^i}{\partial x^2} \right) st + o(s^2+t^2) \\ &= {R_{21j}}^i w^j st + o(s^2+t^2). \end{aligned}$$

Since ${R_{21j}}^i = -{R_{12j}}^i$, this completes the proof. □

Note the analogies with Proposition 5.4.10. The endomorphism

$$R(p)(u,v,\cdot)\colon T_pM \longrightarrow T_pM$$

whose coordinates are ${R_{ijk}}^l u^i v^j$ measures the second-order deviation of the holonomy along a small quadrilateral with sides u and v.

11.1.3. The Riemann tensor in normal coordinates. Recall from Section 10.2 that the exponential map $\exp_p$ furnishes some nice normal coordinates around each point $p \in M$, such that $\Gamma_{ij}^k = 0$ at the point. In these coordinates the expression (38) simplifies and we get

$$(39) \qquad {R_{ijk}}^l = \frac{\partial \Gamma_{jk}^l}{\partial x^i} - \frac{\partial \Gamma_{ik}^l}{\partial x^j}.$$

Of course this equation is valid only at the point p. If (M,g) is a pseudo-Riemannian manifold, we can deduce a reasonable expression for R_{ijkl} directly in terms of the metric tensor:

Proposition 11.1.5. *At the point p, in normal coordinates we have*

$$(40) \qquad R_{ijkl} = \frac{1}{2}\left(\frac{\partial^2 g_{jl}}{\partial x^i \partial x^k} + \frac{\partial^2 g_{ik}}{\partial x^j \partial x^l} - \frac{\partial^2 g_{il}}{\partial x^j \partial x^k} - \frac{\partial^2 g_{jk}}{\partial x^i \partial x^l} \right).$$

Proof. In normal coordinates the first derivatives of g in p vanish. Then

$$\begin{aligned} R_{ijkl} = g_{lm}{R_{ijk}}^m &= g_{lm}\left(\frac{\partial \Gamma_{jk}^m}{\partial x^i} - \frac{\partial \Gamma_{ik}^m}{\partial x^j} \right) \\ &= \frac{1}{2} g_{lm} g^{hm} \left(\frac{\partial}{\partial x^i}\left(\frac{\partial g_{kh}}{\partial x^j} + \frac{\partial g_{hj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l} \right) - \frac{\partial}{\partial x^j}\left(\frac{\partial g_{kh}}{\partial x^i} + \frac{\partial g_{hi}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^l} \right) \right) \\ &= \frac{1}{2}\left(\frac{\partial^2 g_{jl}}{\partial x^i \partial x^k} + \frac{\partial^2 g_{ik}}{\partial x^j \partial x^l} - \frac{\partial^2 g_{il}}{\partial x^j \partial x^k} - \frac{\partial^2 g_{jk}}{\partial x^i \partial x^l} \right). \end{aligned}$$

The proof is complete. □

Note the absence of repeated indices: the element R_{ijkl} is just the sum of four second partial derivatives of the metric g. We could not have hoped for a simpler formula. Of course the use of normal coordinates is crucial here.

We have expressed the Riemann tensor in function of the metric and of the Christoffel symbols. Now we study the converse problem and try to express the metric tensor in terms of the Riemann tensor. Recall that in normal coordinates $g_{ij}(0) = \eta_{ij} = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}$ and $\frac{\partial g_{ij}}{\partial x^k}(0) = 0$. The first interesting terms in the Taylor expansion for g_{ij} are the second order derivatives, and these are precisely R_{ikjl} up to a constant:

Proposition 11.1.6. *In normal coordinates we have*

$$g_{ij}(x) = \eta_{ij} + \frac{1}{3} R_{ikjl}(0) x^k x^l + o\big(\|x\|^2\big).$$

Proof. We would like to express the second derivatives or g_{ij} or the first derivatives of Γ_{ij}^k in terms of the Riemann tensor. The equations (39) and (40) are very useful, but they only do the converse job and we are not able to invert them. We need an additional relation between the partial derivatives of the Γ_{ij}^k, furnished by Proposition 10.2.2 that says

$$\frac{\partial \Gamma_{jk}^i}{\partial x^l}(0) + \frac{\partial \Gamma_{kl}^i}{\partial x^j}(0) + \frac{\partial \Gamma_{lj}^i}{\partial x^k}(0) = 0.$$

Combining this with (39) we get

$$R_{ijk}{}^l(0) + R_{ikj}{}^l(0) = 3 \frac{\partial \Gamma_{jk}^l}{\partial x^i}(0).$$

Now we can express the first derivatives of the Christoffel symbols in terms of the Riemann tensor. We write the Taylor expansion

$$\begin{aligned}
g_{ij}(x) &= \eta_{ij} + \frac{1}{2} \frac{\partial^2 g_{ij}}{\partial x^l \partial x^k}(0) x^k x^l + o\big(\|x\|^2\big) \\
&= \eta_{ij} + \frac{1}{2} \frac{\partial}{\partial x^l} \left(\Gamma_{ki}^m g_{mj} + \Gamma_{kj}^m g_{mi} \right) \Big|_{x=0} x^k x^l + o\big(\|x\|^2\big) \\
&= \eta_{ij} + \frac{1}{2} \left(\frac{\partial \Gamma_{ki}^m}{\partial x^l}(0) \eta_{mj} + \frac{\partial \Gamma_{kj}^m}{\partial x^l}(0) \eta_{mi} \right) x^k x^l + o\big(\|x\|^2\big) \\
&= \eta_{ij} + \frac{1}{6} \big(R_{lkij}(0) + R_{likj}(0) + R_{lkji}(0) + R_{ljki}(0) \big) x^k x^l + o\big(\|x\|^2\big) \\
&= \eta_{ij} + \frac{1}{6} \big(R_{likj}(0) + R_{ljki}(0) \big) x^k x^l + o\big(\|x\|^2\big) \\
&= \eta_{ij} + \frac{1}{6} \big(R_{kilj}(0) + R_{ljki}(0) \big) x^k x^l + o\big(\|x\|^2\big) \\
&= \eta_{ij} + \frac{1}{3} R_{ikjl}(0) x^k x^l + o\big(\|x\|^2\big).
\end{aligned}$$

We have used that $g_{ij}(0) = \delta_{ij}$, $\frac{\partial g_{ij}}{\partial x^k}(0) = 0$, and the equalities $R_{lkij} + R_{lkji} = 0$ and $R_{kilj} = R_{ljki} = R_{ikjl}$, that are easy consequences of Proposition 11.1.5 and will be highlighted in the next section. □

In normal coordinates, the Riemann tensor measures the second-order deviation of g_{ij} from the constant metric η_{ij}.

11.1.4. Symmetries. Being a $(1,3)$-tensor field, we expect the Riemann tensor R to contain a tremendous amount of information on g, and this is what really happens. To help mastering this huge amount of data, we start by unraveling some symmetries.

Proposition 11.1.7. *The following symmetries hold in any coordinate chart:*

(1) $R_{ijkl} = -R_{jikl} = -R_{ijlk}$,
(2) $R_{ijkl} = R_{klij}$,
(3) $R_{ijk}{}^{l} + R_{jki}{}^{l} + R_{kij}{}^{l} = 0$.

Before entering in the proof, note that these symmetries may be stated more intrinsically as follows: for every $p \in M$ and $u, v, w, z \in T_pM$ we get

(1) $R(p)(u, v, w, z) = -R(p)(v, u, w, z) = -R(p)(u, v, z, w)$,
(2) $R(p)(u, v, w, z) = R(p)(w, z, u, v)$,
(3) $R(p)(u, v, w) + R(p)(v, w, u) + R(p)(w, u, v) = 0$.

In the first two we interpret R as a $(0,4)$ tensor field, while in the last we take the original $(1,3)$ tensor field. We will use R slightly ambiguously in this way.

From this intrinsic description we deduce immediately that if some of the above relations (1)-(3) is verified for some basis of T_pM, then it is automatically verified with respect to any basis.

Another intrinsic description consists in saying that some tensor obtained by symmetrising or antisymmetrising some (not all) indices vanishes. We can write (1) and (3) as follows:

$$R_{(ij)kl} = R_{ij(kl)} = 0, \qquad R_{(ijk)}{}^{l} = 0.$$

The symmetry (2) is harder to write in this way because it involves the symultaneous antisymmetrisation of non-adjacent indices.

Proof. Take normal coordinates at p. There R_{ijkl} has the convenient expression (40), which displays (1) and (2) immediately. Analogously for $R^{i}{}_{jkl}$ we use (39) to deduce (3) easily. The proof is complete. $\square$

The Riemann tensor has a priori n^4 independent components, but thanks to its symmetries these reduce to a smaller number. In normal coordinates we can lower the index of symmetry (3) and work fully with tensors of type $(0,4)$.

Proposition 11.1.8. *The* $(0,4)$ *tensors on* $\mathbb{R}^n$ *satisfying the symmetries*

$$R_{ijkl} = -R_{jikl} = -R_{ijlk}, \qquad R_{ijkl} = R_{klij}, \qquad R_{ijkl} + R_{iklj} + R_{iljk} = 0$$

form a vector subspace of $\mathcal{T}^4(\mathbb{R}^n)$ *of dimension*

$$\frac{1}{12}n^2(n^2-1).$$

Proof. The formula is

$$\binom{n}{2} + 3\binom{n}{3} + \frac{4!}{8}\cdot\frac{2}{3}\binom{n}{4} = \frac{1}{12}n^2(n^2-1).$$

The addenda count the number of independent components with 2, 3, and 4 distinct indices respectively. Those with 2 and 3 distinct components transform to R_{abab} and R_{abac} via the symmetries (1) and (2), so counting them is easy. Those with 4 components are $4!\binom{n}{4}$ in total; the symmetries (1)-(2) produce orbits with 8 elements, while (3) contributes by canceling one orbit out of three, hence with a $\frac{2}{3}$ factor. $\square$

Example 11.1.9. In dimension 2, the space has dimension 1 and is governed by R_{1212}. In dimension 3, the space has dimension 6 and is determined by

$$R_{1212},\ R_{1313},\ R_{2323},\ R_{1213},\ R_{2123},\ R_{3132}.$$

In dimension 4, the dimension is 20 and governed by the coordinates

$$R_{1212},\ R_{1313},\ R_{1414},\ R_{2323},\ R_{2424},\ R_{3434},\ R_{1213},\ R_{2123},\ R_{3132}, R_{1214},$$

$$R_{2124},\ R_{4142},\ R_{1314},\ R_{3134},\ R_{4143},\ R_{2324},\ R_{3234},\ R_{4243},\ R_{1234},\ R_{1432}.$$

11.1.5. The Bianchi identity. Let M be equipped with a connection ∇. If ∇ is the Levi-Civita connection of some metric tensor g, then $\nabla g = 0$. What about ∇R? The covariant derivative of R is typically not zero. We can interpret ∇R as a tensor field of type $(1,4)$, with coordinates $\nabla_a R_{ijk}{}^l$. This complicated symmetric tensor inherits all the symmetries of R, plus one more called the *Bianchi identity*:

Proposition 11.1.10 (Bianchi identity). *We have*

$$\nabla_a R_{ijk}{}^l + \nabla_i R_{jak}{}^l + \nabla_j R_{aik}{}^l = 0.$$

Proof. Take normal coordinates at $p = 0$. The Christoffel symbols vanish in 0 and hence by Proposition 9.2.11 the covariant derivatives of R in 0 coincide with the directional derivatives. Therefore

$$\nabla_a R_{ijk}{}^l = \frac{\partial R_{ijk}{}^l}{\partial x^a} = \frac{\partial}{\partial x^a}\left(\frac{\partial \Gamma^l_{jk}}{\partial x^i} - \frac{\partial \Gamma^l_{ik}}{\partial x^j} + \Gamma^l_{im}\Gamma^m_{jk} - \Gamma^l_{jm}\Gamma^m_{ik}\right) = \frac{\partial^2 \Gamma^l_{jk}}{\partial x^a \partial x^i} - \frac{\partial^2 \Gamma^l_{ik}}{\partial x^a \partial x^j}$$

where we have used (38) and the vanishing of the Christoffel symbols at $p = 0$. The conclusion is now a straightforward computation. $\square$

11.1.6. Flat implies Euclidean. We have already noticed that the Riemann tensor measures the local deviation of a metric tensor g from the Euclidean metric. We now show that it does so in a complete way: we prove that a metric tensor g is locally Euclidean if and only if the Riemann tensor vanishes. Let us first fix some definitions.

We say that a Riemannian manifold M is *Euclidean* if it is locally isometric to $\mathbb{R}^n$, that is every $p \in M$ has an open neighbourhood $U(p) \subset M$ which is isometric to some open subset of the Euclidean $\mathbb{R}^n$.

We say that M is *flat* if its Riemann tensor vanishes everywhere.

Theorem 11.1.11. *A Riemannian manifold M is Euclidean $\iff$ it is flat.*

Proof. It is easy to check that Euclidean implies flat: on an open set of $\mathbb{R}^n$ equipped with the Euclidean tensor we have $\Gamma_{ij}^k = 0$ and hence $R = 0$.

We now prove the converse. Let M be flat. Pick a point in M and represent a small neighbourhood of it via normal coordinates $B(0, r) \subset \mathbb{R}^n$. Pick a small cube $(-\varepsilon, \varepsilon)^n$ contained in $B(0, r)$.

We now extend the orthonormal basis $e_1, \ldots, e_n$ at 0 to a frame on the cube, as follows: we first parallel-transport the basis along the axis x_1, then along x_2, and so on until x_n. At the i-th step the frame is defined only on the slice $S_i = \{x_{i+1} = \ldots = x_n = 0\}$ of the cube, and at the end it is defined everywhere. It is smooth because parallel transport depends smoothly on the initial data. We have thus constructed a frame $X_1, \ldots, X_n$ that is an orthonormal basis at every point of the cube, such that $X_i(0) = e_i$. By construction we have

$$\nabla_{e_i} X_k = 0 \quad \text{on } S_i \quad \forall k.$$

We now prove that in fact

$$\nabla_{e_j} X_k = 0 \quad \text{on } S_i \quad \forall k, \forall j \leq i.$$

We show this by induction on i. The case $i = 1$ is done, so we suppose that it holds for i and prove it for $i + 1$. We already know that $\nabla_{e_{i+1}} X_k = 0$ on S_{i+1}. If $j \leq i$, by our induction hypothesis we have $\nabla_{e_j} X_k = 0$ on the hyperplane S_i. To conclude it suffices to check that $\nabla_{e_{i+1}}(\nabla_{e_j} X_k) = 0$ on S_{i+1}. The coordinate fields $e_1, \ldots, e_n$ commute, hence flatness $R = 0$ gives

$$\nabla_{e_{i+1}}(\nabla_{e_j} X_k) = \nabla_{e_j}(\nabla_{e_{i+1}} X_k) = \nabla_{e_j}(0) = 0.$$

The inductive proof is completed and when $i = n$ it shows that

$$\nabla_{e_j} X_k = 0 \quad \forall k, j$$

everywhere on the cube. Since ∇ is symmetric we have

$$[X_i, X_j] = \nabla_{X_i} X_j - \nabla_{X_j} X_i = 0 - 0 = 0.$$

By Proposition 5.4.13 there is a chart $\varphi \colon U \to V$ with $U \subset (-\varepsilon, \varepsilon)^n$ that straightens these vector fields, that is that transports X_i into e_i. The map φ is an isometry between U and V with its Euclidean metric, because it sends pointwise an orthonormal basis $X_1, \ldots, X_n$ to the orthonormal basis $e_1, \ldots, e_n$. The proof is complete. $\square$

11.1.7. Family of curves. We will soon discover a tight relation between the Riemann tensor and the spreading behaviour of families of geodesics. For the moment, we simply prove that R measures the non-commutativity of the covariant derivative also on vector fields on families of curves.

Let M be equipped with a connection ∇ and $f \colon (-\varepsilon, \varepsilon) \times I \to M$ be a family of curves. We defined in Section 10.2.11 the notions of vector field X along f, and its covariant derivatives $D_s X$ and $D_t X$. The coordinate vector fields are S and T. Many manipulations of vector fields in M extend trivially to this context: in particular, given vector fields X_1, X_2, X_3 along f, it makes sense to define a fourth one $R(X_1, X_2, X_3)$ using the Riemann tensor R as

$$R(X_1, X_2, X_3)(s, t) = R(X_1(s, t), X_2(s, t), X_3(s, t)).$$

The following fact should not be too surprising.

Proposition 11.1.12. *For every vector field X on f we have*

$$D_sD_tX - D_tD_sX = R(S, T, X).$$

Proof. If f is an embedding, then S, T, X may be considered as vector fields on the image of f, and the equality is just the definition of R together with the fact that S and T commute.

As for Lemma 10.2.28, for a more general f we work in coordinates. Now f has image in $\mathbb{R}^n$ and we write $x(s, t) = f(s, t)$. From (26) we get

$$\begin{aligned} D_tX &= \frac{\partial X}{\partial t} + \frac{\partial x^i}{\partial t} X^j \Gamma_{ij}^k e_k, \\ D_sD_tX &= \frac{\partial^2 X}{\partial s \partial t} + \frac{\partial}{\partial s}\left(\frac{\partial x^i}{\partial t} X^j \Gamma_{ij}^k e_k\right) + \frac{\partial x^i}{\partial s}(D_tX)^j \Gamma_{ij}^k e_k. \end{aligned}$$

We now use normal coordinates at the point $x(s, t)$. We gratefully obtain $\Gamma_{ij}^k = 0$ at the point and there the expression becomes

$$D_sD_tX = \frac{\partial^2 X}{\partial s \partial t} + \frac{\partial x^i}{\partial t} X^j \frac{\partial \Gamma_{ij}^k}{\partial s} e_k = \frac{\partial^2 X}{\partial s \partial t} + \frac{\partial x^i}{\partial t} X^j \frac{\partial x^l}{\partial s} \frac{\partial \Gamma_{ij}^k}{\partial x^l} e_k.$$

The equality follows from the expression (39) of R in normal coordinates. □

11.2. Sectional curvature

We have seen in Section 11.1.2 that the Riemann curvature tensor measures the second order displacement of vectors that are parallel-transported along small quadrilaterals. We now propose a related geometric interpretation where quadrilaterals are replaced by small surfaces, or more punctually by planes in T_pM. This geometric interpretation is called the *sectional curvature*.

11.2.1. Definition. Let M be a pseudo-Riemannian manifold and R be its Riemann curvature tensor field in the $(0, 4)$ version. Let $p \in M$ be a point and $\sigma \subset T_pM$ be a *non-degenerate tangent plane*, that is a two dimensional linear subspace where the restriction $g(p)|_\sigma$ is non-degenerate. We now assign to σ a number $K(\sigma)$ called the *sectional curvature* along σ, as follows.

Let $u, v \in \sigma$ be arbitrary generators. We define

$$K(\sigma) = \frac{R(p)(u, v, v, u)}{Q(u, v)}$$

where

$$Q(u, v) = \langle u, u\rangle\langle v, v\rangle - \langle u, v\rangle^2$$

is not zero since σ is non-degenerate. When $g(p)|_\sigma$ is positive definite, this is the square of the area of the parallelogram spanned by u and v.

Proposition 11.2.1. *The sectional curvature $K(\sigma)$ is well-defined.*

Proof. Thanks to the symmetries of R, the quantity $K(\sigma)$ does not change (exercise) if we substitute (u, v) with one of the following:

$$(v, u), \qquad (\lambda u, v), \qquad (u + \lambda v, v).$$

By composing such moves we can transform (u, v) into any other basis. $\square$

11.2.2. K **determines** R**.** The Riemann tensor of course determines the sectional curvatures by definition; we now see that also the converse holds:

Proposition 11.2.2. *The sectional curvatures $K(\sigma)$ along non-degenerate planes $\sigma \subset T_pM$ determine the Riemann tensor $R(p)$.*

Proof. The sectional curvatures determine $R(p)(u, v, v, u)$ for all pairs of vectors $u, v \in T_pM$ that generate a non-degenerate plane; since these are easily proved to be dense in the set of all pairs of vectors, the sectional curvature determines $R(p)(u, v, v, u)$ for any pair u, v. The vector $R(p)(u + w, v, v, u + w)$ is therefore determined, and it equals

$$R(p)(u, v, v, u) + 2R(p)(u, v, v, w) + R(p)(w, v, v, w).$$

Therefore the sectional curvatures also determine $R(p)(u, v, v, w)$ $\forall u, v, w$. Analogously, the vector $R(p)(u, v + z, v + z, w)$ is determined and it equals

$$R(p)(u, v, v, w) + R(p)(u, v, z, w) + R(p)(u, z, v, w) + R(p)(u, z, z, w)$$

so the sectional curvatures determine the value of

$$R(p)(u, v, z, w) + R(p)(u, z, v, w) = R(p)(u, v, z, w) - R(p)(z, u, v, w)$$

for all u, v, w, z. If we look at the three numbers

$$R(p)(u, v, z, w), \quad R(p)(v, z, u, w), \quad R(p)(z, u, v, w)$$

we see that their sum is zero and their differences are determined: hence the three numbers are also determined. $\square$

We are not losing any information if we consider sectional curvatures instead of the Riemann tensor.

11.2.3. Constant sectional curvature. A pseudo-Riemannian manifold (M, g) has *constant sectional curvature* K if $K(\sigma) = K$ for every tangent plane $\sigma \subset T_pM$ at every point $p \in M$. This seems a very restrictive hypothesis – and indeed it is – however, it turns out that there are plenty of constant sectional curvature manifolds around.

Proposition 11.2.3. *The manifold M has constant sectional curvature K $\iff$ the Riemann tensor may be written as*

$$R(u,v,w,z) = K\big(\langle u,z\rangle\langle v,w\rangle - \langle u,w\rangle\langle v,z\rangle\big).$$

Proof. If R is of this type, we easily get $K(\sigma) = K$. Conversely, if $K(\sigma) = K$ for all non-degenerate σ, then R must be of this type by Proposition 11.2.2 (one only has to check that the proposed R has all the symmetries listed in Proposition 11.1.7). □

In coordinates, we write the expression for R as

$$R_{ijkl} = K(g_{il}g_{jk} - g_{ik}g_{jl}).$$

Proposition 11.2.4. *If* Isom(M) *acts transitively on the frames of M, then it has constant sectional curvature.*

Proof. Since Isom(M) acts transitively on frames, it acts transitively on all the non-degenerate tangent planes $\sigma \subset T_p(M)$ at all points $p \in M$ having the same signature $(2,0)$, $(1,1)$, or $(0,2)$. Therefore non-degenerate planes σ with the same signature (a,b) have the same curvature $K_{(a,b)}$. We now prove that $K_{(2,0)} = K_{(1,1)} = K_{(0,2)}$ using the fact that R is smooth.

At a fixed p, the pairs (u,v) of vectors generating a non-degenerate plane with fixed signature (a,b) form an open subset $U_{(a,b)} \subset T_pM \times T_pM$. We get

$$R(u,v,u,v) = K_{(a,b)}\big(\langle u,u\rangle\langle v,v\rangle - \langle u,v\rangle^2\big)$$

for every $(u,v) \in U_{(a,b)}$. The open subset $U_{(2,0)} \cup U_{(1,1)} \cup U_{(0,2)}$ is dense in $T_pM \times T_pM$. Since R is smooth, we get $K_{(2,0)} = K_{(1,1)} = K_{(0,2)}$. □

As a corollary, the manifolds $\mathbb{R}^{p,q}$, S^n, and H^n have constant sectional curvature K. We already know that $K = 0$ in the first case, and we will soon discover that $K = +1$ and $K = -1$ for the sphere and the hyperbolic space.

11.3. The Ricci tensor

The Riemann tensor has four indices and contains a huge amount of information. In many contexts we may wish to reduce this data to a more manageable object: with tensor fields, this information reduction can be accomplished in a very natural way by contracting some pair of indices. There is essentially only one way to do this here, and it leads to a tensor field of type $(0,2)$ called the *Ricci tensor*.

11.3.1. Definition. The Riemann curvature tensor R is a tensor field of type $(1,3)$ and it is of course natural to study its contractions, that are tensor fields of type $(0,2)$. There are three possible contractions of ${R_{ijk}}^l$, namely:

$${R_{ijk}}^i, \qquad {R_{ijk}}^j, \qquad {R_{ijk}}^k.$$

Using the symmetries of R we see easily that the first two differ only by a sign and the third vanishes. Therefore there is essentially only one way to get a non-trivial tensor field by contraction, and this yields the *Ricci tensor*:

$$R_{ij} = R_{kij}{}^{k}.$$

This is a tensor field of type $(0,2)$. Since Ricci has the same initial as Riemann, we indicate it by Ric, but we denote its components simply by R_{ij}. The Ricci tensor also defines a $C^\infty(M)$-bilinear map

$$Ric\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \longrightarrow C^\infty(M).$$

Proposition 11.3.1. *The Ricci tensor is symmetric.*

Proof. We have

$$R_{ij} = R_{kij}{}^{k} = R_{kijh}g^{hk} = R_{hjik}g^{hk} = R_{hji}{}^{h} = R_{ji}.$$

The proof is complete. □

Like the metric tensor, the Ricci tensor is a symmetric tensor field of type $(0,2)$. Note however that the Ricci tensor need not be positive definite and may also be degenerate: for instance, on an open set $U \subset \mathbb{R}^n$ with the Euclidean metric, all the tensors that we introduce vanish, including Ricci.

11.3.2. In normal coordinates. Let (M,g) be a pseudo-Riemannian manifold. What geometric information is carried by the Ricci tensor? In normal coordinates, it measures the second order variation of the determinant of g, much as the Riemann tensor measures the second order variation of g itself. Set as usual $\eta = \left(\begin{smallmatrix} I_p & 0 \\ 0 & -I_q \end{smallmatrix}\right)$ where (p,q) is the signature of g.

Proposition 11.3.2. *In normal coordinates we have*

$$\det g_{ij}(x) = \det \eta \left(1 - \frac{1}{3}R_{ij}(0)x^i x^j\right) + o\big(\|x\|^2\big).$$

Proof. For any $n \times n$ matrix A we have

$$\det(\eta + A) = \det \eta \det(I + \eta^{-1}A) = \det \eta (1 + \operatorname{tr}(\eta^{-1}A)) + o\big(\|A\|\big).$$

Combining this with Proposition 11.1.6 we get

$$\begin{aligned} \det g_{ij}(x) &= \det \eta \left(1 + \frac{1}{3}R_{ikjl}(0)\eta^{ij}x^k x^l\right) + o\big(\|x\|^2\big) \\ &= \det \eta \left(1 - \frac{1}{3}R_{iklj}(0)\eta^{ij}x^k x^l\right) + o\big(\|x\|^2\big) \\ &= \det \eta \left(1 - \frac{1}{3}R_{kl}(0)x^k x^l\right) + o\big(\|x\|^2\big). \end{aligned}$$

The proof is complete. □

Let ω be the volume form determined by g. As a consequence, the Ricci tensor measures (in normal coordinates) the second order variation of ω.

Corollary 11.3.3. *In normal coordinates we have*

$$\omega = \left(1 - \frac{1}{6}R_{ij}(0)x^i x^j + o\big(\|x\|^2\big)\right) dx^1 \wedge \cdots \wedge dx^n.$$

Proof. This follows by applying the formula

$$\omega = \sqrt{|\det g_{ij}|}dx^1 \wedge \cdots \wedge dx^n$$

together with $\sqrt{1+t} = 1 + \frac{1}{2}t + o(|t|)$. □

Pick a point $p \in M$ and a non-zero tangent vector $v \in T_pM$. Corollary 11.3.3 implies that the volume of a small cone of geodesics exiting from p around v is smaller or bigger than the corresponding Euclidean cone, according to the sign of $Ric(v, v)$.

Remark 11.3.4. If (M, g) is a Riemannian manifold, by the spectral theorem at every $p \in M$ we can find a basis for T_pM that is simultaneously orthonormal for $g(p)$ and orthogonal for $Ric(p)$. Therefore we can choose normal coordinates at p where $g_{ij}(0) = \delta_{ij}$ and $R_{ij}(0)$ is a diagonal matrix.

11.3.3. Sum of sectional curvatures. Let (M, g) be a pseudo-Riemannian manifold. Pick $p \in M$ and an orthonormal basis $e_1, \ldots, e_n$ of T_pM.

Proposition 11.3.5. *We have*

$$Ric(e_i, e_i) = \langle e_i, e_i\rangle \sum_{j\neq i} K(\sigma(e_i, e_j))$$

where $\sigma(e_i, e_j)$ is the plane generated by e_i and e_j.

Proof. The left hand-side equals

$$\sum_{j=1}^n R_{jii}{}^j = \sum_{j=1}^n R_{jiij}\langle e_j, e_j\rangle = \langle e_i, e_i\rangle \sum_{j=1}^n \frac{R_{jiij}}{\langle e_i, e_i\rangle\langle e_j, e_j\rangle} = \langle e_i, e_i\rangle \sum_{j\neq i} K(\sigma(e_i, e_j)).$$

The proof is complete. □

The number $Ric(e_i, e_i)$ is $\langle e_i, e_i\rangle$ times the sum of the sectional curvatures of the $n-1$ coordinate planes containing the vector e_i. More generally if v is a unit vector we deduce that $Ric(v, v)$ is $\langle v, v\rangle$ times the sum of the sectional curvatures along the planes that contain v and the other vectors of any fixed orthonormal basis containing v.

Corollary 11.3.6. *If (M, g) has constant sectional curvature K, then*

$$Ric = (n-1)Kg.$$

11.4. The scalar curvature

If you think that a tensor of type $(0, 2)$ is yet too complicated an invariant, on a pseudo-Riemannian manifold you can still contract it and get an interesting number, called the *scalar curvature*.

11.4.1. Definition. Let (M, g) be a pseudo-Riemannian manifold. The *scalar curvature* at a point $p \in M$ is

$$R = g^{ij} R_{ij}.$$

Note that we need the metric g here: the scalar curvature is *not* defined for a general connection ∇. With respect to an orthonormal basis, the scalar curvature is simply the trace of the Ricci tensor. The scalar curvature is still indicated with the same letter R as the Riemann and Ricci curvature: the number and position of the indices are enough to distinguish from objects like R, R_{ij}, $R_{ijk}{}^{l}$, and R_{ijkl}.

Which kind of geometric information is conveyed by the scalar curvature? On a Riemannian manifold (M, g), it furnishes some data on the volumes of small geodesic balls. Let $p \in M$ be a point and $B(p, r)$ a geodesic ball of radius r centered at p (remember that this notion is well defined only for sufficiently small $r > 0$). We recall that the volume of a Euclidean ball $B(0, r) \subset \mathbb{R}^n$ is

$$\mathrm{Vol}\big(B(0, r)\big) = V_n(r) = \frac{\pi^{\frac{n}{2}}}{\Gamma\big(\frac{n}{2} + 1\big)} r^n$$

where Γ is Euler's gamma function.

Proposition 11.4.1. *We have*

$$\mathrm{Vol}\big(B(p, r)\big) = V_n(r) \cdot \left(1 - \frac{1}{6(n+2)} R(p) r^2 + o(r^3)\right). \tag{41}$$

Proof. Following Remark 11.3.4, we work in normal coordinates around $p = 0$ where the Ricci tensor $R_{ij}(0)$ is diagonal with entries $\lambda_1, \ldots, \lambda_n$. The scalar curvature is its trace $R(0) = \lambda_1 + \cdots + \lambda_n$. By Corollary 11.3.3 we get

$$\mathrm{Vol}\big(B(p, r)\big) = \int_{B(0,r)} \omega = \int_{B(0,r)} \left(1 - \frac{1}{6} R_{ij}(0) x^i x^j + o\big(\|x\|^2\big)\right) dx^1 \wedge \cdots \wedge dx^n.$$

We now compute

$$\begin{aligned}
\int_{B(0,r)} R_{ij}(0) x^i x^j dx^1 \wedge \cdots \wedge dx^n &= \int_{B(0,r)} (\lambda_1 (x^1)^2 + \cdots + \lambda_n (x^n)^2) dx^1 \wedge \cdots \wedge dx^n \\
&= \sum_{i=1}^n \lambda_i \left(\int_{B(0,r)} (x^i)^2 dx^1 \wedge \cdots \wedge dx^n \right) \\
&= \left(\sum_{i=1}^n \lambda_i \right) \frac{1}{n} \int_{B(0,r)} \rho^2 dx^1 \wedge \cdots \wedge dx^n
\end{aligned}$$

where $\rho^2 = (x^1)^2 + \cdots (x^n)^2$. Let $d\Omega$ be the volume form in the Euclidean S^{n-1}. The last expression equals

$$\begin{aligned}\frac{R(0)}{n}\int_{B(0,r)}\rho^2\cdot\rho^{n-1}d\rho\wedge d\Omega &= \frac{R(0)}{n}\left(\int_0^r \rho^{n+1}d\rho\right)\left(\int_{S^{n-1}} d\Omega\right)\\ &= \frac{R(0)}{n}\cdot\frac{r^{n+2}}{n+2}\mathrm{Vol}(S^{n-1}).\end{aligned}$$

With similar calculations, the volume of the Euclidean ball of radius r is

$$V_n(r) = \frac{r^n}{n}\mathrm{Vol}(S^{n-1})$$

so we find

$$\int_{B(0,r)} R_{ij}(0)x^i x^j dx^1\wedge\cdots\wedge dx^n = \frac{R(0)}{n+2}V_n(r)r^2.$$

Finally, we get

$$\mathrm{Vol}\big(B(p,r)\big) = V_n(r)\left(1-\frac{R(0)}{6(n+2)}r^2+o(r^3)\right).$$

The proof is complete. □

The scalar curvature measures the second order deviation of the ratio between volumes of small geodesic balls and Euclidean balls with the same small radius. Note that this is an intrinsic property of a point $p\in M$, that is coordinates independent. In particular, if $R(p)$ is negative (respectively, positive), geodesic balls of small radius r centered at p have strictly larger (respectively, smaller) volume than the Euclidean ones with the same radius r.

Example 11.4.2. On a surface, the equation (41) becomes

$$\mathrm{Vol}\big(B(p,r)\big) = \pi r^2\left(1-\frac{R(p)}{24}r^2+o(r^3)\right) = \pi r^2 - \frac{R(p)}{24}\pi r^4 + o(r^5).$$

On a 3-manifold, we get

$$\mathrm{Vol}\big(B(p,r)\big) = \frac{4}{3}\pi r^3\left(1-\frac{R(p)}{30}r^2+o(r^3)\right) = \frac{4}{3}\pi r^3 - \frac{2R(p)}{45}\pi r^5 + o(r^6).$$

11.4.2. The contracted Bianchi identity. By contracting the Bianchi identity twice, we get the following formula that relates the divergence of Ric with the covariant derivative of R. Here $R^a{}_j = g^{ak}R_{kj}$ as usual.

Corollary 11.4.3. *We have*

$$\nabla_a R^a{}_j = \frac{1}{2}\nabla_j R$$

Proof. The operation of raising or lowering some indices commutes with ∇ since $\nabla g = 0$. Therefore the Bianchi identity can be written as

$$\nabla_a R_{ij}{}^{kl} + \nabla_i R_{ja}{}^{kl} + \nabla_j R_{ai}{}^{kl} = 0.$$

By contracting twice we get

$$0 = \nabla_a R_{kl}{}^{kl} + \nabla_k R_{la}{}^{kl} + \nabla_l R_{ak}{}^{kl} = -\nabla_a R_{lk}{}^{kl} + \nabla_k R_{la}{}^{kl} + \nabla_l R_{ka}{}^{lk}$$
$$= -\nabla_a R + 2\nabla_k R_a{}^k$$

whence the conclusion. □

11.4.3. Effects of a metric rescaling. Let (M, g) be a as usual pseudo-Riemannian manifold. If we rescale the metric g by a factor $\lambda \neq 0$, we get a new metric tensor $g' = \lambda g$ with the same Levi-Civita connection ∇ as g, see Remark 9.3.10. Therefore we get the same geodesics, the same parallel transport, the same Riemann curvature tensor R, and the same Ricci tensor Ric. Much of the geometry of the manifold is unaltered.

Beware that the Ricci curvature tensor R that is unaffected is the original $(1, 3)$ version, that is purely defined using ∇. The $(0, 4)$ version is then obtained by lowering an index via g, and hence it is altered as $R' = \lambda R$.

Similarly we find that the sectional curvature, the scalar curvature, and the volume form change as follows:

$$K'(\sigma) = \frac{1}{\lambda} K(\sigma), \qquad R' = \frac{1}{\lambda} R, \qquad \omega' = \lambda^{\frac{n}{2}} \omega.$$

11.4.4. Low dimensions. In dimensions 2 and 3 the information carried by the curvature tensors reduce considerably and is more manageable.

Let S be a surface equipped with a Riemannian metric. At every point $p \in S$ the tangent plane T_pS has a sectional curvature $K(p)$, and the whole Riemann tensor is determined by this number by Proposition 11.2.2. Therefore all the information encoded by the Riemann tensor reduces to a more comfortable smooth function $K: S \to \mathbb{R}$, which is in fact equal to half the scalar curvature R: on an orthonormal basis e_1, e_2 for T_pS we get

$$R(p) = 2R_{1212}(p) = 2K(p).$$

Let M be a Riemannian 3-manifold. At a point $p \in M$ we fix an orthonormal basis e_1, e_2, e_3 for T_pM and note that the components R_{ijkl} of the Riemann tensor are determined by the Ricci tensor: at least two of the four indices i, j, k, l must coincide and therefore R_{ijkl} is either zero or equal to an entry of the Ricci tensor R_{ij}. Summing up, we have discovered the following.

Proposition 11.4.4. *The Riemann curvature tensor is determined by the scalar curvature in dimension $n = 2$ and by the Ricci tensor in dimension $n = 3$.*

11.5. Pseudo-Riemannian submanifolds

Let $N \subset M$ be a pseudo-Riemannian submanifold of a pseudo-Riemannian manifold (M, g). The submanifold N has two kinds of geometrical properties: the *intrinsic* ones depend only on the manifold $(N, g|_N)$ itself, while the *extrinsic* ones describe how N is embedded in M. The same N may have different intrinsic and extrinsic properties: for instance, it may be intrinsically flat and extrinsically curved, or viceversa.

11.5.1. Second fundamental form. We have seen in Section 9.3.6 that the tangential part of the connection ∇^M on N is the connection of N:

$$\nabla^N = \pi \circ \nabla^M.$$

We are now interested in the *normal* part of ∇^M. Let νN be the normal bundle of N in M. A section s of the bundle $\mathcal{T}_2 N \otimes \nu N$ is the datum, for every $p \in N$, of a bilinear map $s(p)\colon T_pN \times T_pN \to \nu N$.

The *second fundamental form* of $N \subset M$ is the section II of $\mathcal{T}_2 N \otimes \nu N$ defined as follows. For every $v, w \in T_pN$, we extend w to a vector field W on N near p and then put

$$\mathrm{II}(p)(v,w) = (\nabla^M_v W)^\perp$$

where $Z^\perp \in \nu_p N$ indicates the normal component of $Z \in T_pM$.

Proposition 11.5.1. *The tensor field* II *is well-defined and symmetric.*

Proof. Let V extend v along N. We get

$$(\nabla^M_v W)^\perp = (\nabla^M_w V)^\perp - [V,W]^\perp = (\nabla^M_w V)^\perp$$

since $[V,W]$ is tangent to N. This shows that $\mathrm{II}(p)(v,w)$ does not depend on the extension W and is symmetric. $\square$

Historically, the "second fundamental form" follows the "first fundamental form", that is just the metric $\mathrm{I}(v,w) = g(v,w)$. Both the first and second fundamental forms are symmetric operators on the tangent spaces, with value respectively in $\mathbb{R}$ and in the normal space.

If M has codimension 1 and is equipped with a normal unit field n, we may identify $\nu_N = N \times \mathbb{R}$ and hence II can be interpreted as a symmetric tensor field of type $(0,2)$ like I. In this case we get a useful formula:

Proposition 11.5.2. *If* $M \subset N$ *is a hypersurface with unit normal field* n*,*

$$\mathrm{II}(p)(v,w) = -\langle w, \nabla^N_v n\rangle n.$$

Proof. We have

$$\big\langle \mathrm{II}(p)(v,w), n\big\rangle = \big\langle \nabla^N_v W, n\big\rangle = -\langle w, \nabla^N_v n\rangle n$$

where we used Exercise **??** and the fact that $\langle W, n\rangle = 0$ everywhere. $\square$

With this formula, it suffices to calculate n in a neighbourhood of p to determine $\mathrm{II}(p)$.

11.5.2. The Gauss equation. By definition, for every pair of tangent fields X, Y in N we get

$$\nabla^M_X Y = \nabla^N_X Y + \mathrm{II}(X, Y).$$

This decomposition, sometimes called the *Gauss formula*, leads to a relation between curvatures and second fundamental form called the *Gauss equation*:

Proposition 11.5.3 (Gauss equation). *For every $u, v, w, z \in T_pN$ we have*

$$R^N(u, v, w, z) = R^M(u, v, w, z) + \big\langle \mathrm{II}(u, z), \mathrm{II}(v, w)\big\rangle - \big\langle \mathrm{II}(u, w), \mathrm{II}(v, z)\big\rangle.$$

Here R^N and R^M are the Riemann tensors of N and M.

Proof. Extend the vectors to vector fields U, V, W, Z locally on N. We may take U and V to commute, so that

$$R^N(u, v, w, z) = \big\langle \nabla^N_U \nabla^N_V W - \nabla^N_V \nabla^N_U W, Z\big\rangle.$$

We get:

$$\begin{aligned}
\big\langle \nabla^N_U \nabla^N_V W, Z\big\rangle &= \big\langle \nabla^N_U \nabla^M_V W, Z\big\rangle - \big\langle \nabla^N_U \big(\mathrm{II}(V, W)\big), Z\big\rangle \\
&= \big\langle \nabla^M_U \nabla^M_V W, Z\big\rangle - U\big\langle \big(\mathrm{II}(V, W)\big), Z\big\rangle + \big\langle \mathrm{II}(V, W), \nabla^N_U Z\big\rangle \\
&= \big\langle \nabla^M_U \nabla^M_V W, Z\big\rangle + \big\langle \mathrm{II}(V, W), \mathrm{II}(U, Z)\big\rangle.
\end{aligned}$$

We have used that Z is tangent to N, that $\mathrm{II}(V, W)$ is normal to N, and Exercise **??**. This leads directly to the Gauss equation. □

Corollary 11.5.4. *If $u, v \in T_pM$ generate a non-degenerate plane σ, then*

$$K^N(\sigma) = K^M(\sigma) + \frac{\big\langle \mathrm{II}(u, u), \mathrm{II}(v, v)\big\rangle - \big\langle \mathrm{II}(u, v), \mathrm{II}(u, v)\big\rangle}{\langle u, u\rangle\langle v, v\rangle - \langle u, v\rangle^2}.$$

11.5.3. Quadrics. We now describe a class of pseudo-Riemannian hypersurfaces with constant curvature that generalise the sphere S^n and the hyperbolic space $\mathbb{H}^n$ in the hyperboloid model. Recall that $\mathbb{R}^{p,q}$ is $\mathbb{R}^{p+q}$ equipped with the constant metric tensor

$$\langle x, y\rangle = -x^1y^1 - \cdots - x^qy^q + x^{q+1}y^{q+1} - \cdots + x^{p+q}y^{p+q}.$$

This is a pseudo-Riemannian manifold with signature (p, q) with constant sectional curvature $K = 0$. Set $Q(x) = \langle x, x\rangle$ and define the quadrics

$$S^{p,q} = \big\{x \in \mathbb{R}^{p+1,q} \mid Q(x) = 1\big\}, \qquad H^{p,q} = \big\{x \in \mathbb{R}^{p,q+1} \mid Q(x) = -1\big\}.$$

Proposition 11.5.5. *Each $S^{p,q} \subset \mathbb{R}^{p+1,q}$ and $H^{p,q} \subset \mathbb{R}^{p,q+1}$ is a pseudo-Riemannian submanifold with signature (p, q). We have*

$$T_xS^{p,q} = x^\perp \quad \forall x \in S^{p,q}, \qquad T_xH^{p,q} = x^\perp \quad \forall x \in H^{p,q}.$$

Proof. We show this with $S^{p,q}$, the proof for $H^{p,q}$ is the same. Note that

$$\langle x+y, x+y\rangle = \langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle$$

implies that $dQ_x(y) = 2\langle x,y\rangle$. We have $S^{p,q} = Q^{-1}(1)$ and 1 is a regular value for $Q\colon \mathbb{R}^{p+1,q} \to \mathbb{R}$, so $S^{p,q}$ is a smooth hypersurface and

$$T_xS^{p,q} = \ker dQ_x = \ker(y \mapsto 2\langle x,y\rangle) = x^\perp$$

for every $x \in S^{p,q}$. Since $\langle x,x\rangle = 1$, the restriction of $\langle , \rangle$ to the hyperplane $x^\perp$ has signature (p,q). $\square$

Of course we have $S^{n,0} = S^n$ and $H^{n,0}$ is isometric to two disjoint copies of the hyperboloid model of hyperbolic space. The topology of these spaces is easily determined.

Proposition 11.5.6. *The following diffeomorphisms hold:*

$$S^{p,q} \cong S^p \times \mathbb{R}^q, \qquad H^{p,q} \cong \mathbb{R}^p \times S^q.$$

Proof. We work with $S^{p,q}$, the case $H^{p,q}$ being similar. The map

$$\Psi\colon \mathbb{R}^q \times S^p \longrightarrow S^{p,q}, \qquad \Psi(x,y) = \left(x, \sqrt{1+\|x\|^2}y\right)$$

is a diffeomorphism, with inverse $(x,y) \mapsto \big(x, (1+\|x\|^2)^{-1/2}y\big)$. $\square$

Remark 11.5.7. The linear isomorphism $\mathbb{R}^{p+1,q} \to \mathbb{R}^{q,p+1}$,

$$\iota(x^1, \ldots, x^q, x^{q+1}, \ldots, x^{p+q+1}) = (x^{q+1}, \ldots, x^{p+q+1}, x^1, \ldots, x^q)$$

sends the metric tensor of $\mathbb{R}^{p+1,q}$ to minus the metric tensor of $\mathbb{R}^{q,p+1}$. This restricts to a diffeomorphism $\iota\colon S^{p,q} \to H^{q,p}$ that sends the metric tensor of the first to minus the metric tensor of the second. Therefore $H^{p,q}$ is isometric to $S^{p,q}$ with the sign of $\langle , \rangle$ inverted, that is rescaled it by a factor $\lambda = -1$. As discussed in Section 11.4.3, the geometries of $S^{p,q}$ and $H^{q,p}$ are pretty much the same, although their sectional and scalar curvatures differ by a sign, and the signatures (p,q) and (q,p) are inverted.

11.5.4. Isometries and curvature of the quadrics. Set $n = p+q$ as usual. We define $\mathrm{O}(p,q) \subset \mathrm{GL}(n,\mathbb{R})$ to be the subgroup of all the linear isometries of $\mathbb{R}^n$ that preserve the scalar product $\langle x,y\rangle = {}^{\mathrm{t}}xI_{p,q}y$ with

$$I_{p,q} = \begin{pmatrix} -I_q & 0 \\ 0 & I_p \end{pmatrix}.$$

That is,

$$\mathrm{O}(p,q) = \left\{A \in \mathrm{GL}(n,\mathbb{R}) \ \middle|\ {}^{\mathrm{t}}AI_{p,q}A = I_{p,q}\right\}$$

Proposition 11.5.8. *The isometry groups of $S^{p,q}$ and $H^{p,q}$ are*

$$\mathrm{Isom}(S^{p,q}) = \mathrm{O}(p+1,q), \qquad \mathrm{Isom}(H^{p,q}) = \mathrm{O}(p,q+1).$$

The group acts freely and transitively on the frames of $S^{p,q}$ and $H^{p,q}$.

Proof. For every $A \in \mathrm{O}(p, q)$, the map $x \mapsto Ax$ preserves $\langle , \rangle$ and hence it restricts to an isometry of both $S^{p,q}$ and $H^{p,q}$. It is a simple linear algebra exercise to show that $\mathrm{O}(p, q)$ acts transitively on the frames of both $S^{p,q}$ and $H^{p,q}$, and hence it coincides to its isometry group. $\square$

The isometry group of $M = S^{p,q}$ or $H^{p,q}$ acts transitively on frames, so it acts transitively on the set of all tangent planes $\sigma \subset T_xM$ at all points $x \in M$. In particular the sectional curvature $K(\sigma) = K$ is constant for every σ. We calculate this number K.

Proposition 11.5.9. *The manifolds $S^{p,q}$ and $H^{p,q}$ have constant sectional curvature $K = 1$ and $K = -1$ respectively.*

Proof. It suffices to work out $S^{p,q}$ since $H^{p,q}$ is the (-1)-rescaling of $S^{q,p}$. The outer normal vector field on $S^{p,q}$ is simply $n(x) = x$, because $T_xS^{p,q} = x^\perp$ for all $x \in S^{p,q}$. For every vector field X in $\mathbb{R}^{p,q}$ we have

$$\nabla_X n = X^i \frac{\partial x^i}{\partial x^i} e_i = X.$$

By applying Proposition 11.5.2 to $S^{p,q} \subset \mathbb{R}^{p,q}$ we get

$$\mathrm{II}(v, w) = -\big\langle w, \nabla_v n \big\rangle n = -\langle v, w \rangle n$$

and therefore by Corollary 11.5.4

$$K = 0 + \frac{\langle u, u\rangle\langle v, v\rangle - \langle u, v\rangle^2}{\langle u, u\rangle\langle v, v\rangle - \langle u, v\rangle^2} = 1.$$

The proof is complete. $\square$

11.5.5. Geodesics in the quadrics. By generalising further the arguments exposed in Section 10.1.2, we can easily prove the following.

Proposition 11.5.10. *Pick $p \in S^{p,q}$ and $v \in p^\perp = T_pS^{p,q}$. We have:*

$$\begin{array}{rcll}
\gamma_v(t) & = & \cos(\|v\|t) \cdot p + \sin(\|v\|t) \cdot \frac{v}{\|v\|} & \text{if } v \text{ is spacelike,} \\
\gamma_v(t) & = & \cosh(\|v\|t) \cdot p + \sinh(\|v\|t) \cdot \frac{v}{\|v\|} & \text{if } v \text{ is timelike,} \\
\gamma_v(t) & = & p + tv & \text{if } v \text{ is lightlike.}
\end{array}$$

The same holds for $H^{p,q}$, with the words "spacelike" and "timelike" inverted.

Proof. In all cases $\gamma_v(t) \in S^{p,q}$ and $\gamma_v''(t)$ is a multiple of $\gamma_v(t)$, hence it is orthogonal to $S^{p,q}$ for all t. $\square$

Corollary 11.5.11. *The quadrics $S^{p,q}$ and $H^{p,q}$ are geodesically complete.*

11.5.6. Totally geodesic submanifolds. Geodesics are the straightest curves in a pseudo-Riemann manifold, and we now introduce the "straightest possible k-submanifolds" in all dimensions $k \geq 2$.

Definition 11.5.12. Let (M, g) be a pseudo-Riemannian manifold. A semi-Riemannian submanifold $N \subset M$ is *totally geodesic* if the second fundamental form vanishes.

A totally geodesic submanifold $N \subset M$ has no extrinsic curvature, but it may pretty well have an intrinsic curvature: in fact, by the Gauss equation (see Proposition 11.5.3) the Riemann tensor R^N of N is just the restriction of that R^M of M, so N has the same intrinsic curvature of M.

We may define a semi-Riemannian manifold using geodesics only.

Proposition 11.5.13. *Let $N \subset M$ be a pseudo-Riemannian submanifold. The following are equivalent:*

(1) *N is totally geodesic.*
(2) *Parallel transport along curves in N of vectors tangent to N is the same with respect to ∇^N and ∇^M.*
(3) *Every geodesic of N is also a geodesic of M.*
(4) *For every $v \in TN$, the geodesic γ_v of M lies initially in N.*

Proof. (1)⇒(2). Let γ be a curve in N. A parallel vector field in N is parallel also in M since $\mathrm{II} = 0$. Therefore parallel transports are the same.

(2)⇒(3). A curve γ is a geodesic $\Leftrightarrow$ γ' is parallel.

(3)⇒(4). The geodesic γ_v^N of N is also a geodesic of M, so by uniqueness of geodesics in M with starting vector v it coincides initially with γ_v.

(4)⇒(1). For every tangent $v \in T_pN$ at any $p \in N$, we have $\nabla_v^M(\gamma_v') = 0$ and hence $\mathrm{II}(p)(v, v) = 0$. Therefore $\mathrm{II} = 0$. □

Example 11.5.14. Every affine subspace of $\mathbb{R}^{p,q}$ whose tangent space is non-degenerate is a totally geodesic submanifold.

Proposition 11.5.15. *The intersection of a non-degenerate vector subspace $W \subset \mathbb{R}^{p,q}$ of signature (p', q') with $S^{p,q}$ or $H^{p,q}$ is a totally geodesic submanifold X isometric to $S^{p',q'}$ or $H^{p',q'}$.*

The submanifold X is actually empty if X is isometric to $S^{0,q'}$ or $H^{p',0}$.

Proof. We work with $S^{p,q}$, the case of $H^{p,q}$ being analogous. The intersection $X = W \pitchfork S^{p,q}$ is transverse since for every $x \in X$ we have $T_xS^{p,q} = x^\perp$ and $x \in T_xW$, hence $T_xS^{p,q} + T_xW = \mathbb{R}^{p,q}$. Therefore X is a submanifold.

If we pick an orthonormal basis of W and complete to one of $\mathbb{R}^{p,q}$, we identify isometrically X with $S^{p',q'}$. Proposition 11.5.10 shows that the geodesic in $S^{p,q}$ starting from $p \in X$ with velocity $v \in T_pX = W \cap T_pS^{p,q}$ is contained in the plane $U \subset W$ generated by p and v and is hence contained in X. By Proposition 11.5.13-(4) the submanifold X is totally geodesic. □

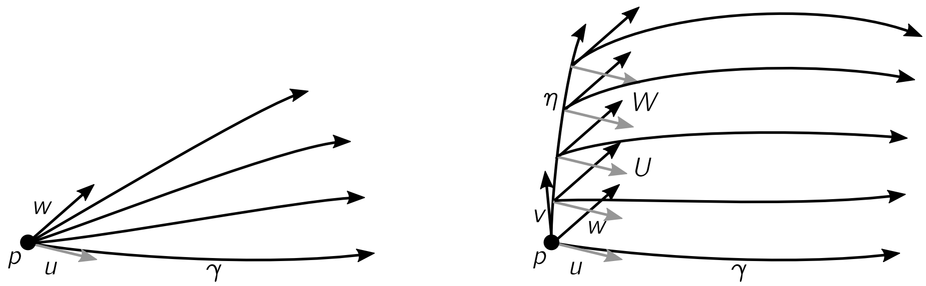

Figure 11.2. How to construct a family of geodesics $\gamma_s(t)$ starting from two vectors u, w (left) and more generally from three vectors u, v, and w (right) based at the same point p. In both cases we write $\gamma = \gamma_0$ and draw few geodesics with $s \in [0, \varepsilon)$.

11.5.7. Warped products. We introduce a class of semi-Riemannian manifolds that appear in various contexts.

Definition 11.5.16. Let B and F be two semi-Riemannian manifolds, and $f : B \to (0, +\infty)$ a smooth function. The *warped product* $M = B \times_f F$ is the semi-Riemannian manifold $B \times F$ equipped with the metric tensor

$$g(p, q) = \begin{pmatrix} g^B(p) & 0 \\ 0 & f^2(p)g^F(q) \end{pmatrix}.$$

If $f \equiv 1$, this is the usual product of pseudo-Riemannian manifolds. In general, we should think at M as fibering over the base B with a fiber F that is shrinked by a factor $f(p)$ above each point $p \in B$.

Example 11.5.17. A surface of revolution in $\mathbb{R}^3$ is a warped product. The Euclidean $\mathbb{R}^n \setminus 0$ is a warped product $(0, +\infty) \times_f S^{n-1}$ with $f(r) = r$.

11.6. Jacobi fields

We now study families of geodesics that depend on one parameter. We prove that these tend to spread when the curvature is negative, and to concentrate when the curvature is positive. The main tool is a kind of vector fields on geodesics called *Jacobi fields*, that measures the first-order variation of families of geodesics.

11.6.1. Families of geodesics. Let M be a manifold equipped with a connection ∇. A *family of geodesics* is a family of curves $f : (-\varepsilon, \varepsilon) \times I \to M$ where $\gamma_s(t) = f(s, t)$ is a geodesic $\forall s$. Recall that f is smooth by assumption.

Example 11.6.1. Fix $p \in M$ and two vectors $u, w \in T_pM$. Then

$$\gamma_s(t) = \exp_p(t(u + sw))$$

is a family of geodesics for $s \in (-\varepsilon, \varepsilon)$. We found a family of this type in the proof of the Gauss Lemma, see Figure 10.8. These geodesics are exiting from p in the direction $u + sw$. See also Figure 11.2-(left).

Example 11.6.2. The previous example can be generalised by allowing the starting point $\gamma_s(0)$ to move along another geodesic η as in Figure 11.2-(right).

Here are the details. Pick three vectors $u, v, w \in T_pM$. Build the geodesic $\eta(s) = \exp_p(v)$ for $s \in (-\varepsilon, \varepsilon)$. Parallel-transport the vectors u and w to two vector fields $U(s)$ and $W(s)$ along η. Define

$$\gamma_s(t) = \exp_{\eta(s)}(t(U(s) + sW(s))).$$

When $v = 0$ this reduces to the previous example.

11.6.2. Jacobi fields. Let γ be a geodesic. A *Jacobi field* is a vector field on γ that describes the first-order variation of a family of geodesics around γ.

More precisely, let M be a manifold equipped with a symmetric connection ∇, and f describe a family of geodesics γ_s with $\gamma = \gamma_0$. The vector field

$$J(t) = df_{(0,t)}\left(\frac{\partial}{\partial s}\right)$$

on γ is the *Jacobi field* of f. It is a vector field on the geodesic γ. The following proposition is crucial because it connects the Riemann tensor R with the first-order variation of families of geodesics, encoded by J.

Proposition 11.6.3. *Every Jacobi field J satisfies the* Jacobi equation

$$D_tD_tJ + R(J, \gamma', \gamma') = 0. \tag{42}$$

Proof. Consider the coordinate fields S, T of the family of geodesics f. By Lemma 10.2.28 we have $D_tS = D_sT$. Since each γ_s is a geodesic, we also get $D_tT = 0$. By combining these with Proposition 11.1.12 we find

$$R(S, T, T) = D_sD_tT - D_tD_sT = -D_tD_tS.$$

At the points $s = 0$ this gives the desired equality. □

11.6.3. Solutions of the Jacobi equation. We now study the solutions of the Jacobi equation. We may write the equation conveniently as follows: pick an arbitrary orthonormal basis $e_1, \ldots, e_n$ at $T_{\gamma(t)}M$ for some $t \in I$ and parallel-transport it all along γ. Now every vector field X on γ may be written as $X = X^ie_i$ and we simply get

$$D_tX = \dot{X}^ie_i, \qquad D_tD_tX = \ddot{X}^ie_i.$$

The Jacobi equation (42) now can be written as

$$\ddot{J}^i + J^j\dot{\gamma}^k\dot{\gamma}^lR_{jkl}{}^i = 0. \tag{43}$$

This is a system of second-order linear differential equations. Therefore for any time $t_0 \in I$ and any pair of tangent vectors $v_1, v_2 \in T_pM$ at the point $p = \gamma(t_0)$ there is a unique solution J of (42) with initial values

$$J(t_0) = v_1, \qquad (D_tJ)(t_0) = v_2.$$

The solutions of the Jacobi equation are parametrised by their initial values $(v_1, v_2) \in T_pM \times T_pM$ at p and hence form a vector space of dimension $2n$.

Proposition 11.6.4. *The Jacobi field J of the family described in Example 11.6.2 and Figure 11.2-(right) has the initial values*

$$J(0) = v, \qquad (D_tJ)(0) = w.$$

Proof. Let S and T be the coordinate vector fields of the family. We have

$$\begin{aligned} J(0) &= S(0,0) = v, \\ (D_tJ)(0) &= (D_tS)(0,0) = (D_sT)(0,0) = w. \end{aligned}$$

The proof is complete. □

We already know that every Jacobi field is a solution of the Jacobi equation, and we now prove the converse (under some mild but necessary hypothesis). Let γ be a geodesic in M.

Proposition 11.6.5. *Suppose that either the geodesic γ is defined on a compact interval I, or the manifold M is geodesically complete.*

Every solution to the Jacobi equation is a Jacobi field on γ.

Proof. Let J be a solution of the Jacobi equation. Pick $t_0 \in I$ and set

$$p = \gamma(t_0), \qquad u = \gamma'(t_0), \qquad v = J(t_0), \qquad w = (D_tJ)(t_0).$$

We use the vectors u, v, and w to construct a family of geodesics γ_s as in Example 11.6.2 (see Figure 11.2). This family exists for all $s \in \mathbb{R}$ if M is geodesically complete, and for all $s \in (-\varepsilon, \varepsilon)$ in any case if I is compact. By Proposition 11.6.4 the Jacobi field of γ_s has the same initial values of J and hence it coincides with J. □

We are mostly interested in the case where the interval $I = [t_0, t_1]$ is compact, so γ is a geodesic connecting $p = \gamma(t_0)$ to $q = \gamma(t_1)$. In this setting the Jacobi fields J on γ are precisely the solutions of the Jacobi equations, and they form naturally a $2n$-dimensional vector space. Each Jacobi field J is determined by its initial values $J(t_0), (D_tJ)(t_0) \in T_pM$ at t_0, which can be arbitrary. More generally, it is determined by any values $J(t), (D_tJ)(t) \in T_{\gamma(t)}M$ at any fixed time $t \in I$.

11.6.4. Tangential and normal Jacobi fields. If (M, g) is a pseudo-RIemannian manifold, we may decompose every Jacobi field into its tangential and normal components, and both these components are again Jacobi fields. We explain this procedure here.

Let $\gamma\colon I \to M$ be a geodesic, defined on some compact interval I. A vector field X on γ is *tangential* (*normal*) if $X(t)$ is tangent to (orthogonal to) $\gamma'(t)$ for all t. Tangential Jacobi fields are easily classified.

Proposition 11.6.6. *Every tangential Jacobi field is of the form*

$$J(t) = a\gamma'(t) + bt\gamma'(t)$$

for some $a, b \in \mathbb{R}$. A Jacobi field J is tangential $\Longleftrightarrow$ both $J(t_0)$ and $(D_tJ)(t_0)$ are tangent for some (and hence all) time $t_0 \in I$.

Proof. The given J is the Jacobi field of the family of geodesics

$$\gamma_s(t) = \gamma((1 + bs)t + as)$$

obtained simply by reparametrising $\gamma = \gamma_0$ linearly. Its initial values at t_0 are

$$J(t_0) = (a + bt_0)\gamma'(t_0), \qquad (D_tJ)(t_0) = b\gamma'(t_0)$$

and by varying $a, b \in \mathbb{R}$ we get all possible pairs of tangent vectors at $\gamma(t_0)$. From this the conclusions easily follow. $\square$

Recall that a geodesic γ in M is lightlike if $\langle\gamma'(t), \gamma'(t)\rangle = 0$ for some (equivalenty, all) time t. If γ is not lightlike, every vector field X on γ decomposes uniquely as $X = X^\perp + X^{||}$ into a normal and a tangential component.

Proposition 11.6.7. *If J is a Jacobi field, both components $J^\perp$, $J^{||}$ also are.*

Proof. We have by definition

$$J^{||} = \frac{\langle J, \gamma'\rangle}{\langle\gamma', \gamma'\rangle}\gamma'.$$

By applying Exercise **??** twice and recalling that $D_t\gamma' = 0$, we get

$$D_tD_tJ^{||} = \frac{\langle D_tD_tJ, \gamma'\rangle}{\langle\gamma', \gamma'\rangle}\gamma'.$$

We deduce easily that $J^{||}$ satisfies the Jacobi equation (which reduces in fact to $D_tD_tJ^{||} = 0$), since J does. By linearity then also $J^\perp$ does. $\square$

Proposition 11.6.8. *The following are equivalent for a Jacobi field J:*

1. *J is normal,*
2. *$J(t)$ and $(D_tJ)(t)$ are both orthogonal to $\gamma'(t)$, for some t,*
3. *$J(t_1)$ and $J(t_2)$ are both orthogonal to γ' for two distinct $t_1 \neq t_2$.*

Proof. For any Jacobi field J, set $g(t) = \langle J(t), \gamma'(t)\rangle$ and prove that $g''(t) = 0$. Therefore $g(t) = at + b$ and this easily implies the assertion. $\square$

The tangent and normal Jacobi fields form two subspaces of dimension 2 and $2n-2$. If γ is not lightlike, these subspaces are transverse. We are mostly interested in normal fields.

11.6.5. Jacobi fields and exponential map. We now look more closely at the families of geodesics that have a common starting point p. These families are in fact all restrictions of the exponential map $\exp_p\colon V_p \to M$, defined on some maximal star-shaped open subset $V_p \subset T_pM$. We now study the tight relations between the exponential map and Jacobi fields.

Let M be a manifold equipped with a symmetric connection ∇. Pick a point $p \in M$ and a vector $v \in V_p \subset T_pM$. This determines a geodesic $\gamma_v\colon I_v \to M$ that can be written as usual as $\gamma_v(t) = \exp_p(tv)$.

For every vector $w \in T_pM$, we may define the family of geodesics

$$\gamma_s(t) = \exp_p(t(v + sw))$$

with $\gamma_0 = \gamma_v$, see Figure 11.2-(left). The Jacobi field J of this family is

$$J(t) = (d\exp_p)_{tv}(tw). \tag{44}$$

This equality is important because it connects Jacobi fields with the differential of the exponential map $\exp_p$ at an arbitrary point $tv \in V_p$. Any information on the Jacobi fields may be used to understand the map $\exp_p$ on its whole domain – not only at the origin as we did until now. In particular in the next pages we will find some conditions that will certify that $\exp_p$ is (or is not) an immersion at any given point in its domain (recall that $\exp_p$ is guaranteed to be an immersion only at the origin).

The Jacobi field J in (44) has initial data $J(0) = 0$ and $(D_tJ)(0) = w$.

11.6.6. Conjugate points. Let M be a manifold equipped with a symmetric connection ∇. Let γ a geodesic connecting two points p and q.

Definition 11.6.9. The points p and q are *conjugate* along γ if there is a non-zero Jacobi field J on γ that vanishes at both endpoints p and q.

Suppose that $\gamma\colon [0, a] \to M$, with $p = \gamma(0)$, $v = \gamma'(0)$ and $q = \gamma(a)$. So

$$\gamma(t) = \exp_p(tv).$$

Proposition 11.6.10. *The points p and q are conjugate along $\gamma \iff av$ is a singular point for* $\exp_p$.

Proof. Any Jacobi field J along γ vanishing at p has initial data $J(0) = 0$ and $(D_tJ)(0) = w$ for some $w \in T_pM$, and by uniqueness it is of the form (44). The formula shows that $J(a) = 0 \Leftrightarrow aw \in \ker(d\exp_p)_{tv}$. □

Remark 11.6.11. The proof also shows that the dimension of all the Jacobi fields J that vanish at both endpoints equals $\dim\ker(d\exp_p)_{av}$. This number is called the *multiplicity* of the conjugate point q. The multiplicity is at most $n-1$, since the space of all Jacobi fields that vanish at p has dimension n and contains γ' that does not vanish at q (unless γ is constant, but in this case we see easily that p and q are not conjugate). If M is a pseudo-Riemannian manifold, a Jacobi field that vanishes at both endpoints must be normal by Proposition 11.6.8-(3).

Example 11.6.12. Pick S^n and a point $p \in S^n$. As shown in Example 10.1.14, when $\|v\| = \pi$ we get $\exp(v) = -p$. Therefore $-p$ is a conjugate point along any geodesic γ exiting from p with maximal multiplicity $n-1$.

Warning 11.6.13. The existence of a Jacobi field on γ that vanishes at the endpoints p, q does not guarantee that there are other nearby geodesics connecting p and q. It only furnishes a family γ_s of geodesics starting from p, whose endpoints are at a distance $o(s)$ from q.

Proposition 11.6.14. *If p and q are not conjugate along γ, for every $v \in T_pM$, $w \in T_qM$ there is a unique Jacobi field J on γ with $J(0) = v, J(a) = w$.*

Proof. Let $\mathcal{J}$ be the $2n$-dimensional vector space of all Jacobi fields on γ. Consider the map $\mathcal{J} \to T_pM \times T_pM$, $J \mapsto J(0), J(a)$. The map is injective since p and q are not conjugate, hence it is surjective. $\square$

11.6.7. The Cartan – Hadamard teorem. In the previous pages we have connected the Riemann tensor R and the Jacobi fields J via the Jacobi equation (42), and then the Jacobi fields with the exponential $\exp_p$ via (44). We now weld these two connections and study the effects of R on $\exp_p$ and on the geometry and topology of M.

We say that a Riemannian manifold M has *negative*, *non-positive*, ecc. *sectional curvature* if the sectional curvature $K(\sigma)$ is negative, non-positive, ecc. for every plane $\sigma \subset T_pM$ at every point $p \in M$.

Theorem 11.6.15 (Cartan – Hadamard). *Let (M, g) be a complete connected Riemannian manifold with non-positive sectional curvature. For every $p \in M$ the exponential map $\exp_p \colon T_pM \to M$ is a smooth covering.*

Proof. Let $\gamma(t)$ be a geodesic emanating from p. Let J be a Jacobi field on γ with $J(0) = 0$. Set $f(t) = \langle J(t), J(t)\rangle$. By Exercise **??** we have

$$\begin{aligned} f'(t) &= 2\langle D_tJ, J\rangle \\ f''(t) &= 2\langle D_tD_tJ, J\rangle + 2\|D_tJ\|^2 \\ &= -2R(J, \gamma', \gamma', J) + 2\|D_tJ\|^2 \geq 0. \end{aligned}$$

Therefore there are no conjugate points and hence $\exp_p$ is a local diffeomorphism. We now equip T_pM with the pull-back metric $g^* = \exp_p^*(g)$, so that $\exp_p$ is promoted to a local isometry between Riemannian manifolds.

A crucial observation here is that the geodesics through 0 for g and g^* are exactly the same. In particular, they exist for al $\mathbb{R}$, and hence (T_pM, g^*) is complete by Proposition 10.3.8. From Exercise 10.4.7 we deduce that $\exp_p$ is a Riemannian covering. $\square$

Corollary 11.6.16. *Let M be a complete connected Riemannian manifold with non-positive sectional curvature. The universal cover of M is diffeomorphic to $\mathbb{R}^n$.*

Proof. The universal cover $\tilde{M}$ inherits from M the structure of a Riemannian manifold with non-positive sectional curvature, that is also complete by Exercise 10.4.8. The exponential map $T_p\tilde{M} \to \tilde{M}$ at any $p \in \tilde{M}$ is a smooth covering, and since $\tilde{M}$ is simply connected it is a diffeomorphism. □

Corollary 11.6.17. *Let M be a simply connected complete Riemannian manifold with non-positive sectional curvature. The exponential map at any point is a diffeomorphism. In particular any two points $p, q \in M$ are joined by a unique geodesic (which is necessarily minimising).*

Corollary 11.6.18. *A compact manifold M with finite fundamental group does not admit any Riemannian metric of non-positive sectional curvature.*

Proof. Since $\pi_1(M)$ is finite, its universal covering $\widetilde{M}$ is also compact, but $\widetilde{M} \cong \mathbb{R}^n$, a contradiction. □

11.6.8. Tidal forces. The Jacobi equation has an immediate physical interpretation. If we construct a geometric model of spacetime where free falling bodies travel along geodesics, then a set of nearby objects falling onto a planet forms a family of geodesics; the Jacobi field J may be interpreted as the mutual distance of two falling bodies, J' as their relative velocity, and J'' as their relative acceleration; in this setting, the Jacobi equation (42) reproduces Newton's second law of motion, saying that the acceleration J'' is equal to the value of some gravitational field that is determined by R. The gravitational field in our spacetime model should somehow be encoded in R.

We are led to the following definition. Let (M, g) be a pseudo-Riemannian manifold. For any non-trivial $v \in T_pM$, we define the *tidal force operator* as

$$F_v \colon v^\perp \longrightarrow v^\perp, \qquad F_v(u) = R(u, v, v).$$

Exercise 11.6.19. The operator F_v is self-adjoint, with trace $-Ric(v, v)$.

Let us say that a vector v is *cospacelike* if $v^\perp$ is positive-definite. This holds in the most interesting cases: when M is Riemannian or when M is Lorentzian and v is timelike. If v is cospacelike, we may use the spectral theorem and find an orthonormal basis $v_1, \ldots, v_{n-1}$ of $v^\perp$ where the tidal force operator is diagonal. If $\|v\| = 1$ we write $\sigma = \langle v, v\rangle = \pm 1$ and easily deduce that the eigenvectors are just the sectional curvatures multiplied by $-\sigma$:

$$F_v(v_i) = -\sigma K(\operatorname{Span}(v, v_i))v_i.$$

11.7. Calculus of variations

On a Riemannian manifold, a geodesic is a curve that locally minimises the distance. Sometimes a geodesic γ connecting p and q minimises the distance also globally, and in this case γ is a minimum of the length functional on all curves going from p to q. In general, a geodesic may not be a minimum for the length functional, not even locally, but it is still a *critical point* with a well-defined *index*, like with Morse functions in Section 6.4.8.

The index is also defined in the slightly more general (and physically interesting) setting of cospacelike geodesics in pseudo-Riemannian manifolds.

11.7.1. First variation of the length and energy. Let M be a pseudo-Riemannian manifold. Let $\alpha\colon [a,b] \to M$ be a timelike or spacelike curve connecting two points p and q. We have $\|\alpha'(t)\| > 0$ for all t by hypothesis. The sign $\sigma = \operatorname{sgn}\langle \alpha', \alpha' \rangle = \pm 1$ depends on whether the curve is spacelike or timelike.

We now consider families of (timelike or spacelike) curves γ_s extending γ_0 with the same endpoints $\gamma_s(a) = p, \gamma_s(b) = q$. We define the *length function*

$$L\colon (-\varepsilon, \varepsilon) \longrightarrow \mathbb{R}, \qquad L(s) = L(\gamma_s).$$

We are interested in particular at its behaviour at 0. Recall the coordinate vector fields S and T of the family of curves. Here we are interested in the *variational vector field* $V(t) = S(0,t)$ on γ.

Lemma 11.7.1. *We have*

$$L'(0) = \sigma \int_a^b \big\langle \gamma'/\|\gamma'\|, D_t V \big\rangle dt.$$

Proof. We get

$$\begin{aligned} L'(s) &= \frac{d}{ds}\int_a^b \|\gamma_s'(t)\| dt = \int_a^b \frac{d}{ds}\|T\| dt = \int_a^b \frac{d}{ds}\sqrt{\sigma\langle T, T\rangle} dt \\ &= \int_a^b \frac{2\sigma\langle D_s T, T\rangle}{2\sqrt{\sigma\langle T,T\rangle}} dt = \sigma\int_a^b \frac{\langle D_t S, T\rangle}{\|T\|} dt. \end{aligned}$$

We have used Lemma 10.2.28. For $s = 0$ we get the result. □

The derivative $L'(0)$ is called the *first variation* of L along the family γ_s. The following proposition shows that the geodesics are precisely the curves with constant speed c whose first variation vanishes on all families.

Proposition 11.7.2. *If $c = \|\gamma'\| > 0$ is constant, we get*

$$L'(0) = -\frac{\sigma}{c}\int_a^b \langle D_t\gamma', V\rangle dt.$$

In particular $L'(0) = 0$ for all families of curves $\gamma_s \iff \gamma$ is a geodesic.

Proof. Since $\frac{d}{dt}\langle\gamma', V\rangle = \langle D_t\gamma', V\rangle + \langle\gamma', D_tV\rangle$, we integrate by parts

$$L'(0) = \frac{\sigma}{c}\int_a^b \big\langle\gamma', D_tV\big\rangle dt = \frac{\sigma}{c}\langle\gamma', V\rangle\Big|_a^b - \frac{\sigma}{c}\int_a^b \langle D_t\gamma', V\rangle dt.$$

The curves γ_s have the same endpoints, hence $V(a) = V(b) = 0$ and the formula is proved. If γ is a geodesic, then $D_t\gamma' = 0$ and therefore $L'(0)$. If γ is not a geodesic, then $D_t\gamma'(t) \neq 0$ for some t, and using a bump function with support near t one constructs easily a vector field V along γ with $V(a) = V(b) = 0$ that gives $L'(0) \neq 0$ when substituted in the formula.

Any V with $V(a) = V(b) = 0$ is the variational vector field of some γ_s, for instance we may take $\gamma_s(t) = \exp_{\gamma(t)}(sV(u))$. $\square$

Corollary 11.7.3. *Let γ be any (spacelike or timelike) curve connecting p and q. We have $L'(0)$ for all families $\gamma_s \Longleftrightarrow \gamma$ is a reparametrised geodesic.*

Proof. If we reparametrise γ so that it has constant speed $c > 0$, and reparametrise correspondingly all the families γ_s, the length functional L is unaffected. So we may suppose that γ has constant speed and then apply Proposition 11.7.2 to conclude. $\square$

It may be useful to think of all the curves connecting p to q as points in some infinite-dimensional manifold, of variations γ_s as paths in this manifold, and of the variation vector fields V at a curve γ as vectors in the (infinite-dimensional) tangent space of γ. With this interpretation, the length L is a function on this manifold, and the critical points for L are precisely the reparametrised geodesics.

If you are annoyed by the fact that any reparametrisation of a geodesic is a critical point, you probably prefer to substitute the length with the *energy*, that is defined as follows. The *energy* of γ is

$$E(\gamma) = \int_a^b \langle\gamma'(t), \gamma'(t)\rangle.$$

Note that the energy may be negative if the scalar product is not definite positive. This quantity has a less appealing geometric meaning than the length, but the removal of the square root has some pleasant consequences. The computation of $E'(0)$ is simpler than that of $L'(0)$ and is left as an exercise.

Exercise 11.7.4. For every family γ_s of curves connecting p and q we get

$$E'(0) = 2\int_a^b \langle\gamma', D_tV\rangle dt = -2\int_a^b \langle D_t\gamma', V\rangle dt.$$

In particular $E'(0) = 0$ for all families of curves $\gamma_s \Longleftrightarrow \gamma$ is a geodesic.

Note that this is valid for any curve γ, not only spacelike or timelike. We needed the hypothesis $\|\gamma_s'\| > 0$ above because the norm is not smooth at zero, but this is not required anymore for $\langle\gamma_s', \gamma_s'\rangle$.

11.7.2. Second variation of the length. Let $\gamma\colon [a,b] \to M$ be a spacelike or timelike geodesic on M connecting two points p, q. We have $L'(0) = 0$ for any family γ_s, so it is natural to look at the second derivative $L''(0)$, called the *second variation* of the length.

We write $\sigma = \mathrm{sgn}\langle\gamma', \gamma'\rangle = \pm 1$ and $c = \|\gamma'\|$. Let S, T be the coordinate fields of the family γ_s and $V(t) = S(0,t)$ be the variational vector field. Recall that every vector field X on γ decomposes as $X^\perp + X^{||}$ into a normal and tangential component. Since γ is a geodesic, we deduce easily that $(D_tX)^\perp = D_t(X^\perp)$ and we can hence write it as $D_tX^\perp$.

Proposition 11.7.5. *We have*

$$L''(0) = \frac{\sigma}{c}\int_a^b \big(\langle D_tV^\perp, D_tV^\perp\rangle - R(V, \gamma', \gamma', V)\big)dt.$$

Proof. We have

$$\begin{aligned} L''(s) &= \int_a^b \frac{d^2}{ds^2}\|\gamma_s'(t)\|dt = \sigma\int_a^b \frac{d}{ds}\frac{\langle D_tS, T\rangle}{\|T\|}dt \\ &= \sigma\int_a^b \frac{\big(\langle D_sD_tS, T\rangle + \langle D_tS, D_sT\rangle\big)\|T\| - \sigma\langle D_tS, T\rangle^2/\|T\|}{\|T\|^2} \\ &= \sigma\int_a^b \frac{R(S,T,S,T) + \langle D_tD_sS, T\rangle + \langle D_tS, D_tS\rangle - \sigma\langle D_tS, T\rangle^2/\|T\|^2}{\|T\|} \end{aligned}$$

In the second equality we used the calculation done during the proof of Lemma 11.7.1. Define the vector field $A(t) = D_sS(0,t)$ on γ'. With $s = 0$ we get

$$\begin{aligned} L''(0) &= \frac{\sigma}{c}\int_a^b -R(V,\gamma',\gamma',V) + \frac{d}{dt}\langle A, \gamma'\rangle + \langle D_tV, D_tV\rangle - \frac{\langle D_tV, \gamma'\rangle^2}{\langle\gamma',\gamma'\rangle} \\ &= \frac{\sigma}{c}\langle A, \gamma'\rangle\Big|_a^b + \frac{\sigma}{c}\int_a^b -R(V,\gamma',\gamma',V) + \langle D_tV^\perp, D_tV^\perp\rangle. \end{aligned}$$

Since $A(a) = A(b) = 0$, the proof is complete. $\square$

Remark 11.7.6. By the symmetries of R, we may write this equality as

$$L''(0) = \frac{\sigma}{c}\int_a^b \big(\langle D_tV^\perp, D_tV^\perp\rangle - R(V^\perp, \gamma', \gamma', V^\perp)\big)dt.$$

Here only the normal component $V^\perp$ is present.

Corollary 11.7.7. *If a Riemannian M has non-positive curvature, we get $L''(0) \geq 0$ for every family γ_s, with a strict inequality if V is not tangential.*

On a Riemannian manifold M with non-positive curvature, every geodesic is shorter than its small perturbations. When M is complete, this may also be deduced by applying Corollary 11.6.17 to the universal cover $\tilde{M}$.

11.7.3. The index form. We now define a bilinear symmetric form on the vector space of all fields V tangent to the geodesic γ. This bilinear form I, called the *index form*, should be interpreted as the hessian of the functional L. Given two fields V, W tangent to γ and vanishing at the endpoints, we set

$$I(V,W) = \frac{\sigma}{c}\int_a^b \big(\langle D_t V^\perp, D_t W^\perp\rangle - R(V,\gamma',\gamma',W)\big)\,dt.$$

The constants σ, c are defined as above. We note immediately that

$$I(V,V) = L''(0)$$

where $L(s) = L(\gamma_s)$ and γ_s is any family of curves with variational vector field V. The first thing to note is that a tangential vector field V lies in the radical of I, that is $I(V,W) = 0$ for any W. In fact we have

$$I(V,V) = I(V^\perp, V^\perp).$$

If we integrate by parts, we find

$$I(V,W) = -\frac{\sigma}{c}\int_a^b \big\langle D_t D_t V^\perp - R(V^\perp,\gamma',\gamma'), W^\perp\big\rangle\,dt. \tag{45}$$

From this we deduce that the Jacobi fields (that vanish at the endpoints) are also in the radical of I. Recall that such a Jacobi field exists (by definition) only when the endpoints are conjugate along γ.

Exercise 11.7.8. The radical consists of those fields V whose normal component $V^\perp$ is a Jacobi field.

11.7.4. Cospacelike geodesics. In some lucky cases the intersection form I is positive or negative semidefinite, that is $I(V,V) \geq 0$ or $I(V,V) \leq 0$ for all V. This holds for instance if the geodesic γ is (correspondingly) the shortest or the longest path joining p and q. These cases may occur only in two distinct settings, both very important from a mathematical and physical perspective.

Proposition 11.7.9. *Suppose that I is positive or negative semidefinite. Then, after possibly substituting the metric tensor g with $-g$, one of the following cases holds:*

(1) *M is Riemannian and I is positive semidefinite.*
(2) *M is Lorentzian, γ is timelike, and I is negative semidefinite.*

Proof. Suppose that there is a unit vector $v \in \gamma'(t)^\perp \subset T_{\gamma(t)}M$ such that the sign $\sigma' = \langle v, v\rangle$ is opposite to the sign σ of $\langle \gamma'(t), \gamma'(t)\rangle$, that is $\sigma\sigma' = -1$. Parallel transport v to a field V on γ, pick $n \in \mathbb{N}$ and define the field

$$X(t) = \frac{1}{n}\sin\frac{2\pi n(t-a)}{b-a}V(t).$$

This field vanishes at the endpoints. We find

$$\begin{aligned} I(X,X) &= \frac{\sigma}{c}\int_a^b \sigma'\left(\frac{2\pi}{b-a}\cos\frac{2\pi n(t-a)}{b-a}\right)^2 - \frac{1}{n^2}\sin^2\frac{2\pi n(t-a)}{b-a}R(V,\gamma',\gamma',V) \\ &= \frac{1}{c}\int_a^b -\left(\frac{2\pi}{b-a}\cos\frac{2\pi n(t-a)}{b-a}\right)^2 - \frac{\sigma}{n^2}\sin^2\frac{2\pi n(t-a)}{b-a}R(V,\gamma',\gamma',V). \end{aligned}$$

If n is sufficiently big we get $I(X,X) < 0$, so I cannot be positive semidefinite. Analogously we find that if v and γ' have the same sign then I cannot be negative semidefinite. □

We denote both cases with a single word by saying that a geodesic γ is *cospacelike* if $\gamma'(t)^\perp$ is a positive definite hyperplane in $T_{\gamma(t)}$ for some (and hence all) t. A geodesic is cospacelike precisely if it is of one of the two types just mentioned: either M is Riemannian, or M is Lorentzian and γ is timelike.

11.7.5. Conjugate points. Let $\gamma\colon [a,b] \to M$ be a cospacelike geodesic on a pseudo-Riemannian manifold M, connecting two points p and q. We have defined the index form I on the space of all variation vector fields on γ, and shown that its radical consists of those fields V whose normal component is a Jacobi field. We write $I^\perp$ to denote the restriction of I to the V that are orthogonal to γ. We now relate the definiteness of $I^\perp$ with the existence of conjugate points on γ.

Theorem 11.7.10. *The following holds.*

(1) *If p has no conjugate points along γ, the index form $I^\perp$ is definite.*
(2) *If q is the only conjugate point of p along γ, then $I^\perp$ is semidefinite and not definite.*
(3) *If p has a conjugate point $\gamma(t_0)$ with $a < t_0 < b$, then $I^\perp$ is indefinite.*

By Proposition 11.7.9, if (1) or (2) holds the index form $I^\perp$ is positive or negative (semi-)definite depending on whether γ is spacelike or timelike.

Proof. (1). Pick Jacobi fields $J_1, \ldots, J_{n-1}$ on γ such that $J_1(0) = \ldots = J_{n-1}(0) = 0$ and $(D_tJ_1)(0), \ldots, (D_tJ_{n-1})(0)$ form a basis of $\gamma'(a)^\perp$. Jacobi fields are orthogonal to γ, and since there are no conjugate points the vectors $J_1(t), \ldots, J_{n-1}(t)$ form a basis of $\gamma'(t)^\perp$ for all $t \in (a,b]$. We remark that

$$\langle J_i, D_tJ_j\rangle = \langle D_tJ_i, J_j\rangle \tag{46}$$

To prove this, we see easily that the derivative along t of the difference of the two members is zero (using the Jacobi equation and the symmetries of R), so this difference is constant, and is actually zero at $t = a$.

Every orthogonal variation field V may be written as $V = V^iJ_i$. With some effort, we will prove below that

$$\langle D_tV, D_tV\rangle - R(V,\gamma',\gamma',V) = \langle \dot V^iJ_i, \dot V^iJ_i\rangle + \frac{d}{dt}\langle V, V^iD_tJ_i\rangle. \tag{47}$$

This will allow to conclude that

$$\begin{aligned}
I(V,V) &= \frac{\sigma}{c}\int_a^b \langle D_tV, D_tV\rangle - R(V,\gamma',\gamma',V) \\
&= \frac{\sigma}{c}\int_a^b \langle \dot V^i J_i, \dot V^i J_i\rangle + \frac{\sigma}{c}\langle V, V^i D_t J_i\rangle\Big|_a^b = \frac{\sigma}{c}\int_a^b \langle \dot V^i J_i, \dot V^i J_i\rangle
\end{aligned}$$

is positive or negative definite according to the sign σ of γ'. We turn to (47):

$$\begin{aligned}
\frac{d}{dt}\langle V, V^i D_t J_i\rangle &= \frac{d}{dt}\langle V^j J_j, V^i D_t J_i\rangle = \langle \dot V^j J_j, V^i D_t J_i\rangle + \langle V^j D_t J_j, V^i D_t J_i\rangle \\
&\quad + \langle V^j J_j, \dot V^i D_t J_i\rangle + \langle V^j J_i, V^i D_t D_t J_i\rangle \\
&= 2\langle \dot V^j J_j, V^i D_t J_i\rangle + \langle V^j D_t J_j, V^i D_t J_i\rangle - R(V,\gamma',\gamma',V) \\
&= \langle D_tV, D_tV\rangle - \langle \dot V^j J_j, \dot V^i J_i\rangle - R(V,\gamma',\gamma',V).
\end{aligned}$$

In the second equality we used (46) and the Jacobi equation.

(2) The same argument above shows that I is semidefinite; the presence of a Jacobi field shows that the radical is non trivial and hence I is not definite.

(3) Let J be a non-trivial Jacobi field with $J(a) = J(t_0) = 0$. Let V be the field on γ that equals J on $[a, t_0]$ and is zero on $[t_0, b]$. This field is only C^0 and not smooth at t_0, so we smoothen it to a normal field by modifying it a little in the interval $[t_0 - \varepsilon, t_0 + \varepsilon]$. For every field W, using (45) we find

$$\begin{aligned}
I(V,W) &= -\frac{\sigma}{c}\int_{t_0-\varepsilon}^{t_0+\varepsilon} \big\langle D_tD_tV - R(V,\gamma',\gamma'), W\big\rangle dt \\
&= -\frac{\sigma}{c}\big\langle (D_tV)(t_0+\varepsilon) - (D_tV)(t_0-\varepsilon), W(t_0)\big\rangle + O(\varepsilon) \\
&= \frac{\sigma}{c}\big\langle (D_tJ)(t_0), W(t_0)\big\rangle + O(\varepsilon).
\end{aligned}$$

In particular $I(V,V) = O(\varepsilon)$. Note that $(D_tJ)(t_0) \neq 0$ because J is non trivial. Pick a field W such that $W(t_0) = (D_tJ)(t_0)$. For every $\delta \in \mathbb{R}$ we have

$$I(V+\delta W, V+\delta W) = 2\delta\frac{\sigma}{c}\big\langle (D_tJ)(t_0), (D_tJ)(t_0)\big\rangle + \delta^2 I(W,W) + O(\varepsilon).$$

Since γ is cospacelike we have $\langle (D_tJ)(t_0), (D_tJ)(t_0)\rangle > 0$. If $\varepsilon > 0$ and $|\delta|$ are sufficiently small, we get both negative and positive numbers, according to the sign of $\delta\sigma$. Therefore I is indefinite. □

Remark 11.7.11. A more geometric proof (or at least intuition) towards Theorem 11.7.10 may be taken in the Riemannian case. We may obtain point (1) as a consequence of Exercise 10.6.8. Point (3) can be accepted intuitively by saying that if we substitute the first segment $[0, t_0]$ of γ with a nearby geodesic of the same length, we get a new curve with the same length as γ but having an angle at t_0 that can be then smoothened, to produce a strictly shorter nearby curve. This argument is however not rigorous: nearby geodesics do not have the same endpoint (see Warning 11.6.13) nor the same length.

11.7.6. Bonnet – Myers Theorem. On a Riemannian manifold M, the Cartan – Hadamard Theorem says that when the curvature is non-positive there are no conjugate points, and hence the exponential map at any point is a covering. As a topological consequence, the universal cover of M is $\mathbb{R}^n$.

In the opposite direction, we now show that positive sectional curvature on a Riemannian M forces the existence of conjugate points. We then deduce as a topological consequence that the universal cover of the manifold is compact.

Actually, we do not really need all the sectional curvatures to be positive, only their averages as measured by the Ricci tensor. Recall that $Ric(v, v)$ may be interpreted as the average of the sectional curvatures of the planes containing v times the sign of $\langle v, v\rangle$, and is also minus the trace of the tidal force operator F_v.

Let M be a pseudo-Riemannian manifold.

Proposition 11.7.12. *Let γ be a unit speed and cospacelike geodesic in M connecting p and q. If*

$$Ric(\gamma', \gamma') \geq (n-1)C^2, \qquad L(\gamma) \geq \pi/C$$

for some $C > 0$, then p has a conjugate point along γ.

Proof. We may suppose that $\gamma\colon [0, \pi/C] \to M$. Let σ be the sign of $\langle\gamma', \gamma'\rangle$. We construct an orthogonal non-trivial vector field V on γ such that $\sigma I(V, V) \leq 0$, and Theorem 11.7.10 then implies that there are conjugate points along γ.

Construct an orthonormal basis $\gamma', e_1, \ldots, e_{n-1}$ at $\gamma(0)$ and then parallel transport it along γ. Consider for $i = 2, \ldots, n-1$ the vector field

$$V_i = \sin(Ct)e_i.$$

Note that V vanishes at the endpoints. We calculate

$$\sigma I(V_i, V_i) = \int_0^{\pi/C} \big(C^2\cos^2(Ct) - \sin^2(Ct)R(e_i, \gamma', \gamma', e_i)\big)\,dt.$$

By summing along the orthogonal indices, since γ is cospacelike we get

$$\begin{aligned}\sum_{i=2}^{n} \sigma I(V_i, V_i) &= \int_0^{\pi/C} \big((n-1)C^2\cos^2(Ct) - \sin^2(Ct)Ric(\gamma', \gamma')\big)\,dt \\ &\leq \int_0^{\pi/C} \big((n-1)C^2\cos^2(Ct) - (n-1)C^2\sin^2(Ct)\big)\,dt = 0.\end{aligned}$$

Therefore there is a V_i such that $\sigma I(V_i, V_i) \leq 0$. □

As for Cartan – Hadamard Theorem, we get a beautiful application for Riemannian manifolds. For a constant $c \in \mathbb{R}$, we write $Ric \geq c$ to indicate that $Ric(v, v) \geq c$ for any unit vector $v \in T_pM$ at any $p \in M$. Equivalently, the tensor $Ric - cg$ is positive semidefinite at any $p \in M$.

Theorem 11.7.13 (Bonnet – Myers). *Let M be a complete connected Riemannian manifold with $Ric \geq (n-1)C^2 > 0$. Then M is compact with diameter $\leq \pi/C$, and $\pi_1(M)$ is finite.*

The *diameter* of a metric space X is the sup of $d(p,q)$ as $p,q \in X$ vary.

Proof. We prove that the diameter is $\leq \pi/C$. By contradiction, suppose that $p,q \in M$ have distance $> \pi/C$. Since M is complete, there is a minimising geodesic γ connecting p and q, see Proposition 10.3.2. Since $L(\gamma) > \pi/C$, the point p has a conjugate along γ by Proposition 11.7.12. Theorem 11.7.10 hence says that $I(\gamma)$ is indefinite and thus γ is not minimising, a contradiction.

Finite diameter and complete easily imply compact. The universal cover $\tilde{M}$ inherits a Riemannian structure with the same inequality $Ric \geq (n-1)C^2 > 0$. By what just said $\tilde{M}$ is compact, hence the covering $\tilde{M} \to M$ has finite index and $\pi_1(M)$ is finite. □

Remark 11.7.14. It is not enough to require $\mathrm{Ric} > 0$, since the paraboloid $z = x^2 + y^2$ in $\mathbb{R}^3$ has $K > 0$ everywhere and is not compact.

11.8. Locally symmetric spaces

We introduce a class of pseudo-Riemannian manifolds that contains the constant curvature ones and can be studied elegantly within a uniform framework. These are the manifolds where the Riemann tensor R is constant and are called *locally symmetric spaces*.

11.8.1. Definition. As alluded in the introduction, we say that a connected pseudo-Riemannian manifold (M,g) is a *locally symmetric space* if $\nabla R = 0$. Here is an important example.

Proposition 11.8.1. *If M has constant curvature, it is locally symmetric.*

Proof. We know from Proposition 11.2.3 that R is constructed from g by tensor products and linear combinations, hence $\nabla g = 0$ implies $\nabla R = 0$. □

The product of two locally symmetric spaces is locally symmetric: note that this is not true for constant curvature manifolds (unless both manifolds have zero constant curvature). Therefore locally symmetric spaces form a strictly larger class than constant curvature ones, as they include for instance $S^2 \times S^2$ or $S^2 \times \mathbb{R}$.

11.8.2. Polar maps. Let (M,g) and (N,h) be two pseudo-Riemannian manifolds. Consider two points $p \in M$ and $q \in N$. Let $f\colon T_pM \to T_qN$ be a given linear map. For every normal neighbourhood Z of p such that $f(\exp_p^{-1}(Z)) \subset V_q$ we have a well-defined *polar map*

$$\varphi = \exp_q \circ f \circ \exp_p^{-1}\colon Z \longrightarrow N.$$

If f is an isomorphism and Z is sufficiently small, this map is a diffeomorphism onto its image, a normal neighbourhood of q. In general we have

$$\varphi(\gamma_v(t)) = \gamma_{f(v)}(t) \tag{48}$$

for any $v \in T_pM$ and for any t such that $\gamma_v(t) \in Z$. We get

$$d\varphi_p = f.$$

If N is complete, the polar map is defined on any normal neighbourhood Z of p. The following proposition says that any local isometry sending p to q must be the extension of a polar map of a linear isometry.

Proposition 11.8.2. *If $\psi\colon M \to N$ is a local isometry sending p to q, then $\varphi = \psi|_Z$ for any polar map $\varphi\colon Z \to N$ of its differential $d\psi_p$.*

Proof. A local isometry ψ sends geodesics to geodesics and therefore (48) is fulfilled also by ψ, with $f = d\psi_p$. □

11.8.3. Local isometries. A nice feature of locally symmetric spaces is that every isometry of tangent spaces that preserves the Riemann tensors may be realised locally by a local isometry of manifolds.

Lemma 11.8.3. *Let (M, g) and (N, h) be locally symmetric spaces. Let $p \in M$, $q \in N$ and $f : T_pM \to T_qN$ be a linear isomorphism. Suppose that f preserves both the metric and the Riemann tensors, that is*

$$f^*(h(q)) = g(p), \qquad f^*(R^N(q)) = R^M(p).$$

Then any polar map φ of f is a local isometry.

Proof. Let $\varphi\colon Z \to \varphi(Z)$ be a polar map. We must show that for every $\exp_p(v) \in Z$ the differential $d\varphi_{\exp_p(v)} = d\varphi_{\gamma_v(1)}$ is an isometry. Recall that

$$\varphi(\gamma_v(t)) = \gamma_{f(v)}(t).$$

Fix an orthonormal basis $e_1, \ldots, e_n$ for T_pM and parallel transport it along the geodesic γ_v. Since f preserves the metric tensor, the basis $e'_1 = f(e_1), \ldots, e'_n = f(e_n)$ of T_qN is also orthonormal and we parallel transport it along the geodesic $\gamma_{f(v)}$.

Since f preserves the Riemann tensors, both R^M and R^N have the same coordinates ${R_{ijk}}^l$ in p and q with respect to the chosen basis. Moreover, since $\nabla R^M = 0$ and $\nabla R^N = 0$, the coordinates of R^M and R^N are constantly the same ${R_{ijk}}^l$ at every point of γ_v and $\gamma_{f(v)}$ with respect to the transported basis.

By what just said, the Jacobi equations along these two geodesics, written in coordinates as in (43), are exactly the same. By (44) the differential $d\varphi_{\gamma_v(1)}$, written in coordinates, is the identity, hence an isometry. □

This lemma has important consequences. We refer to Section 10.4.3.

Corollary 11.8.4. *Every locally symmetric space M is locally homogeneous.*

Proof. For every $p \in M$, let $A_p \subset M$ be the subset consisting of all $q \in M$ such that there is an isometry $\varphi\colon U(p) \to V(q)$ of neighbourhoods sending p to q. We prove that $A_p = M$.

Let Z be a normal neighbourhood of p. By parallel-transporting along radial geodesics we construct for every $q \in Z$ an isomorphism $T_pM \to T_qM$ that preserves both the metric and the Riemann tensor (since $\nabla g = 0$ and $\nabla R = 0$). By Lemma 11.8.3 we get $q \in A_p$. Therefore $Z \subset A_p$ and hence A_p is open. The connected manifold M is partitioned in open non-empty subsets $\{A_p\}_{p\in M}$, hence $A_p = M$ for every $p \in M$. □

Two locally homogeneous manifolds M and N are *locally isometric* if the disjoint union $M \sqcup N$ is again locally homogeneous. That is, for some (equivalently, every) $p \in M, q \in N$ there is an isometry $U(p) \to V(q)$ of neighbourhoods sending p to q.

Corollary 11.8.5. *Two symmetric spaces M, N are locally isometric $\Longleftrightarrow$ there is a curvature-preserving linear isometry $T_pM \to T_qN$ for some p, q.*

Corollary 11.8.6. *Every constant curvature manifold M is locally homogeneous. Two pseudo-Riemannian manifolds M and N with the same signature and the same constant sectional curvature K are locally isometric.*

Proof. The signature and K determine R, see Proposition 11.2.3. □

If a pseudo-Riemannian manifold M has constant curvature K, it is harmless to suppose that $K \in \{-1, 0, 1\}$, since this can always be achieved by rescaling the metric by an appropriate factor.

Corollary 11.8.7. *A pseudo-Riemannian M with signature (p, q) and constant curvature $1, 0, -1$ is locally isometric respectively to $S^{p,q}$, $\mathbb{R}^{p,q}$, $H^{p,q}$.*

The following corollary explains the term "locally symmetric" and furnishes an alternative curvature-free definition of locally symmetric spaces.

Corollary 11.8.8. *A semi-Riemannian manifold M is locally symmetric $\Longleftrightarrow$ at every $p \in M$ any polar map of $-\mathrm{id}\colon T_pM \to T_pM$ is a local isometry.*

Proof. ($\Rightarrow$). The linear map $-\mathrm{id}$ preserves both the metric and the Riemann tensor, so by the lemma any polar map is an isometry.

($\Leftarrow$). The polar map is an isometry and hence preserves R and ∇R. Since its differential at p is $-\mathrm{id}$ we deduce that for every $u, v, w, z \in T_pM$

$$\nabla_u R(v, w, z) = -\nabla_{-u} R(-v, -w, -z) = -\nabla_u R(v, w, z)$$

and therefore $\nabla R(p) = 0$. This holds at every $p \in M$. □

In other words, M is locally symmetric $\Longleftrightarrow$ at every p there is a local isometry that fixes p and whose differential at T_pM is $-\mathrm{id}$ (such a local isometry must restrict to a polar map).

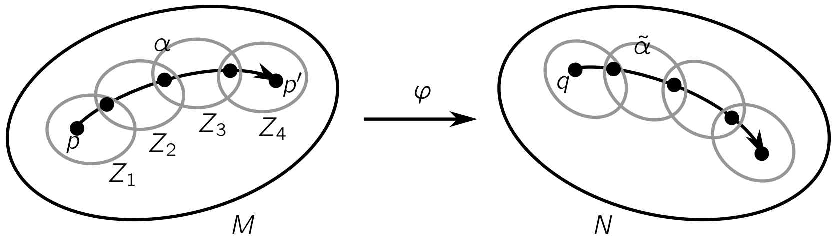

Figure 11.3. The developing map φ. The dots on the left indicate the points $p = \alpha(0)$, $\alpha(t_1), \alpha(t_2), \alpha(t_3), \alpha(1) = q$

11.8.4. Developing map. We would like to extend the local isometries found in the previous section to a global map. To accomplish this task we need a simply connected domain and a geodesically complete target.

Theorem 11.8.9. *Let M, N be locally symmetric spaces. If M is simply connected and N is geodesically complete, any curvature-preserving linear isometry $f : T_pM \to T_qN$ is the differential of a unique local isometry $\varphi\colon M \to N$.*

Proof. The construction of φ from f is similar to that (usually taught in the topology courses) of the lift of a map $X \to Y$ to a $X \to \tilde{Y}$ when $\tilde{Y} \to Y$ is a covering and X is simply connected. The role of "well-covered subsets" is somehow played here by the "totally normal subsets".

For any point $p' \in M$, we define $\varphi(p')$ as follows. Pick an arc $\alpha\colon [0, 1] \to M$ joining p and p'. Cover the arc with finitely many totally normal sets $Z_1, \ldots, Z_k$. We may suppose that $0 = t_0 < t_1 < \cdots < t_k = 1$ and $\alpha([t_{i-1}, t_i]) \subset Z_i$ for all i, see Figure 11.3. We define inductively a local isometry $\varphi_i\colon Z_i \to N$ for $i = 1, \ldots, k$, by applying Lemma 11.8.3 to f for φ_1 and then to $(d\varphi_i)_{t_i}$ for φ_{i+1} for each $i > 0$.

By construction the local isometries φ_i glue along α and project it to a new smooth curve $\tilde{\alpha}\colon [0, 1] \to N$. More precisely, we set $\tilde{\alpha}(t) = \varphi_i(t)$ if $t \in [t_{i-1}, t_i]$, see Figure 11.3. We define $\varphi(p') = \tilde{\alpha}(1)$.

We leave as an exercise to show that the curve $\tilde{\alpha}$ does not depend on the chosen covering $\{Z_i\}$. If we pick another curve α' joining p to p', since M is simply connected there is a homotopy (fixing the endpoints) from α to α', which can be projected (exercise) to a homotopy (fixing the endpoints) from $\tilde{\alpha}$ to $\tilde{\alpha}'$. Therefore $\tilde{\alpha}(1) = \tilde{\alpha}'(1)$ and $\varphi(p')$ is uniquely defined.

We have defined a map $\varphi\colon M \to N$. By prolonging α smoothly with curves that lie in Z_k we easily get that $\varphi|_{Z_k} = \varphi_k$ is a local isometry near p'. Since p' is generic, we get that φ is a local isometry. $\square$

The local isometry φ is sometimes called the *developing map* of f.

Corollary 11.8.10. *Two geodesically complete simply connected locally symmetric spaces M and N are isometric $\Longleftrightarrow$ there is a curvature-preserving linear isometry $T_pM \to T_qN$ for some p, q.*

Proof. ($\Rightarrow$) is obvious. ($\Leftarrow$) From the theorem we get a local isometry $\varphi\colon M \to N$, which is a pseudo-Riemannian covering by Exercise 10.4.7. Since M and N are both simply connected, φ is an isometry. $\square$

11.8.5. Manifolds with constant curvature. Theorem 11.8.9 has a really strong impact to the theory of constant curvature pseudo-Riemannian manifolds. We start by analysing the Riemannian ones.

Corollary 11.8.11. *Every complete simply connected Riemannian manifold with constant curvature $-1, 0, 1$ is isometric to $\mathbb{H}^n$, $\mathbb{R}^n$, S^n respectively.*

Completeness is of course crucial: many open subsets of $\mathbb{H}^n$, $\mathbb{R}^n$, S^n are simply connected and not complete, and have constant curvature: these are certainly not isometric to $\mathbb{H}^n$, $\mathbb{R}^n$, S^n.

Before stating a similar result for more general pseudo-Riemannian manifolds, let us recall that $S^{p,q}$ and $H^{p,q}$ are geodesically complete, and also simply connected (see Proposition 11.5.6) in all cases except $S^{0,n} \cong H^{n,0} \cong \mathbb{R}^n \sqcup \mathbb{R}^n$ and $S^{1,n-1} \cong H^{n-1,1} \cong S^1 \times \mathbb{R}^{n-1}$. Therefore we define $\tilde{S}^{0,n}$ and $\tilde{H}^{n,0}$ as one of the two components of $S^{0,n}$ and $H^{n,0}$, and $\tilde{S}^{1,n-1}$, $\tilde{H}^{n-1,1}$ as the universal covers of $S^{1,n-1}$, $H^{n-1,1}$, with the induced metric (which is still geodesically complete by Exercise 10.4.8).

Corollary 11.8.12. *Every geodesically complete simply connected pseudo-Riemannian manifold with signature (p, q) and constant curvature $-1, 0, 1$ is isometric to $H^{p,q}$ ($\tilde{H}^{p,q}$ if $q \leq 1$), $\mathbb{R}^{p,q}$, $S^{p,q}$ ($\tilde{S}^{p,q}$ if $p \leq 1$) respectively.*

Let us denote for simplicity by $\mathbb{X}_K^{(p,q)}$ the unique simply connected pseudo-Riemannian manifold with signature (p, q) and constant curvature $K = -1, 0, 1$

We now drop the simply connected hypothesis and obtain an elegant theorem that uses many of the techniques that we have seen to characterise algebraically all the complete manifolds with constant curvature and arbitrary signature.

Theorem 11.8.13. *Every geodesically complete pseudo-Riemannian manifold with signature (p, q) and constant curvature $K \in \{-1, 0, 1\}$ is obtained as a quotient $\mathbb{X}_K^{(p,q)}/_\Gamma$ for some subgroup $\Gamma < \mathrm{Isom}\big(\mathbb{X}_K^{(p,q)}\big)$ acting freely and property discontinuously. We get a 1-1 correspondence*

$$\left\{\begin{array}{c}\text{geodesically complete manifolds}\\ \text{with constant curvature } K\\ \text{and signature } (p,q)\\ \text{up to isometry}\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{subgroups } \Gamma < \mathrm{Isom}\big(\mathbb{X}_K^{(p,q)}\big)\\ \text{acting freely}\\ \text{and properly discontinuously}\\ \text{up to conjugation}\end{array}\right\}.$$

Proof. Every $M = \mathbb{X}^{(p,q)}_K/_\Gamma$ with $\Gamma < \mathrm{Isom}(\mathbb{X}^{(p,q)})$ acting freely and properly discontinuously is a smooth manifold that inherits a constant curvature structure, and is geodesically complete by Exercise 10.4.8. Conversely, given a geodesically complete constant curvature M, its universal cover also inherits a geodesically complete constant curvature structure and is simply connected, hence it is isometric to $\mathbb{X}^{(p,q)}_K$ by Corollary 11.8.12. Since the universal covering is regular, by Proposition 1.2.8 we have $M = \mathbb{X}^{(p,q)}/_\Gamma$ where Γ is the deck transformations group, that acts freely and properly discontinuously. Moreover Γ acts and by isometries by construction.

When passing from the manifold M to the group Γ, the only choice we made is an isometry between the universal cover of M and $\mathbb{X}^{(p,q)}_K$. Different choices produce conjugate groups Γ. This shows the 1-1 correspondence. $\square$

11.9. Miscellaneous facts

11.9.1. Killing fields on manifolds with negative Ricci curvature. Let X be a Killing field on a Riemannian manifold M. We study the function

$$f(p) = \frac{1}{2}\|X(p)\|^2.$$

We may compute its gradient, Hessian, and Laplacian.

Lemma 11.9.1. *The following holds:*

(1) $\mathrm{grad} f = -\nabla_X X$,
(2) $(\nabla^2 f)(v,v) = \|\nabla_v X\|^2 - R(v,X,X,v)$,
(3) $\Delta f = -\|\nabla X\|^2 + Ric(X,X)$.

Proof. Extend v to a vector field V. Recall from Proposition 10.4.13 that ∇X is a skew-adjoint $(1,1)$ tensor field and $\langle \nabla_v X, v\rangle = 0$. Therefore

$$\begin{aligned}
\nabla_v f &= \langle \nabla_v X, X\rangle = -\langle v, \nabla_X X\rangle, \\
\mathrm{grad} f &= -\nabla_X X, \\
(\nabla^2 f)(v,v) &= \langle \nabla_v(\mathrm{grad} f), v\rangle = -\langle \nabla_v \nabla_X X, v\rangle \\
&= -\langle \nabla_X \nabla_V X, V\rangle - \langle \nabla_{[V,X]} X, V\rangle - R(V,X,X,V) \\
&= -X\langle \nabla_V X, V\rangle + \langle \nabla_V X, \nabla_X V\rangle + \langle [V,X], \nabla_V X\rangle - R(V,X,X,V) \\
&= \langle \nabla_V X, \nabla_V X\rangle - R(V,X,X,V),
\end{aligned}$$

If we pick an orthonormal basis at the point we find

$$\begin{aligned}
\Delta f = -g^{ij}\nabla_i \nabla_j f &= -\sum_{i=1}^n \|\nabla_i X\|^2 + \sum_{i=1}^n R(e_i, X, X, e_i) \\
&= -\|\nabla X\|^2 + Ric(X,X).
\end{aligned}$$

The proof is complete. $\square$

Here is an interesting consequence.

Theorem 11.9.2 (Bochner). *If M is a compact oriented Riemannian manifold with $Ric \leq 0$, every Killing field is parallel. If $Ric < 0$, there are no non-trivial Killing fields.*

Proof. Let X be a Killing field and set $f(p) = \frac{1}{2}\|X(p)\|^2$. Corollary 10.5.13 and Lemma 11.9.1 give

$$0 = \int_M \Delta f = \int_M -\|\nabla X\|^2 + \int_M Ric(X, X).$$

If $Ric \leq 0$ we get $\nabla X = 0$ and $Ric(X, X) = 0$. If $Ric < 0$ we get $X = 0$. □

Compactness is required, since the Euclidean $\mathbb{R}^2$ has the Killing field $X(x, y) = (-y, x)$, that is clearly not parallel. Note also that every constant vector field in $\mathbb{R}^n$ descends to a parallel Killing field in the flat torus $T = \mathbb{R}^n/_{\mathbb{Z}^n}$. Hence $Ric \leq 0$ is not enough to exclude the presence of non-trivial Killing fields.

11.9.2. Killing fields on manifolds with positive sectional curvature. We now prove the following.

Proposition 11.9.3 (Berger). *On an even-dimensional compact Riemannian manifold M with positive sectional curvature, every Killing field X has a zero.*

Proof. Consider again $f(p) = \frac{1}{2}\|X(p)\|^2$. Suppose by contradiction that X has no zeroes. Then f has a global positive minimum at some p. Being a minimum, we have $\mathrm{grad} f(p) = 0$ and $\nabla^2 f(p) \geq 0$. Therefore at p we get

$$\nabla_X X = 0, \qquad \|\nabla_v X\|^2 - R(v, X, X, v) \geq 0$$

for every $v \in T_pM$. Since M is even-dimensional, the kernel of the skew-adjoint operator ∇X at p is also even-dimensional. It contains $X(p)$, and hence also some other vector v linear independent from $X(p)$. Hence $\nabla_v X = 0$, and the hypothesis on the sectional curvature gives $R(v, X, X, v) > 0$, a contradiction. □

This fact is not true on odd-dimensional manifolds, since on S^{2n-1} we may have the Killing field $X(x^1, \ldots, x^{2n}) = (x^2, -x^1, \ldots, x^{2n}, -x^{2n-1})$.

11.10. Exercises

Exercise 11.10.1. Let $\varphi\colon M \to M$ be an isometry of a Riemannian manifold. Show that the fixed points form a disjoint union of closed geodesic submanifolds of M (possibly of different dimensions).

Exercise 11.10.2. Let X be a Killing vector field on a Riemannian manifold (M, g). Show that X restricts to a Jacobi field on any geodesic. Deduce that X is determined by the values of $X(p)$ and $\nabla X(p)$ at any point $p \in M$. Remembering that $\nabla X(p)$ is antisymmetric, deduce that the Killings field form a Lie algebra of dimension at most $(n+1)n/2$.

Exercise 11.10.3. Let X be a Killing vector field on a Riemannian manifold (M, g). The zero set of X is a disjoint union of geodesic submanifolds of even codimension.

CHAPTER 12

Lie groups

A *Lie group* is a group that is also a smooth manifold. Lie groups are everywhere: most symmetry groups that one encounters in geometry are naturally Lie groups. The fundamental examples are matrix groups like $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{O}(n)$.

12.1. Basics

We define the Lie groups and start to investigate their properties.

12.1.1. Definition. A *Lie group* is a smooth manifold G equipped with a group structure, such that the multiplication and inverse maps

$$G \times G \longrightarrow G, \qquad (g, h) \longmapsto gh,$$
$$G \longrightarrow G, \qquad g \longmapsto g^{-1}$$

are both smooth. This is equivalent to requiring the map $G \times G \to G$, $(g, h) \mapsto gh^{-1}$ to be smooth.

Here are some important examples.

Example 12.1.1 (Abelian). The first examples of Lie groups are $\mathbb{R}^n$ with the sum operation and S^1 with the product, where we see $S^1 \subset \mathbb{C}$ as the unit complex numbers. These Lie groups are abelian.

Example 12.1.2 (Linear and orthogonal groups). A more elaborated and equally important example is the *general linear group* $\mathrm{GL}(n, \mathbb{R})$ of all $n \times n$ invertible matrices with the product operation. This Lie group contains also many other interesting Lie groups, such as the *special linear group* $\mathrm{SL}(n, \mathbb{R})$, the *orthogonal group* $\mathrm{O}(n)$, and the *special orthogonal group* $\mathrm{SO}(n)$. We studied the topology of these manifolds in Section 3.9.

Example 12.1.3 (Products). The product $G \times H$ of two Lie groups is naturally a Lie group. For instance, the n-torus $S^1 \times \cdots \times S^1$ is an abelian compact Lie group of dimension n.

Example 12.1.4 (Affine transformations). Another example is the group $\mathrm{Aff}(\mathbb{R}^n) = \mathrm{GL}(n, \mathbb{R}) \rtimes \mathbb{R}^n$ of all affine transformations of $\mathbb{R}^n$. As a set, we have $\mathrm{Aff}(\mathbb{R}^n) = \mathrm{GL}(n, \mathbb{R}) \times \mathbb{R}^n$ and we use this bijection to assign a smooth manifold structure to $\mathrm{Aff}(\mathbb{R}^n)$. The group structure is not a direct product, but the group operations are smooth nevertheless.

A Lie group of dimension 0 is called *discrete*. Every countable group G like $\mathbb{Z}$ may be given the structure of a Lie group by assigning it the discrete topology. Of course a discrete Lie group is connected if and only if it is trivial.

12.1.2. Homomorphisms. A *Lie group homomorphism* is a smooth homomorphism $f\colon G \to H$ between Lie groups. As usual, this is an *isomorphism* if f is invertible, that is if f is a diffeomorphism, and an *automorphism* if in addition $G = H$. For instance, every conjugation $G \to G, x \mapsto g^{-1}xg$ by some fixed element $g \in G$ is an automorphism of the Lie group G.

Example 12.1.5. The Lie groups S^1 and $\mathrm{SO}(2)$ are isomorphic, via the map

$$S^1 \longrightarrow \mathrm{SO}(2), \qquad e^{i\theta} \longmapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

12.1.3. Left and right multiplication. If $g \in G$, the *left and right multiplications by g* are the maps

$$L_g\colon G \to G, \quad x \mapsto gx,$$
$$R_g\colon G \to G, \quad x \mapsto xg.$$

Both maps are diffeomorphisms, with inverses $L_{g^{-1}}$ and $R_{g^{-1}}$, but are not Lie group isomorphisms, unless $g = e$. The maps L_g and $R_{g'}$ commute for all $g, g' \in G$. Conjugation by g is just $L_{g^{-1}} \circ R_g$.

12.1.4. Lie subgroups. Let G be a Lie group. A *Lie subgroup* of G is the image of any injective Lie group homomorphism $H \hookrightarrow G$ that is also an immersion. We identify H with its image and write $H < G$. For instance, $\mathrm{O}(n)$ is a Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$.

We require H to be "injectively immersed" in G instead of the stronger and nicer "embedded" because we do not want to rule out the following types of Lie subgroups:

Example 12.1.6. Pick $\lambda \in \mathbb{R} \setminus \mathbb{Q}$ and consider the injective immersion $\mathbb{R} \to S^1 \times S^1, t \mapsto (e^{2\pi i t}, e^{2\pi i \lambda t})$. The image is a dense Lie subgroup of $S^1 \times S^1$. See Exercise 5.5.4.

The reason for allowing non-embedded Lie subgroups will be apparent in the next section. We exhibit more examples.

Example 12.1.7. The Lie group $\mathrm{Aff}(\mathbb{R}^n)$ may be embedded as a Lie subgroup of $\mathrm{GL}(n+1, \mathbb{R})$, by representing the affine transformation $x \mapsto Ax + b$ via the matrix

$$\begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}.$$

Example 12.1.8. The *Heisenberg group* is the Lie subgroup of $\mathrm{SL}(3,\mathbb{R})$ formed by all the matrices

$$\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$$

where $a, b, c \in \mathbb{R}$ vary. It is diffeomorphic to $\mathbb{R}^3$, but it is not abelian.

12.1.5. Identity connected component. Let G be a Lie group. We denote by $G^0 \subset G$ the connected component of G containing the identity $e \in G$. The following may be seen as the first interesting result in Lie groups theory. The proof mixes topological and group theory arguments.

Proposition 12.1.9. *The component G^0 is a normal Lie subgroup.*

Proof. For every $g \in G$, the left multiplication L_g is a diffeomorphism and hence permutes the connected components of G. If $g \in G^0$, then L_g sends e to g and hence sends G^0 to itself. Therefore $gh \in G^0$ for all $g, h \in G^0$, so G^0 is closed under multiplication.

Analogously, the inverse map $g \mapsto g^{-1}$ permutes the connected components of G and fixes e, hence leaves G^0 invariant. Therefore G^0 is a subgroup. Along the same line, for every $g \in G$ the conjugation $x \mapsto g^{-1}xg$ is a diffeomorphism that fixes e and hence leaves G^0 invariant. So G^0 is normal. □

The quotient G/G^0 is naturally a discrete Lie group.

Example 12.1.10. We have $\mathrm{O}(n)^0 = \mathrm{SO}(n)$, while $\mathrm{GL}(n,\mathbb{R})^0$ consists of all invertible matrices with positive determinant.

12.1.6. Identity neighbourhoods. Let G be a Lie group. If $U, V \subset G$ are subsets, we construct more subsets as follows:

$$UV = \{uv \mid u \in U, v \in V\}, \qquad U^{-1} = \{u^{-1} \mid u \in U\}.$$

If U, V are neighbourhoods of the identity, then both UV and U^{-1} also are. We can use this to prove the following.

Proposition 12.1.11. *If G is connected, any neighbourhood U of the identity generates G.*

Proof. We can suppose that U is open and $U = U^{-1}$, otherwise we substitute U with $U \cap U^{-1}$. The subgroup generated by U is $H = \cup_{n=1}^{\infty} U^n$. Each U^n is open, so H is an open subgroup of G. Its left cosets are also open. Since G is connected, we get $G = H$. □

12.1.7. Universal cover. Let G be a connected Lie group, and $\tilde{G}$ be its universal cover. We show that the Lie group structure lifts from G to $\tilde{G}$.

Proposition 12.1.12. *There is a natural Lie group structure on $\tilde{G}$ such that the cover $\pi\colon \tilde{G} \to G$ is a Lie groups homomorphism.*

Proof. We fix an arbitrary identity $\tilde{e} \in \pi^{-1}(e)$. Since $\tilde{G}$ is simply connected, both the product $G \times G \to G$ and the inversion $G \to G$ lift to two smooth maps $\tilde{G} \times \tilde{G} \to \tilde{G}$ and $\tilde{G} \to \tilde{G}$ between the universal covers, such that $(\tilde{e}, \tilde{e})$ goes to $\tilde{e}$ and $\tilde{e}$ goes to $\tilde{e}$, respectively. These define a product and inverse structure on $\tilde{G}$. Using the unique lift property of paths we can prove that these indeed satisfy the group axioms (exercise). $\square$

We have discovered that every connected Lie group has a universal cover. The universal cover of S^1 is of course $\mathbb{R}$. For $n \geq 3$, the *spin group* is defined as the universal cover of $\mathrm{SO}(n)$:

$$\mathrm{Spin}(n) = \widetilde{\mathrm{SO}(n)}.$$

12.1.8. Coverings. Let a *covering of Lie groups* be a homomorphism of connected Lie groups $G \to H$ that is also a smooth covering. The universal cover $\tilde{G} \to G$ constructed above is one example. In general, it is quite easy to understand when a Lie group homomorphism is a covering.

Proposition 12.1.13. *A Lie group homomorphism $f\colon G \to H$ between connected Lie groups is a smooth covering $\Longleftrightarrow$ df_e is invertible.*

Proof. The implication $\Rightarrow$ is obvious, so we prove $\Leftarrow$. Since df_e is invertible, there are open neighbourhoods U and V of $e \in G$ and $e \in H$ such that f maps diffeomorphically U to V.

For every $h \in H$, and every $g \in f^{-1}(h)$, we define

$$V_h = L_h(V), \qquad U_g = L_g(U).$$

These are open neighbourhoods of h and g, and one sees easily that

$$f^{-1}(V_h) = \bigsqcup_{g \in f^{-1}(h)} U_g.$$

The restriction of f to U_g is a diffeomorphism onto V_h, therefore f is a smooth covering. $\square$

Here is a concrete way to build coverings of Lie groups:

Proposition 12.1.14. *Let G be a Lie group and $\Gamma < Z(G)$ be a discrete central subgroup. The quotient G/Γ is naturally a Lie group and $G \to G/\Gamma$ is a regular covering of Lie groups, with deck transformation group Γ.*

Proof. The action of Γ on G by multiplication is smooth, free, and properly discontinuous (exercise). Proposition 3.5.5 applies. $\square$

We now want to prove a converse of this proposition.

Proposition 12.1.15. *Let G be a connected Lie group. Every discrete normal subgroup $\Gamma \subset G$ is central.*

Proof. Pick $\gamma \in \Gamma$. For every $g \in G$, choose a path $g_t \in G$ connecting $g_0 = e$ and $g_1 = g$. By normality $g_t^{-1}\gamma g_t$ is a path in Γ, that must be constant, so $g^{-1}\gamma g = \gamma$ for all $g \in G$. □

Here is a converse for Proposition 12.1.14:

Proposition 12.1.16. *Every covering of Lie groups $G \to H$ is as in Proposition 12.1.14. That is, $\Gamma = \ker G$ is discrete and central and $H = G/\Gamma$.*

Proof. The kernel Γ is the fibre of e and is hence discrete. Being also normal, it is central by the previous proposition. □

By assembling all our discoveries, we obtain the following.

Corollary 12.1.17. *Every connected Lie group is a quotient G/Γ of a simply connected Lie group G along some discrete central subgroup Γ.*

The classification of connected Lie groups hence reduces to the classification of simply connected ones (and their discrete central subgroups). The classification of simply connected Lie groups is hence a fundamental topological problem, that is elegantly transformed into an algebraic one through the fundamental notion of *Lie algebra* that we introduce in the next section.

We close our investigation with a corollary.

Corollary 12.1.18. *The fundamental group of every Lie group is abelian.*

12.2. Lie algebra

One of the most important aspects of Lie groups G is the leading role played by the tangent space T_eG at the identity $e \in G$, that has a natural structure of *Lie algebra*, see Definition 5.4.2.

12.2.1. Left-invariant vector fields. Let G be a Lie group. We now consider the tangent space T_eG at the identity $e \in G$. We note that for every $g \in G$ the differential of L_g yields an isomorphism

$$(dL_g)_e \colon T_eG \longrightarrow T_gG$$

on tangent spaces. Therefore we can use left-multiplication to identify canonically all the tangent spaces to T_eG, and this is a crucial aspect of Lie groups.

In particular, every fixed vector $v \in T_eG$ extends canonically to a vector field X in G by left-multiplication, as follows:

$$X(g) = (dL_g)_e(v).$$

The vector field X is *left-invariant*, that is it is invariant under the diffeomorphisms L_h, for all $h \in G$. Indeed we have

$$X(hg) = (dL_{hg})_e(v) = (dL_h)_g \circ (dL_g)_e(v) = (dL_h)_g(X(g)).$$

Every left-invariant vector field is clearly constructed in this way. We have obtained a natural isomorphism between T_eG and the subspace of $\mathfrak{X}(G)$ consisting of all the left-invariant vector fields. (Recall that $\mathfrak{X}(G)$ is the space of all vector fields in G.) We will henceforth identify these two spaces along this isomorphism.

By replacing L_g with R_g in the construction we would get analogously a natural isomorphism between T_eG and the subspace of all *right*-invariant vector fields. Note that a left-invariant vector field is not necessarily right-invariant, so the two subspaces of $\mathfrak{X}(G)$ may differ.

12.2.2. Parallelisability. The first important consequence that we can draw form our discovery is the following.

Proposition 12.2.1. *Every Lie group G is parallelisable.*

Proof. Every basis $v_1, \ldots, v_n$ of T_eG extends by left-multiplication to n left-invariant vector fields $X_1, \ldots, X_n$ on G that trivialise the bundle. □

Corollary 12.2.2. *Every Lie group G is orientable.*

12.2.3. Lie algebra. Let G be a Lie group. We have identified T_eG with the subspace of left-invariant vector fields in $\mathfrak{X}(G)$. We now note the following.

Proposition 12.2.3. *If $X, Y \in \mathfrak{X}(G)$ are left-invariant, then $[X, Y]$ also is.*

Proof. If two vector fields X, Y are invariant under some diffeomorphism, then their bracket also is. □

This observation shows that the space T_eG of all left-invariant vector fields is closed under the Lie bracket $[,]$. In other words T_eG is a *Lie subalgebra* of $\mathfrak{X}(G)$, and it is such an important object that it deserves a new symbol:

$$\mathfrak{g} = T_eG.$$

This is the *Lie algebra* of the Lie group G. The Lie algebra of Lie groups like $\mathrm{GL}(n, \mathbb{R})$, $\mathrm{O}(n)$, etc. is usually denoted as $\mathfrak{gl}(n, \mathbb{R})$, $\mathfrak{o}(n)$, etc.

12.2.4. Examples. On $\mathbb{R}^n$, a vector field is left-invariant if and only if it is constant, and the bracket of two constant vector fields is zero. Therefore the Lie algebra of $\mathbb{R}^n$ is $\mathbb{R}^n$ with the trivial Lie bracket. A Lie algebra with trivial Lie bracket is called *abelian*.

Analogously, the Lie algebra of S^1 is $\mathbb{R}$ with trivial Lie bracket. The Lie algebra of a product of Lie groups is just the product of their Lie algebras: in particular the Lie algebra of $S^1 \times \cdots \times S^1$ is again $\mathbb{R}^n$ with the trivial Lie bracket.

A more interesting example is $\mathrm{GL}(n, \mathbb{R})$. Being an open subset of the vector space $M(n)$ of all $n \times n$ matrices, its Lie algebra $\mathfrak{gl}(n, \mathbb{R})$ is $M(n)$ as a vector space, and we only need to understand the Lie bracket.

Proposition 12.2.4. *The Lie bracket of* $A, B \in \mathfrak{gl}(n, \mathbb{R})$ *is*

$$[A, B] = AB - BA.$$

Proof. Since $\mathrm{GL}(n, \mathbb{R})$ is an open subset of $M(n)$, a vector field is simply a map $\mathrm{GL}(n, \mathbb{R}) \to M(n)$. Every vector $A \in M(n)$ tangent at the origin extends by left-multiplication to the vector field $X \mapsto XA$. Similarly to Exercise 5.4.7, one can check (exercise) that the bracket of two vector fields $X \mapsto XA$ and $X \mapsto XB$ is $X \mapsto X(AB - BA)$. $\square$

In particular, the Lie algebra $\mathfrak{gl}(n, \mathbb{R})$ is non-abelian as soon as $n \geq 2$.

12.2.5. Homomorphisms. Every Lie group homomorphism $f : G \to H$ induces a linear map $f_* : \mathfrak{g} \to \mathfrak{h}$ which is just the differential $f_* = df_e$.

Proposition 12.2.5. *The map* $f_* : \mathfrak{g} \to \mathfrak{h}$ *is a Lie algebra homomorphism.*

Proof. The homomorphism f commutes with left-multiplication, that is

$$f \circ L_g = L_{f(g)} \circ f$$

for every $g \in G$. This implies that a left-invariant vector field $X \in \mathfrak{g}$ and its image $f_*(X) \in \mathfrak{h}$ are f-related. Exercise 5.4.8 says that for every $X, Y \in \mathfrak{g}$ the vector fields $[X, Y]$ and $[f_*(X), f_*(Y)]$ are also f-related, so $f_*([X, Y]) = [f_*(X), f_*(Y)]$ as required. $\square$

During the proof we have also discovered that for every $X \in \mathfrak{g}$ the vector fields X and $f_*(X)$ are f-related.

12.2.6. Lie subgroups. A Lie subgroup $H < G$ is by definition the image of an injective immersion and homomorphism, so by the previous discussion the Lie algebra $\mathfrak{h}$ of H is naturally a Lie subalgebra of $\mathfrak{g}$.

This implies in particular that the Lie algebra of any Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$ is completely determined as soon as we know its tangent space at the identity: there is no need of computing the Lie bracket again since it will always be $[A, B] = AB - BA$.

For instance, we know from Propositions 3.9.1 and 3.9.2 that

$$\mathfrak{sl}(n, \mathbb{R}) = \left\{ A \in M(n, \mathbb{R}) \mid \mathrm{tr} A = 0 \right\},$$
$$\mathfrak{o}(n, \mathbb{R}) = \mathfrak{so}(n, \mathbb{R}) = \left\{ A \in M(n, \mathbb{R}) \mid {}^{t}A = -A \right\},$$

where both $\mathfrak{sl}(n, \mathbb{R})$ and $\mathfrak{o}(n, \mathbb{R})$ are subalgebras of $\mathfrak{gl}(n, \mathbb{R})$. One verifies easily that they are indeed both closed under the Lie bracket multiplication.

12.2.7. From Lie subalgebras to Lie subgroups. Here is a striking application of the Frobenius Theorem.

Theorem 12.2.6. *Let G be a Lie group. For every subalgebra $\mathfrak{h} \subset \mathfrak{g}$ there is a unique connected Lie subgroup $H < G$ whose Lie algebra is $\mathfrak{h}$.*

Proof. A subalgebra $\mathfrak{h} \subset \mathfrak{g}$ is in particular a subspace of $\mathfrak{g} = T_eG$, and by left-multiplication it extends to a distribution D in G, defined as

$$D_g = (dL_g)_e(\mathfrak{h}) \subset T_gG \tag{49}$$

for every $g \in G$. Since $\mathfrak{h}$ is a subalgebra, the distribution D is involutive. To prove this, pick k left-invariant vector fields $X_1, \ldots, X_k$ generating $\mathfrak{h}$. By construction they are tangent to D. Since $\mathfrak{h}$ is a subalgebra, their brackets $[X_i, X_j]$ are still in $\mathfrak{h}$ and hence are also tangent to D. Now Exercise 5.5.10 shows that D is involutive.

By the Frobenius Theorem 5.5.9, there is a foliation $\mathscr{F}$ of G tangent to D. Let H be the leaf of $\mathscr{F}$ containing the identity e. It is an injectively immersed manifold in G, with tangent space $T_eH = \mathfrak{h}$. For every $g \in G$, the diffeomorphism L_g preserves D and hence permutes the leaves of $\mathscr{F}$. If $h \in H$, then $L_{h^{-1}}$ sends $h \in H$ to $e \in H$ and hence preserves the leaf H. This implies that H is a subgroup, and hence a Lie subgroup.

If $H < G$ is connected with Lie algebra $\mathfrak{h}$, we leave as an exercise to show that H must be obtained from $\mathfrak{h}$ in the way just described, so it is unique. $\square$

We have discovered a beautiful natural 1-1 correspondence:

$$\big\{\text{connected Lie subgroups of } G\big\} \longleftrightarrow \big\{\text{Lie subalgebras of } \mathfrak{g}\big\}.$$

We note that the subgroup $H < G$ corresponding to $\mathfrak{h}$ is not guaranteed to be embedded, and there is no easy way to understand from $\mathfrak{h}$ alone whether $H < G$ is embedded or not. In fact, the pleasure of obtaining such a powerful and elegant theorem is the main reason for allowing non-embedded Lie subgroups in our definition.

12.2.8. Foliations. The proof of Theorem 12.2.6 also displays a nice geometric phenomenon that is worth emphasising. Let G be a Lie group. Given a subalgebra $\mathfrak{h} \subset \mathfrak{g}$, by left-multiplication we get an integrable distribution D as in (49), and hence a foliation $\mathscr{F}$ of G. We write $\mathscr{F}_\mathfrak{h}$ to stress its dependence on $\mathfrak{h}$. The construction implies easily the following fact.

Proposition 12.2.7. *Let $H < G$ be a Lie subgroup with Lie subalgebra $\mathfrak{h} \subset \mathfrak{g}$. The left cosets of H are unions of leaves of the foliation $\mathscr{F}_\mathfrak{h}$.*

Corollary 12.2.8. *Every embedded Lie subgroup $H < G$ is closed.*

Proof. Every embedded union of leaves in a foliation is closed. $\square$

12.2.9. Local homomorphisms. We now pass from subgroups to homomorphisms; that is, we ask ourselves if every Lie algebra homomorphism should be induced by some Lie group homomorphism. This is true only locally.

A *local homomorphism* between two Lie groups G and H is a smooth map $f\colon U \to H$ defined on some neighbourhood U of $e \in G$, such that

$$f(ab) = f(a)f(b) \quad \forall a, b, ab \in U.$$

Here is a partial converse to Proposition 12.2.5.

Theorem 12.2.9. *Let G, H be Lie groups and $F: \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra homomorphism. There is a local homomorphism $f: U \to H$ with $df_e = F$.*

Proof. The graph of the map F is

$$\mathfrak{f} = \big\{\big(X, F(X)\big) \bigm| X \in \mathfrak{g}\big\} \subset \mathfrak{g} \times \mathfrak{h}$$

and it is a Lie subalgebra of $\mathfrak{g} \times \mathfrak{h}$, the Lie algebra of $G \times H$. By Theorem 12.2.6 there is a Lie subgroup $K \subset G \times H$ with Lie algebra $\mathfrak{f}$.

The projections $\pi_1: K \to G$ and $\pi_2: K \to H$ are Lie group homomorphisms. The differential of π_1 at $(e, e) \in K$ is invertible (it is $(X, F(X)) \mapsto X$) so π_1 is a local diffeomorphism at (e, e). Thus we can define on some open neighbourhood U of $e \in G$ the local homomorphism

$$f: U \to H, \quad f = \pi_2 \circ \pi_1^{-1}.$$

Its differential is clearly F. $\square$

With similar techniques we obtain also a uniqueness result.

Proposition 12.2.10. *Let G, H be Lie groups. If G is connected, two homomorphisms $f, f': G \to H$ with the same differentials $f_* = f'_*$ must coincide.*

Proof. Following the previous proof, the graphs of f and f' are two connected Lie subgroups $K, K' \subset G \times H$ with the same Lie subalgebra $\mathfrak{f}$, and hence must coincide, that is $f = f'$. $\square$

If G is simply connected, existence is also achieved.

Proposition 12.2.11. *Let G, H be Lie groups. If G is simply connected, every Lie algebra homomorphism $\mathfrak{g} \to \mathfrak{h}$ is the differential of a unique Lie group homomorphism $G \to H$.*

Proof. In the proof of Theorem 12.2.9, the map $\pi_1: K \to G$ is a smooth covering by Proposition 12.1.13. Being G simply connected, the map π_1 is an isomorphism, so we can define $f = \pi_2 \circ \pi_1^{-1}: G \to H$ and conclude. $\square$

12.2.10. Simply connected Lie groups. The results just stated have the following important consequence.

Corollary 12.2.12. *Two simply connected Lie groups are isomorphic $\Longleftrightarrow$ their Lie algebras are.*

Proof. Every isomorphism $\mathfrak{g} \to \mathfrak{h}$ gives rise to two homomorphisms $G \to H$ and $H \to G$, whose composition is the identity because its differential is. $\square$

Remember that Corollary 12.1.17 reduces the problem of classifying connected Lie groups to the simply connected ones. Now Corollary 12.2.12 in turn translates this task into the purely algebraic problem of classifying all the Lie algebras (to be precise, only the Lie algebras that arise from some Lie groups are important for us).

Two Lie groups G, H are *locally isomorphic* if there are neighbourhoods U and V of $e \in G$ and $e \in H$ and a diffeomorphism $f: U \to V$ such that $f(ab) = f(a)f(b)$ whenever $a, b, ab \in U$.

Corollary 12.2.13. *Let G, H be two connected Lie groups. The following are equivalent:*

- *G and H are locally isomorphic;*
- *G and H have isomorphic universal covers;*
- *$\mathfrak{g}$ and $\mathfrak{h}$ are isomorphic Lie algebras.*

12.2.11. Abelian Lie groups. We now apply the techniques just introduced to classify all the abelian Lie groups. We will need the following.

Proposition 12.2.14. *The differentials of the multiplication $m: G \times G \to G$ and the inverse $i: G \to G$ are*

$$\mathfrak{g} \times \mathfrak{g} \longrightarrow \mathfrak{g}, \quad (X, Y) \longmapsto X + Y, \qquad \mathfrak{g} \longrightarrow \mathfrak{g}, \quad X \longmapsto -X.$$

Proof. For the first, by linearity it suffices to prove that $(X, 0) \mapsto X$, which is obvious since $ge = g$. The second follows from $m(g, i(g)) = g$. □

Here is a smart application.

Proposition 12.2.15. *If a Lie group G is abelian, then $\mathfrak{g}$ also is.*

Proof. Since G is abelian, the map $G \to G$, $g \mapsto g^{-1}$ is an endomorphism. Therefore its derivative $\mathfrak{g} \to \mathfrak{g}$, $X \mapsto -X$ is a Lie algebra endomorphism. Hence for every $X, Y \in \mathfrak{g}$ we get

$$-[X, Y] = [-X, -Y] = [X, Y]$$

which implies $[X, Y] = 0$. □

Recall that in every dimension n there is a unique abelian Lie algebra $\mathbb{R}^n$. We will also need the following.

Exercise 12.2.16. Let $\Gamma < \mathbb{R}^n$ be a discrete subgroup. There is a basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ where $v_1, \ldots, v_k$ generate Γ. In particular $\Gamma \cong \mathbb{Z}^k$.

Here is a complete classification of abelian Lie groups.

Theorem 12.2.17. *Every abelian Lie group is isomorphic to*

$$\underbrace{S^1 \times \cdots \times S^1}_{k} \times \mathbb{R}^{n-k}$$

for some $0 \leq k \leq n$.

Proof. By Proposition 12.2.15 the Lie algebra of an abelian group G is $\mathbb{R}^n$, which is also the Lie algebra of the Lie group $\mathbb{R}^n$. By Corollary 12.2.12 then $\tilde{G} = \mathbb{R}^n$, and by Corollary 12.1.17 we have $G = \mathbb{R}^n/\Gamma$ for some discrete $\Gamma < \mathbb{R}^n$. Now Exercise 12.2.16 applies. □

12.3. Examples

Having proved a number of general theorems, it is due time to exhibit and study more examples of Lie groups.

12.3.1. Complex matrices. We introduce some Lie groups using complex matrices. To this purpose we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ in the usual way, by sending $(z_1, \ldots, z_n)$ to $(\Re z_1, \Im z_1, \ldots, \Re z_n, \Im z_n)$. We consider every complex endomorphism of $\mathbb{C}^n$ as a particular real endomorphism of $\mathbb{R}^{2n}$ and thus see $M(n, \mathbb{C})$ as a linear subspace of $M(2n, \mathbb{R})$, and more than that as a subalgebra with respect to matrix multiplication.

Our first example is the *complex general linear group*

$$\mathrm{GL}(n, \mathbb{C}) = \{A \in M(n, \mathbb{C}) \mid \det A \neq 0\}.$$

This is an open subset of $M(n, \mathbb{C})$ and hence a Lie group of dimension $2n^2$. It is a Lie subgroup of $\mathrm{GL}(2n, \mathbb{R})$, with Lie algebra

$$\mathfrak{gl}(n, \mathbb{C}) = M(n, \mathbb{C})$$

where we see $M(n, \mathbb{C})$ as a Lie subalgebra of $M(2n, \mathbb{R})$, with the same Lie bracket $[A, B] = AB - BA$. Note the Lie subgroup inclusions:

$$\mathrm{GL}(n, \mathbb{R}) \subset \mathrm{GL}(n, \mathbb{C}) \subset \mathrm{GL}(2n, \mathbb{R}).$$

These Lie groups have dimensions n^2, $2n^2$, and $4n^2$ respectively. When $n = 1$ these reduce to

$$\mathbb{R}^* \subset \mathbb{C}^* \subset \mathrm{GL}(2, \mathbb{R}).$$

In the second inclusion, every element $\rho e^{i\theta} \in \mathbb{C}^*$ is interpreted as the product of a ρ-dilation with a θ-rotation:

$$\rho \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The determinant is a Lie group homomorphism $\det\colon \mathrm{GL}(n, \mathbb{C}) \to \mathbb{C}^*$. As in the real case, the *complex special linear group* is its kernel

$$\mathrm{SL}(n, \mathbb{C}) = \{A \in \mathrm{GL}(n, \mathbb{C}) \mid \det A = 1\}.$$

This is a Lie subgroup, with Lie algebra

$$\mathfrak{sl}(n, \mathbb{C}) = \{A \in M(n, \mathbb{C}) \mid \mathrm{tr} A = 0\}.$$

The Lie group $\mathrm{GL}(n, \mathbb{C})$ contains the *unitary group* $\mathrm{U}(n)$, that consists of all unitary matrices:

$$\mathrm{U}(n) = \{A \in \mathrm{GL}(n, \mathbb{C}) \mid {}^t\bar{A}A = I\}.$$

Exercise 12.3.1. The unitary group is a Lie subgroup of $\mathrm{GL}(n, \mathbb{C})$ of dimension n^2, whose Lie algebra consists of all the $n \times n$ skew-Hermitian matrices:

$$\mathfrak{u}(n) = \{A \in M(n, \mathbb{C}) \mid {}^t\bar{A} + A = 0\}.$$

Hint. Adapt the proof of Proposition 3.9.2 to the complex case. □

Finally, the *special unitary group* is

$$\mathrm{SU}(n) = \left\{ A \in \mathrm{GL}(n, \mathbb{C}) \;\middle|\; {}^{\mathrm{t}}\bar{A}A = I, \det A = 1 \right\}.$$

Exercise 12.3.2. This is a Lie subgroup of dimension $n^2 - 1$ with Lie algebra

$$\mathfrak{su}(n) = \left\{ A \in M(n, \mathbb{C}) \;\middle|\; {}^{\mathrm{t}}\bar{A} + A = 0, \mathrm{tr} A = 0 \right\}.$$

We note that

$$\mathrm{SU}(n) = \mathrm{U}(n) \cap \mathrm{SL}(n, \mathbb{C}).$$

Exercise 12.3.3. The Lie groups $\mathrm{GL}(n, \mathbb{C})$, $\mathrm{SL}(n, \mathbb{C})$, $\mathrm{U}(n)$, and $\mathrm{SU}(n)$ are all connected.

12.3.2. More matrix Lie groups. We further introduce some Lie subgroups of $\mathrm{GL}(n, \mathbb{R})$ that are widely used in geometry.

Example 12.3.4 (Indefinite orthogonal groups). Remember that

$$\mathrm{O}(p, q) = \left\{ A \in \mathrm{GL}(n, \mathbb{R}) \;\middle|\; {}^{\mathrm{t}}A I_{p,q} A = I_{p,q} \right\}$$

where $I_{p,q} = \left(\begin{smallmatrix} I_p & 0 \\ 0 & -I_q \end{smallmatrix}\right)$. Similarly to the proof of Proposition 3.9.2, we check easily that the group of matrices $\mathrm{O}(p, q)$ is indeed a submanifold of $\mathrm{GL}(n, \mathbb{R})$ of dimension $\frac{n(n-1)}{2}$ with Lie algebra

$$\mathfrak{o}(p, q) = \left\{ A \in M(n) \;\middle|\; {}^{\mathrm{t}}A I_{p,q} + I_{p,q} A = 0 \right\}.$$

Every matrix in $\mathrm{O}(p, q)$ has determinant ± 1, and $\mathrm{SO}(p, q)$ is the index-two subgroup consisting of those with determinant 1. We have $\mathfrak{so}(p, q) = \mathfrak{o}(p, q)$.

The Lie groups $\mathrm{O}(p, q)$ and $\mathrm{O}(q, p)$ are isomorphic. If $p, q > 0$, the Lie group $\mathrm{O}(p, q)$ is not compact (exercise).

Example 12.3.5 (Indefinite unitary groups). Proceeding exactly as above with the standard hermitian product of signature (p, q) on $\mathbb{C}^{p+q}$, we construct the Lie groups

$$\mathrm{U}(p, q) = \left\{ A \in \mathrm{GL}(n, \mathbb{C}) \;\middle|\; {}^{\mathrm{t}}\bar{A} I_{p,q} A = I_{p,q} \right\}$$

with Lie algebra

$$\mathfrak{u}(p, q) = \left\{ A \in M(n, \mathbb{C}) \;\middle|\; {}^{\mathrm{t}}\bar{A} I_{p,q} + I_{p,q} A = 0 \right\}.$$

The matrices of $\mathrm{U}(p, q)$ with unit determinant form a Lie subgroup $\mathrm{SU}(p, q)$, with Lie algebra

$$\mathfrak{su}(p, q) = \left\{ A \in M(n, \mathbb{C}) \;\middle|\; {}^{\mathrm{t}}\bar{A} I_{p,q} + I_{p,q} A = 0, \mathrm{tr} A = 0 \right\}.$$

We have $\dim \mathrm{U}(p, q) = n^2$ and $\dim \mathrm{SU}(p, q) = n^2 - 1$, with $n = p + q$.

Example 12.3.6 (Symplectic groups). Let $\mathbb{R}^{2n}$ or $\mathbb{C}^{2n}$ be equipped with the standard symplectic (that is, antisymmetric and non-degenerate) form

$$\omega(x, y) = {}^{\mathrm{t}}x J y$$

where

$$J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}.$$

Let $\mathrm{Sp}(2n, \mathbb{R})$ or $\mathrm{Sp}(2n, \mathbb{C})$ be group of all linear isomorphism preserving the symplectic form. That is,

$$\mathrm{Sp}(2n, \mathbb{R}) = \left\{ A \in \mathrm{GL}(n, \mathbb{R}) \;\middle|\; {}^{\mathrm{t}}AJA = J \right\}.$$

The Lie algebra is

$$\mathfrak{sp}(2n, \mathbb{R}) = \left\{ A \in M(n) \;\middle|\; {}^{\mathrm{t}}AJ + JA = 0 \right\}.$$

The complex case is analogous. The dimensions of $\mathrm{Sp}(2n, \mathbb{R})$ and $\mathrm{Sp}(2n, \mathbb{C})$ are $n(2n+1)$ and $2n(2n+1)$ respectively.

Example 12.3.7 (Affine extensions). For every Lie subgroup $G < \mathrm{GL}(n, \mathbb{R})$ we may consider its affine extension

$$G \rtimes \mathbb{R}^n = \left\{ x \mapsto Ax + b \;\middle|\; A \in G, b \in \mathbb{R}^n \right\} \subset \mathrm{Aff}(\mathbb{R}^n).$$

This is a Lie subgroup of $\mathrm{Aff}(\mathbb{R}^n)$, which is in turn a Lie subgroup of $\mathrm{GL}(n+1, \mathbb{R})$, recall Example 12.1.7. Its Lie algebra is the subalgebra of $\mathfrak{gl}(n+1, \mathbb{R})$ consisting of all matrices

$$\begin{pmatrix} A & b \\ 0 & 0 \end{pmatrix}$$

where $A \in \mathfrak{g}$ and $b \in \mathbb{R}^n$.

12.3.3. Low dimensions. We now try to embark on a more systematic classification of connected Lie groups with increasing dimension. We use the powerful *Lie groups – Lie algebra correspondence* proved in the previous pages, which can be reassumed as follows:

(i) Every connected Lie group is the quotient G/Γ of a simply connected Lie group G by a discrete central subgroup $\Gamma < G$.
(ii) Every simply connected Lie group G is totally determined by its Lie algebra $\mathfrak{g}$.

An optimistic strategy to produce all connected Lie groups would be the following:

(1) Classify all Lie algebras $\mathfrak{g}$.
(2) Try to build a simply connected Lie group G for each Lie algebra $\mathfrak{g}$.
(3) Quotient G by its central discrete subgroups.

Dimension one. The only one-dimensional Lie algebra is the abelian $\mathbb{R}$, so the 1-dimensional connected Lie groups are $\mathbb{R}$ and S^1.

Dimension two. In dimension two, we find two Lie algebras:

- The abelian $\mathbb{R}^2$.
- The Lie algebra $\mathfrak{aff}(\mathbb{R})$ of $\mathrm{Aff}(\mathbb{R})$.

Proposition 12.3.8. *These are the only two 2-dimensional Lie algebras up to isomorphism.*

Proof. Let $\mathfrak{a}$ be a 2-dimensional Lie algebra. Pick a basis $X, Y \in \mathfrak{a}$ and note that the whole structure is determined by the element $[X, Y]$. If $[X, Y] = 0$ then $\mathfrak{a}$ is abelian. Otherwise, after changing the basis we easily reduce to the case $[X, Y] = Y$ and we get $\mathfrak{aff}(\mathbb{R})$. Indeed, We see $\mathrm{Aff}(\mathbb{R}) \subset \mathrm{GL}(2, \mathbb{R})$ as the set of all matrices

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$$

with $a, b \in \mathbb{R}$. Its Lie algebra is generated by the matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We have $[A, B] = B$, so $\mathfrak{aff}(\mathbb{R}) \cong \mathfrak{a}$. □

The simply connected Lie group with algebra $\mathfrak{aff}(\mathbb{R})$ is $\mathrm{Aff}(\mathbb{R})^0$. We can easily classify the two-dimensional connected Lie groups up to isomorphism:

Proposition 12.3.9. *The two-dimensional connected Lie groups are*

$$\mathbb{R}^2, \quad S^1 \times \mathbb{R}, \quad S^1 \times S^1, \quad \mathrm{Aff}(\mathbb{R})^0.$$

Proof. Since the centre of $\mathrm{Aff}(\mathbb{R})^0$ is trivial, there is no other connected Lie group with Lie algebra $\mathfrak{aff}(\mathbb{R})$ except $\mathrm{Aff}(\mathbb{R})^0$ itself. □

Dimension three. In dimension three we find many more Lie algebras. Here are some:

(1) The abelian $\mathbb{R}^3$.
(2) The *Heisenberg algebra*, which is the subalgebra of $\mathfrak{sl}(3, \mathbb{R})$ formed by the matrices
$$\begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}$$
with $a, b, c \in \mathbb{R}$. This is the Lie algebra of the Heisenberg group.
(3) The direct product $\mathbb{R} \oplus \mathfrak{aff}(\mathbb{R})$.
(4) The Lie algebra of the affine isometries of $\mathbb{R}^2$.
(5) The Lie algebra of the affine isometries of $\mathbb{R}^{1,1}$.
(6) The Lie algebra $\mathfrak{sl}(2, \mathbb{R})$.
(7) The Lie algebra $\mathfrak{so}(3)$.

Each of these seven algebras is the Lie algebra of some Lie group. Unfortunately, this is not the end of the story: the are uncountably many Lie algebras in dimension three, as the following exercise shows.

to be checked

Exercise 12.3.10. Consider $\mathbb{R}^3$ with basis X, Y, T and Lie bracket defined by

$$[T, X] = X, \qquad [T, Y] = tY, \qquad [X, Y] = 0.$$

This defines a Lie algebra $\mathfrak{g}_t$ for all $t \in \mathbb{R}$. If $tu \neq 1$ then $\mathfrak{g}_t$ and $\mathfrak{g}_u$ are not isomorphic. Every $\mathfrak{g}_t$ is a subalgebra of $\mathfrak{gl}(n, \mathbb{R})$ for some n and is hence the Lie algebra of some Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$.

It is actually possible to classify all the three-dimensional Lie algebras: this was done by Bianchi in 1898 who subdivided them into 11 classes, two of which are continuous families. However, these examples already suggest that it is practically impossible to classify all connected Lie groups without adding further assumptions like, for instance, that the Lie group should be compact, or abelian, or some weaker assumption.

We now write some isomorphisms between some notable three-dimensional Lie algebras. Let $\times$ be the cross product of vectors in $\mathbb{R}^3$.

Proposition 12.3.11. *The Lie algebras* $\mathfrak{so}(3)$ *and* $\mathfrak{su}(2)$ *are both isomorphic to the algebra* $(\mathbb{R}^3, \times)$.

Proof. A basis for $\mathfrak{so}(3)$ is given by the matrices

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

We have

$$[A, B] = C, \qquad [B, C] = A, \qquad [C, A] = B.$$

Therefore $\mathfrak{so}(3) \cong (\mathbb{R}^3, \times)$. Analogously $\mathfrak{su}(2)$ is generated by the matrices

$$A = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad B = \frac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \qquad C = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

whose Lie brackets are again as above. □

This implies that SO(3) and SU(2) have the same universal cover. In fact, we will write an explicit double cover $\mathrm{SU}(2) \to \mathrm{SO}(3)$ soon.

Proposition 12.3.12. *The Lie algebras* $\mathfrak{sl}(2, \mathbb{R})$ *and* $\mathfrak{so}(2, 1)$ *are isomorphic.*

Proof. A basis for $\mathfrak{sl}(2, \mathbb{R})$ is

$$A = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad B = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad C = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

We have

$$[A, B] = C, \qquad [B, C] = -A, \qquad [C, A] = -B.$$

The Lie algebra $\mathfrak{so}(2, 1)$ consists of matrices of the form

$$\begin{pmatrix} M & b \\ {}^{t}b & 0 \end{pmatrix}$$

with ${}^{t}M + M = 0$. A basis is

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

Their Lie brackets are as above. □

The *derived algebra* $[\mathfrak{g}, \mathfrak{g}]$ of a Lie algebra $\mathfrak{g}$ is the subalgebra generated by all the brackets $[X, Y]$ as $X, Y \in \mathfrak{g}$ varies. The derived algebra is trivial $\Longleftrightarrow$ $\mathfrak{g}$ is abelian.

Exercise 12.3.13. In the seven Lie algebras listed above, the dimension of $[\mathfrak{g}, \mathfrak{g}]$ is zero for (1), one for (2, 3), two for (4,5), and three for (6,7).

12.4. The exponential map

Similar to Riemannian manifolds, Lie groups G are equipped with an *exponential map* $\mathfrak{g} \to G$. For matrix groups, this is the usual matrix exponential, and this finally explains the reason for adopting this name...

12.4.1. Definition. Let G be a Lie group. Pick an arbitrary left-invariant vector field $X \in \mathfrak{g}$.

Proposition 12.4.1. *The vector field X is complete.*

Proof. Let $\gamma_g : I_g \to G$ be the maximal integral curve of X at g. Since X is left-invariant, we have $\gamma_g = L_g \circ \gamma_e$ and $I_g = I_e$ for all $g \in G$. By Lemma 5.2.4 the vector field is complete. □

Being complete, the vector field $X \in \mathfrak{g}$ induces a flow $\Phi_X : G \times \mathbb{R} \to G$.

Definition 12.4.2. The *exponential map* $\exp : \mathfrak{g} \longrightarrow G$ is

$$\exp(X) = \Phi_X(e, 1).$$

The map exp is smooth because $\Phi_X(e, 1)$ depends smoothly on the initial values X of the system.

12.4.2. One-parameter subgroups. In the Riemannian case, the restrictions of the exponential map to the vector lines are geodesics; here, these are "one-parameter subgroups."

Let G be a Lie group. For every $X \in \mathfrak{g}$ we consider the curve $\gamma_X : \mathbb{R} \to G$,

$$\gamma_X(t) = \exp(tX).$$

As in the Riemannian case, by construction we have $\gamma_{\lambda X}(t) = \gamma_X(\lambda t)$.

Proposition 12.4.3. *The map $\gamma_X : \mathbb{R} \to G$ is the integral curve of the left-invariant field X with $\gamma_X(0) = e$. It is a Lie group homomorphism.*

Proof. We have

$$\gamma_X(t) = \exp(tX) = \Phi_{tX}(e, 1) = \Phi_X(e, t)$$

so γ_X is the integral curve for X with $\gamma_X(0) = e$. Since X is left-invariant,

$$\gamma_X(s)\gamma_X(t) = L_{\gamma_X(s)}(\gamma_X(t)) = \gamma_X(s + t).$$

Therefore γ_X is a Lie groups homomorphism. $\square$

A Lie group homomorphism $\mathbb{R} \to G$ is called a *one-parameter subgroup* of G. It turns out that every one-parameter subgroup arises in this way.

Proposition 12.4.4. *Every one-parameter subgroup of G is a γ_X for some element $X \in \mathfrak{g}$.*

Proof. Given $f : \mathbb{R} \to G$, we set $X = f_*(1)$. Since $f_* = (\gamma_X)_*$, we have $f = \gamma_X$ by Proposition 12.2.10. $\square$

The Lie algebra $\mathfrak{g}$ thus parametrises all the one-parameter subgroups in G.

12.4.3. Properties. We now list some properties of the exponential map.

Proposition 12.4.5. *Let G be a Lie group. The following hold.*

- *The differential $d\exp_0 : \mathfrak{g} \to \mathfrak{g}$ is the identity. Hence the exponential map is a local diffeomorphism at 0.*
- *If $f : G \to H$ is a Lie group homomorphism, the following diagram commutes:*

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "f_*"] \arrow[d, "\exp"'] & \mathfrak{h} \arrow[d, "\exp"] \\ G \arrow[r, "f"'] & H \end{tikzcd}$$

Proof. Everything follows readily if we interpret $\mathfrak{g}$ and $\mathfrak{h}$ as sets of one-parameter subgroups. $\square$

In particular, if $H \subset G$ is a subgroup, the exponential map $\mathfrak{h} \to H$ is just the restriction of the exponential map $\mathfrak{g} \to G$.

12.4.4. Matrix exponential. We finally motivate the use of the term "exponential map". Recall that the exponential of a square matrix A is

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

If A and B commute, then $e^{A+B} = e^A e^B = e^B e^A$. In particular e^A is invertible with inverse e^{-A}.

Proposition 12.4.6. *The exponential map* $\exp : \mathfrak{gl}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R})$ *is*

$$\exp(A) = e^A.$$

Proof. For every $A \in \mathfrak{gl}(n,\mathbb{R})$ consider the curve $\alpha\colon \mathbb{R} \to \mathrm{GL}(n,\mathbb{R})$, $\alpha(t) = e^{tA}$. We can differentiate it and find $\alpha'(t) = Ae^{tA}$. So α is a smooth curve and in fact a one-parameter subgroup of $\mathrm{GL}(n,\mathbb{R})$. By Proposition 12.4.4 we have $\alpha = \gamma_{\alpha'(0)} = \gamma_A$. In particular $e^A = \exp(A)$. □

By restriction, the same exponential map works for all the Lie subgroups of $\mathrm{GL}(n,\mathbb{R})$ like $\mathrm{SL}(n,\mathbb{R})$ or $\mathrm{O}(n)$. We discover in particular that the exponential of an antisymmetric matrix is orthogonal, and that of a traceless matrix has determinant one; these facts follow also from the following exercise.

Exercise 12.4.7. We have $e^{{}^tA} = {}^t(e^A)$ and $\det e^A = e^{\operatorname{tr} A}$.

From these examples we discover that, as in the Riemannian case, the exponential map needs not to be surjective, not even if G is connected.

Proposition 12.4.8. *The exponential map* $\mathfrak{sl}(2,\mathbb{R}) \to \mathrm{SL}(2,\mathbb{R})$ *is not surjective.*

Proof. If $g = \exp(A)$, it has a square root $\sqrt{g} = \exp\left(\frac{A}{2}\right)$. However

$$B = \begin{pmatrix} -4 & 0 \\ 0 & -\frac{1}{4} \end{pmatrix}$$

has no square root (exercise: use Jordan normal form). □

12.4.5. Applications. In the rest of this section we will use the exponential map to prove these remarkable non-trivial facts. Let G be a Lie group. Then:

(1) Every closed subgroup $H < G$ is a Lie subgroup.
(2) If $H \triangleleft G$ is closed and normal, the quotient G/H is a Lie group.
(3) The kernel and the image of any homomorphism $G \to H$ of Lie groups are Lie subgroups of G and H.

12.4.6. The closed subgroup theorem. As promised, we start by proving the following powerful theorem, which transforms a purely topological condition (closeness) into a much stronger differential one (being a smooth embedded submanifold).

Theorem 12.4.9. *Let G be a Lie group. Every closed subgroup $H \subset G$ is an embedded Lie subgroup.*

To prove this theorem we need a lemma. Recall that $\exp(X+Y) \neq \exp(X)\exp(Y)$ in general.

Lemma 12.4.10. *Let G be a Lie group. For every $X, Y \in \mathfrak{g}$ we have*

$$\exp(X+Y) = \lim_{n\to\infty}\left(\exp\frac{X}{n}\exp\frac{Y}{n}\right)^n.$$

Proof. When t is sufficiently small we have

$$\exp(tX)\exp(tY) = \exp\big(\psi(t)\big)$$

where ψ is the smooth map

$$\psi\colon \mathbb{R} \xrightarrow{\gamma_X \times \gamma_Y} G \times G \xrightarrow{m} G \xrightarrow{\exp^{-1}} \mathfrak{g}.$$

Here m is the multiplication and $\exp^{-1}$ is defined only in a neighbourhood of e. The map ψ is defined only near 0 and $\psi'(0) = X + Y$. Therefore we have

$$\psi(t) = t(X+Y) + t^2 Z(t)$$

for some smooth map Z defined only near 0. This implies

$$\exp(tX)\exp(tY) = \exp\big(\psi(t)\big) = \exp\big(t(X+Y) + t^2 Z(t)\big).$$

If n is sufficiently big, we deduce that

$$\begin{aligned}\left(\exp\frac{X}{n}\exp\frac{Y}{n}\right)^n &= \left(\exp\left(\frac{1}{n}(X+Y) + \frac{1}{n^2}Z\left(\frac{1}{n}\right)\right)\right)^n \\ &= \exp\left(X + Y + \frac{1}{n}Z\left(\frac{1}{n}\right)\right).\end{aligned}$$

This completes the proof. □

We can now turn back to the proof of Theorem 12.4.9

Proof. We must prove that $H \subset G$ is an embedded submanifold. Let $\mathfrak{h} \subset \mathfrak{g}$ be the subset defined as

$$\mathfrak{h} = \big\{X \in \mathfrak{g} \ \big|\ \exp(tX) \in H \ \forall t \in \mathbb{R}\big\}.$$

We first prove that $\mathfrak{h}$ is a subspace of $\mathfrak{g}$. To do so, we pick $X, Y \in \mathfrak{h}$, and prove that $X + Y \in \mathfrak{h}$. We know that $\exp\frac{tX}{n}, \exp\frac{tY}{n} \in H$, hence $\big(\exp\frac{tX}{n}\exp\frac{tY}{n}\big)^n \in H$. Since H is closed, by the previous lemma we get $\exp(t(X+Y)) \in H$ for every $t \in \mathbb{R}$ and therefore $X + Y \in \mathfrak{h}$.

We now construct neighbourhoods U and W of $0 \in \mathfrak{g}$ and $e \in G$ such that $\exp|_U\colon U \to W$ is a diffeomorphism and

$$\exp(\mathfrak{h} \cap U) = H \cap W. \tag{50}$$

This shows that H is an embedded submanifold near e, and hence everywhere by left multiplication.

Let $\mathfrak{k} \subset \mathfrak{g}$ be a complementary subspace for $\mathfrak{h}$. We leave as an exercise to prove that there is an open neighbourhood $U_\mathfrak{k}$ of $0 \in \mathfrak{k}$ such that

$$H \cap \exp\big(U_\mathfrak{k} \setminus \{0\}\big) = \varnothing. \tag{51}$$

Instead of the exponential map, it is now convenient to consider the map

$$f\colon \mathfrak{h} \times \mathfrak{k} \longrightarrow \mathfrak{g}, \qquad f(X,Y) = \exp(X)\exp(Y).$$

We still have $df_0 = \mathrm{id}$, so there are neighbourhoods $U_\mathfrak{h}, U_\mathfrak{k}$ of $0 \in \mathfrak{h}, \mathfrak{k}$ such that

$$f\colon U_\mathfrak{h} \times U_\mathfrak{k} \longrightarrow G$$

is a diffeomorphism onto its image. We suppose that $U_{\mathfrak{f}}$ also satisfies (51). We now set $U = U_{\mathfrak{h}} \times U_{\mathfrak{f}}$ and prove that

$$f(\mathfrak{h} \cap U) = H \cap f(U). \tag{52}$$

We have $\mathfrak{h} \cap U = U_{\mathfrak{h}}$ and $\exp(U_{\mathfrak{h}}) \subset H$, therefore $f(\mathfrak{h} \cap U) \subset H \cap f(U)$. On the other hand, if $h \in H \cap f(U)$ then $h = \exp(X)\exp(Y)$ with $X \in U_{\mathfrak{h}}$ and $Y \in U_{\mathfrak{f}}$. Now $h, \exp(X) \in H$ implies that $\exp(Y) \in H$ and hence by (51) we get $Y = 0$. Therefore $h \in \exp(U_{\mathfrak{h}})$.

We have proved (52), which in turn implies (50) by taking $W = \exp(U)$. This concludes the proof. □

By combining the theorem with Corollary 12.2.8 we get

Corollary 12.4.11. *Let G be a Lie group. A subgroup $H < G$ is an embedded Lie subgroup $\Longleftrightarrow$ it is closed.*

12.4.7. Kernel. Here is an immediate application of the closed subgroup theorem.

Proposition 12.4.12. *Let $f\colon G \to H$ be a homomorphism of Lie groups. The kernel* $\ker f$ *is an embedded Lie subgroup of G.*

Proof. It is closed since f is continuous. Theorem 12.4.9 applies. □

We want to prove an analogous theorem for the image. It is more convenient to first study the quotients of Lie groups.

12.4.8. Quotient of Lie groups. We now recycle the proof of the closed subgroup theorem to obtain the following.

Theorem 12.4.13. *Let G be a Lie group and $H < G$ a closed subgroup. The quotient G/H has a natural structure of smooth manifold such that $\pi\colon G \to G/H$ is a fibre bundle.*

Proof. We know that G is foliated into the the cosets of H. Since H is closed, it is embedded, and hence its cosets also are. We now need to show that the cosets fit like fibers in a bundle.

As in the proof of Theorem 12.4.9 we pick a complementary subspace $\mathfrak{f}$ for $\mathfrak{h} \subset \mathfrak{g}$ and consider the map

$$f\colon \mathfrak{f} \times \mathfrak{h} \longrightarrow \mathfrak{g}, \qquad f(X, Y) = \exp(X)\exp(Y).$$

Let $U_{\mathfrak{f}}, U_{\mathfrak{h}}$ be neighbourhoods of $0 \in \mathfrak{f}, \mathfrak{h}$ such that

$$f\colon U_{\mathfrak{f}} \times U_{\mathfrak{h}} \longrightarrow G$$

is a diffeomorphism onto its image and $\operatorname{Im} f \cap H = f(0 \times U_{\mathfrak{h}}) = \exp(U_{\mathfrak{h}})$. We now pick a smaller neighbourhood $U'_{\mathfrak{f}} \subset U_{\mathfrak{f}}$ such that $u_1, u_2 \in U'_{\mathfrak{f}} \Rightarrow u_1 - u_2 \in U_{\mathfrak{f}}$. This implies that

$$\exp(U'_{\mathfrak{f}})\big(\exp(U'_{\mathfrak{f}})\big)^{-1} \subset \exp(U_{\mathfrak{f}}).$$

We consider the multiplication map

$$m\colon \exp(U'_{\mathfrak{f}}) \times H \longrightarrow G, \qquad m(g,h) = gh.$$

The map m is injective: if $g_1h_1 = g_2h_2$, then $g_2g_1^{-1} = h_2^{-1}h_1 \in H$, but since $g_2g_1^{-1} \in \exp(U_{\mathfrak{f}})$ we deduce that $g_2g_1^{-1} = e$, so $g_1 = g_2$ and $h_1 = h_2$.

The map m is an open embedding, after replacing $U'_{\mathfrak{f}}$ with a smaller open neighbourhood: we have $dm_{(e,e)} = \mathrm{id}$, so $dm_{(g,e)}$ is invertible for every $g \in \exp(U'_{\mathfrak{f}})$ up to taking a smaller $U'_{\mathfrak{f}}$. Hence $dm_{(g,h)}$ is invertible by right-multiplication for every $h \in H$.

Finally, we assign to G/H its quotient topology. The map

$$U'_{\mathfrak{f}} \longrightarrow G/H, \qquad X \longmapsto \exp(X)H$$

is a homeomorphism onto its image. More generally, for every $g \in G$ the map $U'_{\mathfrak{f}} \to G/H, X \mapsto g\exp(X)H$ is a homeomorphism onto its image and we use these maps as charts to give G/H a smooth structure.

The space G/H is now a smooth manifold and the map $G \to G/H$ is a fibre bundle, with fibre diffeomorphic to H. □

When H is a normal subgroup, things of course improve.

Corollary 12.4.14. *Let G be a Lie group and $H \triangleleft G$ a closed normal subgroup. The quotient G/H has a natural structure of Lie group, and $G \to G/H$ is a Lie group homomorphism.*

12.4.9. Image. After taking care of kernels and quotients, we can finally consider images of Lie group homomorphisms. It is remarkable how many non-trivial theorems are necessary to prove this reasonable-looking fact.

Proposition 12.4.15. *Let $f: G \to H$ be a homomorphism of Lie groups. The image* $\mathrm{Im}\, f$ *is a Lie subgroup of H.*

Proof. Since $\ker f$ is closed and normal, the quotient $G/\ker f$ is a Lie group. The induced map $G/\ker f \to H$ is an injective immersion: hence its image is an injectively immersed manifold and a subgroup of H, that is a Lie subgroup. □

The image is of course not guaranteed to be embedded.

Remark 12.4.16. The use of the term *one-parameter subgroup* in Section 12.4.2 for any Lie group homomorphism $\mathbb{R} \to G$ is now fully legitimated, since its image is indeed a Lie subgroup of G.

12.5. Lie group actions

Lie groups arise often as symmetry groups, and are more generally designed to act on spaces of various kind.

12.5.1. Definition. Let M be a smooth manifold and G a Lie group. A *Lie group action* of G on M is a homomorphism

$$G \longrightarrow \mathrm{Diffeo}(M)$$

that is also smooth in the following sense: the induced map

$$G \times M \longrightarrow M, \qquad (g, x) \longmapsto g(x)$$

should be smooth. A manifold M equipped with a Lie group action of G is sometimes called a *G-manifold*.

Here are some important examples:

- The group $\mathrm{GL}(n, \mathbb{R})$ or $\mathrm{Aff}(\mathbb{R}^n)$ acts on $\mathbb{R}^n$.
- The group $\mathrm{O}(n)$ acts on $S^{n-1} \subset \mathbb{R}^n$.
- The group $\mathrm{U}(n)$ acts on $S^{2n-1} \subset \mathbb{C}^n$.
- Every Lie group G acts on itself by left-multiplication $g(x) = gx$, by right-multiplication $g(x) = xg^{-1}$, and by conjugation $g(x) = gxg^{-1}$.

An action of $\mathbb{R}$ on M was called a *one-parameter group of diffeomorphisms* in Section 5.2.2.

12.5.2. Lie algebras. As usual, Lie algebras are there to help us, by encoding elegantly the infinitesimal side of the story. Let $\rho\colon G \to \mathrm{Diffeo}(M)$ be a Lie group action on M. This induces a homomorphism

$$\rho_*\colon \mathfrak{g} \longrightarrow \mathfrak{X}(M)$$

as follows. For every $p \in M$ we have a map

$$G \longrightarrow M, \qquad g \longmapsto g(p)$$

whose image is the orbit of p. The differential of this map at $e \in G$ is a linear map $\mathfrak{g} \to T_p(M)$. By collecting all these linear maps as $p \in M$ varies we get our homomorphism $\rho_*\colon \mathfrak{g} \to \Gamma(TM) = \mathfrak{X}(M)$.

Exercise 12.5.1. For every $X \in \mathfrak{g}$, the vector field $\rho_*(X)$ on M is complete with flow $\Phi_t\colon M \to M$. We have $\Phi_t(p) = \exp(tX)(p)$ for every $p \in M$.

In some sense $\mathrm{Diffeo}(M)$ is an infinite-dimensional Lie group and $\mathfrak{X}(M)$ is its Lie algebra. A morphism ρ of Lie groups should then induce one ρ_* of Lie algebras: we leave a rigorous proof of this fact as an exercise.

Pare sia in realtà un anti-homomorphism. Controllare

Exercise 12.5.2. The homomorphism ρ_* is a Lie algebra homomorphism.

Exercise 12.5.3. Let ρ be the action of $\mathrm{GL}(n, \mathbb{R})$ on $\mathbb{R}^n$. For every $A \in \mathfrak{gl}(n, \mathbb{R}) = M(n)$ the vector field $\rho_*(A)$ is $\mathbb{R}^n \to \mathbb{R}^n$, $X \mapsto AX$.

12.5.3. Stabilisers and orbits. When dealing with group actions, the first thing to do is always to investigate stabilisers and orbits. Let a Lie group G act on a smooth manifold M.

Proposition 12.5.4. *For every $p \in M$ the stabiliser $G_p < G$ is an embedded Lie subgroup, whose Lie algebra is*

$$\mathfrak{g}_p = \big\{X \in \mathfrak{g} \bigm| \rho_*(X)(p) = 0\big\}.$$

Moreover the induced map

$$G/G_p \longrightarrow M, \qquad g \longmapsto g(p)$$

is an injective immersion, whose image is the orbit of p.

Proof. The stabiliser G_p is closed (exercise), so it is an embedded Lie subgroup. By Exercise 12.5.1 we have $\rho_*(X)(p) = 0$ for every $X \in \mathfrak{g}_p$. Conversely, if $\rho_*(X)(p) = 0$ then $p = \Phi_t(p) = \exp(tX)(p)$ for all t and hence $\exp(tX) \in G_p$ for all t, so $X \in \mathfrak{g}_p$.

The map $G/G_p \to M$ is smooth because $G \to M$ is. Its differential at e is injective because if $X \in \mathfrak{g} \setminus \mathfrak{g}_p$ then $\rho_*(X)(p) \neq 0$. It is hence injective everywhere by left-multiplication. □

We have discovered that stabilisers are Lie subgroups, and orbits are immersed submanifolds. The manifold M is hence partitioned into immersed submanifolds (the orbits) that may have varying dimension.

Example 12.5.5. Let S^1 act on $\mathbb{R}^2$ by rotations. The orbits are the circles centered at the origin, and the origin itself.

Example 12.5.6. Every similarity or congruence class of matrices in the space $M(n)$ of all $n \times n$ real matrices is an immersed submanifold. This holds because each such class is an orbit of the action of $\mathrm{GL}(n, \mathbb{R})$ by conjugation or congruence.

For the same reason, every conjugacy class in a Lie group G is an immersed submanifold.

As usual, one wonders whether injective immersions can be promoted to embeddings. The usual counterexample shows that non-embedded orbits may occur: the action

$$\mathbb{R} \longrightarrow \mathrm{Diffeo}(S^1 \times S^1), \qquad s \mapsto \big((e^{it}, e^{iu}) \mapsto e^{i(t+s)}, e^{i(u+\lambda s)}\big)$$

has dense orbits if $\lambda \notin \mathbb{Q}$. Things improve if an additional hypothesis is fulfilled.

12.5.4. Proper actions. Let G be a Lie group acting on a manifold M

Definition 12.5.7. The action is *proper* if the following map is:

$$G \times M \longrightarrow M \times M, \qquad (g, p) \longmapsto \big(g(p), p\big).$$

If the action is proper, the stabilisers $G_p < G$ are compact for every $p \in M$. The orbits are also nicer.

Proposition 12.5.8. *If the action is proper, orbits are embedded and closed.*

Proof. The induced map $G/G_p \to M$, $g \mapsto g(p)$ is proper. By Exercise 3.8.5 A proper injective immersion is an embedding and has closed image. □

If G is compact, then every action of G is proper.

12.5.5. Homogeneous spaces. Recall that a G-manifold is a manifold M equipped with the action of a Lie group G.

Definition 12.5.9. If the action is transitive, the G-manifold M is called a *homogeneous space*.

Example 12.5.10. Let G be a Lie group and $H < G$ a closed subgroup. The left action of G on G/H is transitive: hence G/H is a homogeneous space.

It turns out that every homogenous space is precisely of this form.

Proposition 12.5.11. *If G acts transtitively on M, for every $p \in M$ the map*

$$G/G_p \longrightarrow M$$

is a G-equivariant diffeomorphism.

Proof. This is a corollary of Proposition 12.5.4. □

In other words, a homogeneous space is just a quotient G/H of a Lie group G by a closed subgroup H. A homogeneous space is one where "all points look the same", since G act transitively on them.

12.5.6. Examples. There are many interesting examples of homogeneous spaces, and we list some here.

Example 12.5.12. The group $\mathrm{SO}(n)$ acts transitively on S^{n-1}, with stabiliser isomorphic to $\mathrm{SO}(n-1)$. Therefore we get the homogeneous space

$$\mathrm{SO}(n)/\mathrm{SO}(n-1) \cong S^{n-1}.$$

By Theorem 12.4.13 we have a fibre bundle $\mathrm{SO}(n) \to S^{n-1}$ with fibre $\mathrm{SO}(n-1)$.

Example 12.5.13. The group $\mathrm{Isom}^+(\mathbb{R}^n)$ of the orientation-preserving Euclidean affine isometries acts transitively on $\mathbb{R}^n$ with stabiliser isomorphic to $\mathrm{SO}(n)$. We get the homogeneous space

$$\mathrm{Isom}^+(\mathbb{R}^n)/\mathrm{SO}(n) \cong \mathbb{R}^n$$

and a fibre bundle $\mathrm{Isom}^+(\mathbb{R}^n) \to \mathbb{R}^n$ with fibre $\mathrm{SO}(n)$.

Example 12.5.14. The group $\mathrm{O}(n)$ acts on the grassmannian $\mathrm{Gr}_k(\mathbb{R}^n)$ with stabiliser isomorphic to $\mathrm{O}(k) \times \mathrm{O}(n-k)$. We get the homogeneous space

$$\mathrm{O}(n)/_{\mathrm{O}(k)\times\mathrm{O}(n-k)} \cong \mathrm{Gr}_k(\mathbb{R}^n)$$

and a fibre bundle $\mathrm{O}(n) \to \mathrm{Gr}_k(\mathbb{R}^n)$ with fibre $\mathrm{O}(k) \times \mathrm{O}(n-k)$.

In fact, we could have used this construction to *define* a natural smooth manifold structure on the grassmannian. We do this with another interesting set. A *flag* on a n-dimensional vector space V is a nested sequence

$$0 \subset V_1 \subset \ldots \subset V_n = V$$

of i-dimensional subspaces $V_i \subset V$. In the following example we build a natural smooth manifold structure on the set of all flags in V.

Example 12.5.15. The group $\mathrm{GL}(n,\mathbb{R})$ acts on the space F of all the flags in $\mathbb{R}^n$. The stabiliser of the coordinate flag $V_i = \mathrm{Span}(e_1, \ldots, e_i)$ is the closed subgroup $H < \mathrm{GL}(n,\mathbb{R})$ of all upper triangular invertible matrices. Therefore the space of all flags in $\mathbb{R}^n$ is naturally identified with the homogeneous manifold $\mathrm{GL}(n,\mathbb{R})/H$.

Exercise 12.5.16. The group $\mathrm{SL}(2,\mathbb{C})$ acts transitively on $\mathbb{P}^1(\mathbb{C})$ as follows:

$$\rho\begin{pmatrix} a & b \\ c & d \end{pmatrix} : [w,z] \longmapsto [aw+bz, cw+dz].$$

The stabiliser is a Lie group diffeomorphic to $\mathbb{C}^* \times \mathbb{C}$. We get a fibre bundle $\mathrm{SL}(2,\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ with fibre $\mathbb{C}^* \times \mathbb{C}$.

12.6. Exercises

Exercise 12.6.1. Show that the exponential map $\exp\colon \mathfrak{so}(n) \to \mathrm{SO}(n)$ is surjective.

Exercise 12.6.2. Show that the only connected Lie groups of $\mathrm{SO}(3)$ are $\{e\}$, $\mathrm{SO}(3)$, and the subgroups isomorphic to S^1 that describe the rotations around some axis.

Index