An Introduction to Riemannian Geometry

with Applications to Mechanics and Relativity

Leonor Godinho and José Natário

Lisbon, 2004

Contents

CHAPTER 1

Differentiable Manifolds

This chapter introduces the basic notions of differential geometry.

The first section studies **topological manifolds** of dimension n, which is the rigorous mathematical concept corresponding to the intuitive notion of "continuous n-dimensional spaces". Several examples are studied, particularly in dimension 2 (surfaces).

Section 2 specializes to **differentiable manifolds**, on which one can define **differentiable functions** (Section 3) and **tangent vectors** (Section 4). Important examples of differentiable maps, namely **immersions** and **embeddings**, are examined in Section 5.

Vector fields and their **flows** are the main topic of Section 6. It is shown that there is a natural differential operation between vector fields, called the **Lie bracket**, which produces a new vector field.

Section 7 is devoted to the important class of differentiable manifolds which are also groups, the so-called **Lie groups**. It is shown that to each Lie group one can associate a **Lie algebra**, i.e. a vector space equipped with a Lie bracket, and the **exponential map**, which maps the Lie algebra to the Lie group.

The notion of **orientability** of a manifold (which generalizes the intuitive notion of "having two sides") is discussed in Section 8.

Finally, **manifolds with boundary** are studied in Section 9.

1. Topological Manifolds

We will begin this section by studying spaces that are locally like $\mathbb{R}^n$, meaning that there exists a neighborhood around each point which is homeomorphic to an open subset of $\mathbb{R}^n$.

DEFINITION 1.1. *A* **topological manifold** *M of dimension n is a topological space with the following properties:*

(i) M is **Hausdorff***, that is, for each pair p_1, p_2 of distinct points of M, there exist neighborhoods V_1, V_2 of p_1 and p_2 such that $V_1 \cap V_2 = \varnothing$.*

(ii) Each point $p \in M$ possesses a neighborhood V homeomorphic to an open subset U of $\mathbb{R}^n$.

(iii) M satisfies the **second countability axiom***, that is, M has a countable basis for its topology.*

Conditions (i) and (iii) are included in the definition to prevent the topology of these spaces from being too strange. In particular, the Hausdorff axiom ensures that the limit of a convergent sequence is unique. This, along with the second countability axiom, guarantees the existence of partitions of unity (cf. Section 7.2 of Chapter 2), which, as we will see, are a fundamental tool in differential geometry.

REMARK 1.2. If the dimension of M is zero, then M is a countable set equipped with the discrete topology (every subset of M is an open set). If $\dim M = 1$, then M is locally homeomorphic to an open interval; if $\dim M = 2$, then it is locally homeomorphic to an open disk, etc.

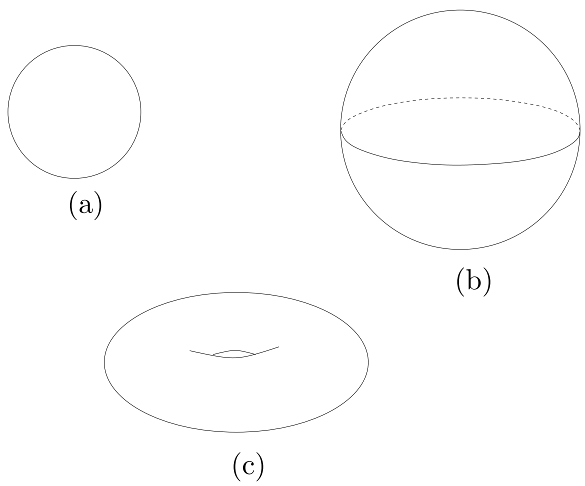

FIGURE 1. (a) S^1, (b) S^2, (c) Torus of revolution.

EXAMPLE 1.3.

(1) Every open subset M of $\mathbb{R}^n$ with the subspace topology (that is, $U \subset M$ is an open set if and only if $U = M \cap V$ with V an open set of $\mathbb{R}^n$) is a topological manifold.

(2) (The **circle** S^1) The circle
$$S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$
with the subspace topology is a topological manifold of dimension 1. Conditions (i) and (iii) are inherited from the ambient space. Moreover, for each point $p \in S^1$ there is at least one coordinate axis which is not parallel to the vector n_p normal to S^1 at p. The projection on this axis is then a homeomorphism between a (sufficiently small) neighborhood V of p and an interval in $\mathbb{R}$.

(3) (The **2-sphere** S^2) The previous example can be easily generalized to show that the **2-sphere**
$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$$

with the subspace topology is a topological manifold of dimension 2.

(4) (The **torus of revolution**) Again as in the previous examples, we can show that the surface of revolution obtained by revolving a circle around an axis that does not intersect it is a topological manifold of dimension 2.

(5) The surface of a cube is a topological manifold (homeomorphic to S^2).

EXAMPLE 1.4. We can also obtain topological manifolds by identifying edges of certain polygons by means of homeomorphisms. The edges of a square, for instance, can be identified in several ways (see Figure 2):

(1) The **torus** T^2 is the quotient of the unit square $Q = [0,1]^2 \subset \mathbb{R}^2$ by the equivalence relation
$$(x,y) \sim (x+1,y) \sim (x,y+1),$$
equipped with the quotient topology (cf. Section 10.1).

(2) The **Klein bottle** K^2 is the quotient of the unit square $Q = [0,1]^2 \subset \mathbb{R}^2$ by the equivalence relation
$$(x,y) \sim (x+1,y) \sim (x,1-y).$$

(3) The **projective plane** $\mathbb{R}P^2$ is the quotient of the unit square $Q = [0,1]^2 \subset \mathbb{R}^2$ by the equivalence relation
$$(x,y) \sim (1-x,y) \sim (x,1-y).$$

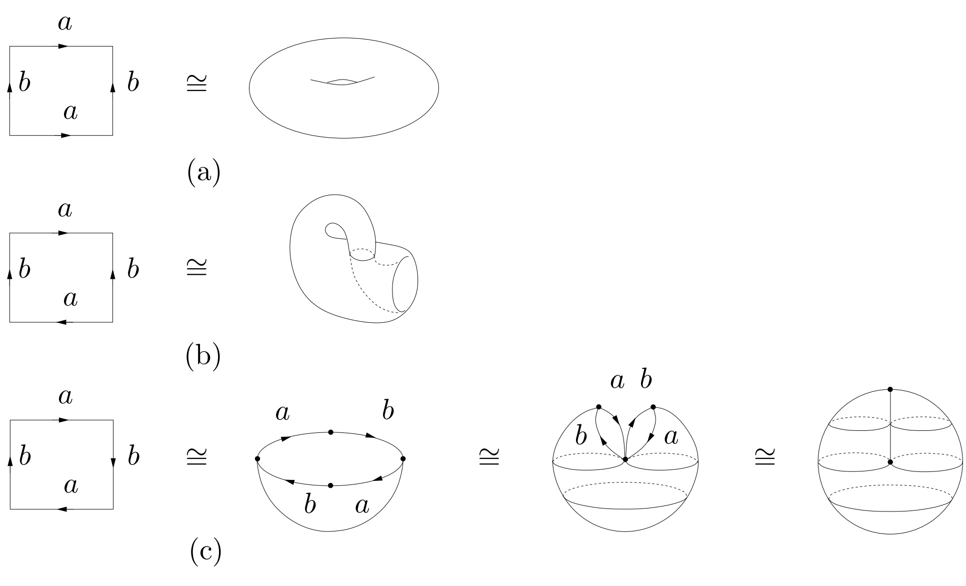

FIGURE 2. (a) Torus (T^2), (b) Klein bottle (K^2), (c) Real projective plane ($\mathbb{R}P^2$).

Remark 1.5.

(1) The only compact connected 1-dimensional topological manifold is the circle S^1 (see [**Mil97**]).

(2) The **connected sum** of two topological manifolds M and N is the topological manifold $M\#N$ obtained by deleting an open set homeomorphic to a ball on each manifold and gluing the boundaries by an homeomorphism (cf. Figure 3). It can be shown that any compact connected 2-dimensional topological manifold is homeomorphic either to S^2 or to connected sums of manifolds from Example 1.4 (see [**Blo96, Mun00**]).

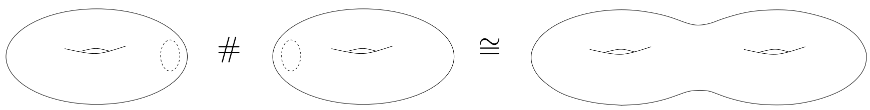

Figure 3. Connected sum of two tori.

If we do not identify all the edges of the square, we obtain a cylinder or a Möbius band (cf. Figure 4). These topological spaces are examples of **manifolds with boundary**.

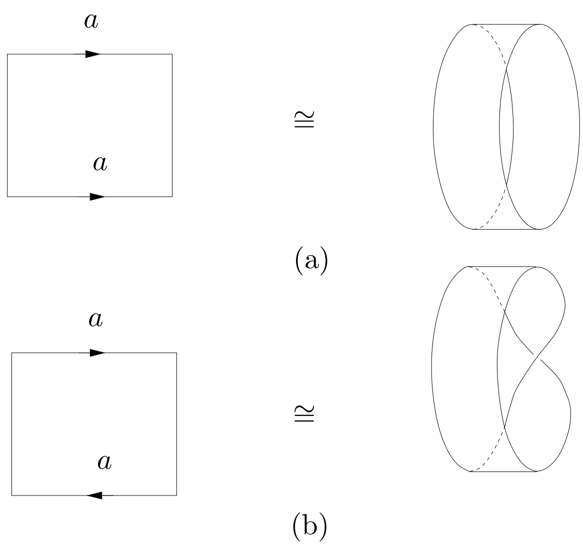

Figure 4. (a) Cylinder, (b) Möbius band.

Definition 1.6. *Consider the* **closed half space**

$$\mathbb{H}^n = \{(x^1, \dots, x^n) \in \mathbb{R}^n : x^n \geq 0\}.$$

A **topological manifold with boundary** *is a Hausdorff space M, with a countable basis of open sets, such that each point $p \in M$ possesses a neighborhood V which is homeomorphic either to an open subset U of $\mathbb{H}^n \backslash \partial\mathbb{H}^n$,* *or to an open subset U of $\mathbb{H}^n$, with the point p identified to a point in $\partial\mathbb{H}^n$. The points of the first type are called* **interior points**, *and the remaining ones are called* **boundary points**.

The set of boundary points ∂M is called **boundary** of M and is a manifold of dimension $(n-1)$.

REMARK 1.7.

1. Making a paper model of the Möbius band, we can easily verify that its boundary is homeomorphic to a circle (not to two disjoint circles), and that it has only one side (cf. Figure 4).
2. Both the Klein bottle and the real projective plane contain Möbius bands (cf. Figure 5). Deleting this band on the projective plane, we obtain a disk (cf. Figure 6). In other words, we can glue a Möbius band to a disk along their boundaries and obtain $\mathbb{R}P^2$.

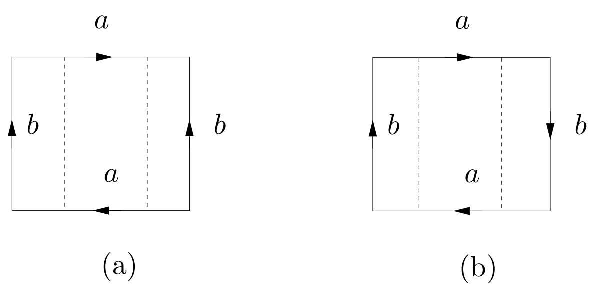

FIGURE 5. (a) Klein bottle, (b) Real projective plane.

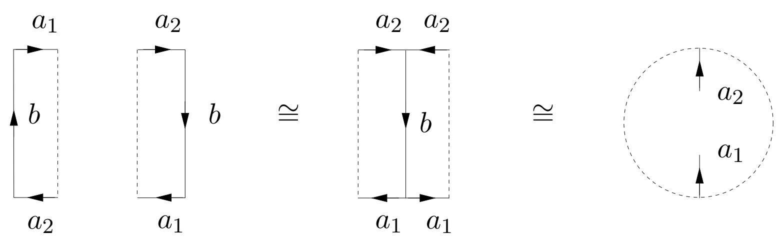

FIGURE 6. Disk inside the real projective plane.

Two topological manifolds are considered the same if they are homeomorphic. For example, spheres of different radii in $\mathbb{R}^3$ are homeomorphic, and so are the two surfaces in Figure 7. Indeed, the knotted torus can be obtained by cutting the torus along a circle, knotting it and gluing it back again. An obvious homeomorphism is then the one which takes each point on the initial torus to its final position after cutting and gluing.

EXERCISES 1.8.

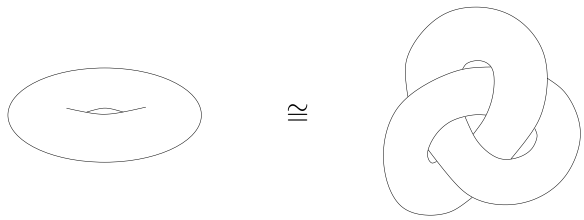

FIGURE 7. Two homeomorphic topological manifolds.

(1) Which of the following sets (with the subspace topology) are topological manifolds?
  (a) $D^2 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$;
  (b) $S^2 \setminus \{p\}$ $(p \in S^2)$;
  (c) $S^2 \setminus \{p, q\}$ $(p, q \in S^2, p \neq q)$;
  (d) $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 = 1\}$;
  (e) $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 = z^2\}$;
(2) Which of the manifolds above are homeomorphic?
(3) Show that the Klein bottle K^2 can be obtained by gluing two Möbius bands together through a homeomorphism of the boundary.
(4) Show that
  (a) $M \# S^2 = M$ for any 2-dimensional topological manifold M;
  (b) $\mathbb{R}P^2 \# \mathbb{R}P^2 = K^2$;
  (c) $\mathbb{R}P^2 \# T^2 = \mathbb{R}P^2 \# K^2$.
(5) A **triangulation** of a 2-dimensional topological manifold M is a decomposition of M in a finite number of triangles (i.e. images of Euclidean triangles by homeomorphisms) such that the intersection of any two triangles is either empty or composed of common edges or common vertices (it is possible to prove that such a triangulation always exists). The **Euler characteristic** of M is

$$\chi(M) := V - E + F,$$

where V, E and F are the number of vertices, edges and faces of a given triangulation. Show that:
  (a) $\chi(M)$ is well defined, i.e., does not depend on the choice of triangulation;
  (b) $\chi(S^2) = 2$;
  (c) $\chi(T^2) = 0$;
  (d) $\chi(K^2) = 0$;
  (e) $\chi(\mathbb{R}P^2) = 1$;
  (f) $\chi(M \# N) = \chi(M) + \chi(N) - 2$.

2. Differentiable Manifolds

Recall that an n-dimensional topological manifold is a Hausdorff space with a countable basis of open sets such that each point possesses a neighborhood homeomorphic to an open subset of $\mathbb{R}^n$. Each pair (U, φ), where U is an open subset of $\mathbb{R}^n$ and $\varphi : U \to \varphi(U) \subset M$ is a homeomorphism of U to an open subset of M, is called a **parametrization**. The inverse φ^{-1} is called a **coordinate system** or **chart**, and the set $\varphi(U) \subset M$ is called a **coordinate neighborhood**. When two coordinate neighborhoods overlap, we have formulas for the associated coordinate change (cf. Figure 8). The idea to obtain differentiable manifolds will be to choose a sub-collection of parametrizations so that the coordinate changes are differentiable maps.

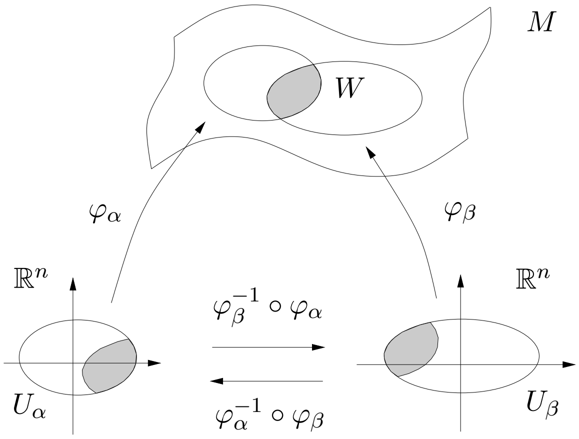

FIGURE 8. Parametrizations and overlap maps.

DEFINITION 2.1. *An n-dimensional* **differentiable** *or* **smooth manifold** *is a topological manifold of dimension n and a family of parametrizations $\varphi_\alpha : U_\alpha \to M$ defined on open sets $U_\alpha \subset \mathbb{R}^n$, such that:*

(i) the coordinate neighborhoods cover M, that is, $\bigcup_\alpha \varphi_\alpha(U_\alpha) = M$;

(ii) for each pair of indices α, β such that

$$W := \varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta) \neq \varnothing,$$

the overlap maps

$$\begin{aligned} \varphi_\beta^{-1} \circ \varphi_\alpha : \varphi_\alpha^{-1}(W) &\to \varphi_\beta^{-1}(W) \\ \varphi_\alpha^{-1} \circ \varphi_\beta : \varphi_\beta^{-1}(W) &\to \varphi_\alpha^{-1}(W) \end{aligned}$$

are C^∞;

(iii) the family $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ is maximal with respect to (i) and (ii), meaning that if $\varphi_0 : U_0 \to M$ is a parametrization such that $\varphi_0^{-1} \circ \varphi$ and $\varphi^{-1} \circ \varphi_0$ are C^∞ for all φ in $\mathcal{A}$, then (U_0, φ_0) is in $\mathcal{A}$.

REMARK 2.2.

(1) Any family $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ that satisfies (i) and (ii) is called a C^∞-**atlas** for M. If $\mathcal{A}$ also satisfies (iii) it is called a **maximal atlas** or a **differentiable structure**.
(2) Condition (iii) is purely technical. Given any atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ on M, there is a unique maximal atlas $\widetilde{\mathcal{A}}$ containing it. In fact, we can take the set $\widetilde{\mathcal{A}}$ of all parametrizations that satisfy (ii) with every parametrization on $\mathcal{A}$. Clearly $\mathcal{A} \subset \widetilde{\mathcal{A}}$, and one can easily check that $\widetilde{\mathcal{A}}$ satisfies (i) and (ii). Also, by construction, $\widetilde{\mathcal{A}}$ is maximal with respect to (i) and (ii). Two atlases are said to be **equivalent** if they define the same differentiable structure.
(3) We could also have defined C^k-manifolds by requiring the coordinate changes to be C^k-maps (a C^0-manifold would then denote a topological manifold).

EXAMPLE 2.3.

(1) The space $\mathbb{R}^n$ with the usual topology defined by the Euclidean metric is a Hausdorff space and has a countable basis of open sets. If, for instance, we consider a single parametrization $(\mathbb{R}^n, id)$, conditions (i) and (ii) of Definition 2.1 are trivially satisfied and we have an atlas for $\mathbb{R}^n$. The maximal atlas that contains this parametrization is usually called the **standard differentiable structure** on $\mathbb{R}^n$. We can of course consider other atlases. Take, for instance, the atlas defined by the parametrization $(\mathbb{R}^n, \varphi)$ with $\varphi(x) = Ax$ for a non-singular $(n \times n)$-matrix A. It is an easy exercise to show that these two atlases are equivalent.
(2) It is possible for a manifold to possess non-equivalent atlases: consider the two atlases $\{(\mathbb{R}, \varphi_1)\}$ and $\{(\mathbb{R}, \varphi_2)\}$ on $\mathbb{R}$, where $\varphi_1(x) = x$ and $\varphi_2(x) = x^3$. As the map $\varphi_2^{-1} \circ \varphi_1$ is not differentiable at the origin, these two atlases define different (though, as we shall see, diffeomorphic) differentiable structures (cf. Exercises 2.5.4 and 3.2.6).
(3) Every open subset V of a smooth manifold is a manifold of the same dimension. Indeed, as V is a subset of M, its subspace topology is Hausdorff and admits a countable basis of open sets. Moreover, if $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ is an atlas for M and we take the U_α's for which $\varphi_\alpha(U_\alpha) \cap V \neq \varnothing$, it is easy to check that the family of parametrizations $\widetilde{\mathcal{A}} = \{(\widetilde{U}_\alpha, \varphi_\alpha|_{\widetilde{U}_\alpha})\}$, where $\widetilde{U}_\alpha = \varphi_\alpha^{-1}(V)$, is an atlas for V.
(4) Let $M_{n\times n}$ be the set of $n \times n$ matrices with real coefficients. Rearranging the entries along one line, we see that this space is just $\mathbb{R}^{n^2}$, and so it is a manifold. By Example 3, we have that $GL(n) = \{A \in M_{n\times n} | \det A \neq 0\}$ is also a manifold of dimension n^2. In fact, the determinant is a continuous map from $M_{n\times n}$ to $\mathbb{R}$, and $GL(n)$ is the preimage of the open set $\mathbb{R}\backslash\{0\}$.

(5) Let us consider the n-sphere

$$S^n = \{(x^1, \ldots, x^{n+1}) \in \mathbb{R}^{n+1} | \, (x^1)^2 + \cdots + (x^{n+1})^2 = 1\}$$

and the maps

$$\begin{array}{rcl} \varphi_i^+ : U \subset \mathbb{R}^n & \to & S^n \\ (x^1, \ldots, x^n) & \mapsto & (x^1, \ldots, x^{i-1}, g(x^1, \ldots, x^n), x^i, \ldots, x^n), \end{array}$$

$$\begin{array}{rcl} \varphi_i^- : U \subset \mathbb{R}^n & \to & S^n \\ (x^1, \ldots, x^n) & \mapsto & (x^1, \ldots, x^{i-1}, -g(x^1, \ldots, x^n), x^i, \ldots, x^n), \end{array}$$

where

$$U = \{(x^1, \ldots, x^n) \in \mathbb{R}^n \mid (x^1)^2 + \cdots + (x^n)^2 < 1\}$$

and

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

Being a subset of $\mathbb{R}^{n+1}$, the sphere (equipped with the subspace topology) is a Hausdorff space and admits a countable basis of open sets. It is also easy to check that the family $\{(U, \varphi_i^+), (U, \varphi_i^-)\}_{i=1}^{n+1}$ is an atlas for S^n, and so this space is a manifold of dimension n (the corresponding charts are just the projections on the hyperplanes $x^i = 0$).

(6) We can define an atlas for the surface of a cube $Q \subset \mathbb{R}^3$ making it a smooth manifold: Suppose the cube is centered at the origin and consider the map $f : Q \to S^2$ defined by $f(x) = x/\|x\|$. Then, considering an atlas $\{(U_\alpha, \varphi_\alpha)\}$ for S^2, the family $\{(U_\alpha, f^{-1} \circ \varphi_\alpha)\}$ defines an atlas for Q.

Remark 2.4. There exist topological manifolds which admit no differentiable structures at all. Indeed, Kervaire presented the first example (a 10-dimensional manifold) in 1960 [**Ker60**], and Smale constructed another one (of dimension 12) soon after [**Sma60**]. In 1956 Milnor [**Mil56b**] had already given an example of a 8-manifold which he believed not to admit a differentiable structure, but that was not proved until 1965 (see [**Nov65**]).

Exercises 2.5.

(1) Show that two atlases $\mathcal{A}_1$ and $\mathcal{A}_2$ for a smooth manifold are equivalent if and only if $\mathcal{A}_1 \cup \mathcal{A}_2$ is an atlas.
(2) Let M be a differentiable manifold. Show that a set $V \subset M$ is open if and only if $\varphi_\alpha^{-1}(V)$ is an open subset of $\mathbb{R}^n$ for every parametrization $(U_\alpha, \varphi_\alpha)$ of a C^∞ atlas.
(3) Show that the two atlases on $\mathbb{R}^n$ from Example 2.3.1 are equivalent.
(4) Consider the two atlases on $\mathbb{R}$ from Example 2.3.2, $\{(\mathbb{R}, \varphi_1)\}$ and $\{(\mathbb{R}, \varphi_2)\}$, where $\varphi_1(x) = x$ and $\varphi_2(x) = x^3$. Show that $\varphi_2^{-1} \circ \varphi_1$ is not differentiable at the origin. Conclude that the two atlases are not equivalent.

(5) Recall from elementary vector calculus that a **surface** $S \subset \mathbb{R}^3$ is a set such that, for each $p \in M$, there is a neighborhood V_p of p in $\mathbb{R}^3$ and a C^∞ map $f_p : U_p \to \mathbb{R}$ (where U_p is an open subset of $\mathbb{R}^2$) such that $S \cap V_p$ is the graph of $z = f_p(x, y)$, or $x = f_p(y, z)$, or $y = f_p(x, z)$. Show that S is a smooth manifold of dimension 2.

(6) (*Product manifold*) Let $\{(U_\alpha, \varphi_\alpha)\}$, $\{(V_\beta, \psi_\beta)\}$ be two atlases for two smooth manifolds M and N. Show that the family $\{(U_\alpha \times V_\beta, \varphi_\alpha \times \psi_\beta)\}$ is an atlas for the product $M \times N$. With the differentiable structure generated by this atlas, $M \times N$ is called the **product manifold** of M and N.

(7) (*Stereographic projection*) Consider the n-sphere S^n with the subspace topology and let $N = (0, \dots, 0, 1)$ and $S = (0, \dots, 0, -1)$ be the north and south poles. The **stereographic projection** from N is the map $\pi_N : S^n \backslash \{N\} \to \mathbb{R}^n$ which takes a point $p \in S^n \backslash \{N\}$ to the intersection point of the line through N and p with the hyperplane $x^{n+1} = 0$ (cf. Figure 9). Similarly, the stereographic projection from S is the map $\pi_S : S^n \backslash \{S\} \to \mathbb{R}^n$ which takes a point p on $S^n \backslash \{S\}$ to the intersection point of the line through S and p with the same hyperplane. Check that $\{(\mathbb{R}^n, \pi_N^{-1}), (\mathbb{R}^n, \pi_S^{-1})\}$ is an atlas for S^n. Show that this atlas is equivalent to the atlas on Example 2.3.5. The maximal atlas obtained from these is called the **standard differentiable structure** on S^n.

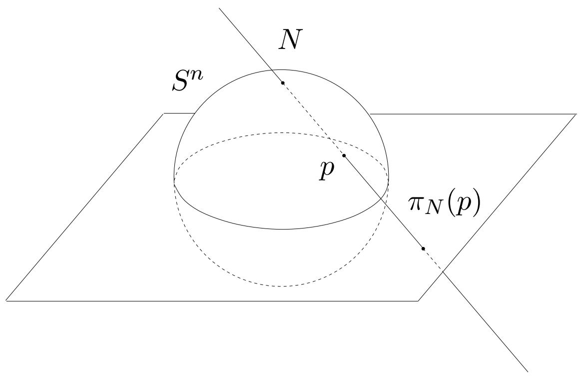

FIGURE 9. Stereographic projection.

(8) (*Real projective space*) The **real projective space** $\mathbb{R}P^n$ is the set of lines through the origin in $\mathbb{R}^{n+1}$. This space can be defined as the quotient space of S^n by the equivalence relation $x \sim -x$ that identifies a point to its antipodal point.
   (a) Show that the quotient space $\mathbb{R}P^n = S^n/\sim$ with the quotient topology is a Hausdorff space and admits a countable basis of open sets (**Hint:** Use Proposition 10.2);

(b) Considering the atlas on S^n defined in Example 2.3.5 and the canonical projection $\pi : S^n \to \mathbb{R}P^n$ given by $\pi(x) = [x]$, define an atlas for $\mathbb{R}P^n$.

(9) We can define an atlas on $\mathbb{R}P^n$ in a different way by identifying it with the quotient space of $\mathbb{R}^{n+1}\backslash\{0\}$ by the equivalence relation $x \sim \lambda x$, with $\lambda \in \mathbb{R}\backslash\{0\}$. For that, consider the sets $V_i = \{[x^1, \dots, x^{n+1}] \,|\, x^i \neq 0\}$ (corresponding to the set of lines through the origin in $\mathbb{R}^{n+1}$ that are not contained on the hyperplane $x^i = 0$) and the maps $\varphi_i : \mathbb{R}^n \to V_i$ defined by

$$\varphi_i(x^1, \dots, x^n) = [x^1, \dots, x^{i-1}, 1, x^i, \dots, x^n].$$

Show that:

(a) the family $\{(\mathbb{R}^n, \varphi_i)\}$ is an atlas for $\mathbb{R}P^n$;

(b) this atlas defines the same differentiable structure as the atlas on Exercise 2.5.8.

(10) (*A non-Hausdorff manifold*) Let M be the disjoint union of $\mathbb{R}$ with a point p and consider the maps $f_i : \mathbb{R} \to M$ $(i = 1, 2)$ defined by $f_i(x) = x$ if $x \in \mathbb{R}\backslash\{0\}$, $f_1(0) = 0$ and $f_2(0) = p$. Show that:

(a) the maps $f_i^{-1} \circ f_j$ are differentiable on their domains;

(b) if we consider an atlas formed by $\{(\mathbb{R}, f_1), (\mathbb{R}, f_2)\}$, the corresponding topology will not satisfy the Hausdorff axiom.

3. Differentiable Maps

In this book the words **differentiable** and **smooth** will be used to mean **infinitely differentiable** (C^∞).

DEFINITION 3.1. *Let M and N be two differentiable manifolds of dimension m and n, respectively. A map $f : M \to N$ is said to be* **differentiable** *(or* **smooth***, or C^∞) at a point $p \in M$ if there exist parametrizations (U, φ) of M at p (i.e. $p \in \varphi(U)$) and (V, ψ) of N at $f(p)$, with $f(\varphi(U)) \subset \psi(V)$, such that the map*

$$\hat{f} := \psi^{-1} \circ f \circ \varphi : U \subset \mathbb{R}^m \to \mathbb{R}^n$$

is differentiable at $\varphi^{-1}(p)$.

The map f is said to be differentiable on a subset of M if it is differentiable at every point of this set.

As coordinate changes are smooth, this definition is independent of the parametrizations chosen at $f(p)$ and p. The map $\hat{f} := \psi^{-1} \circ f \circ \varphi : U \subset \mathbb{R}^m \to \mathbb{R}^n$ is called a **local representation** of f and is the expression of f on the local coordinates defined by φ and ψ. The set of all smooth functions $f : M \to N$ is denoted $C^\infty(M, N)$, and we will simply write $C^\infty(M)$ for $C^\infty(M, \mathbb{R})$.

A differentiable map $f : M \to N$ between two manifolds is continuous (cf. Exercise 3.2.2). Moreover, it is called a **diffeomorphism** if it is bijective and its inverse $f^{-1} : N \to M$ is also differentiable. The differentiable

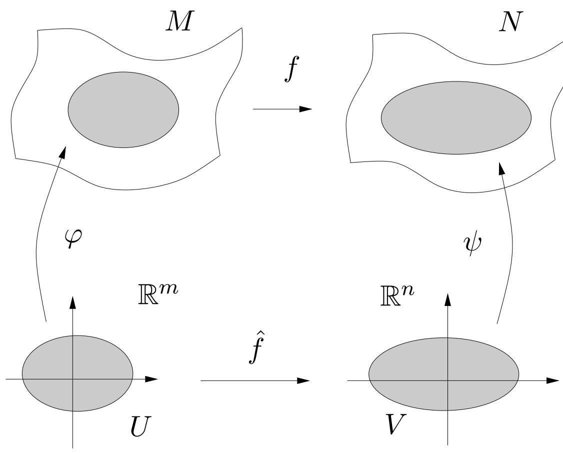

FIGURE 10. Local representation of a map between manifolds.

manifolds M and N will be considered the same if they are **diffeomorphic**, i.e. if there exists a diffeomorphism $f : M \to N$. A map f is called a **local diffeomorphism** at a point $p \in M$ if there are neighborhoods V of p and W of $f(p)$ such that $f|_V : V \to W$ is a diffeomorphism.

For a long time it was thought that, up to a diffeomorphism, there was only one differentiable structure for each topological manifold (the two different differentiable structures in Exercises 2.5.4 and 3.2.6 are diffeomorphic – cf. Exercise 3.26). However, in 1956, Milnor [**Mil56a**] presented examples of manifolds that were homeomorphic but not diffeomorphic to S^7. Later, Milnor and Kervaire [**Mil59, KM63**] showed that more spheres of dimension greater than 7 admitted several differentiable structures. For instance, S^{19} has 73 distinct smooth structures and S^{31} has $16,931,177$. More recently, in 1982 and 1983, Freedman [**Fre82**] and Gompf [**Gom83**] constructed examples of non-standard differentiable structures on $\mathbb{R}^4$.

EXERCISES 3.2.

(1) Prove that Definition 3.1 does not depend on the choice of parametrizations.
(2) Show that a differentiable map $f : M \to N$ between two smooth manifolds is continuous.
(3) Show that if $f : M_1 \to M_2$ and $g : M_2 \to M_3$ are differentiable maps between smooth manifolds M_1, M_2 and M_3, then $g \circ f : M_1 \to M_3$ is also differentiable.
(4) Show that the antipodal map $f : S^n \to S^n$, defined by $f(x) = -x$, is differentiable.
(5) Using the stereographic projection from the north pole $\pi_N : S^2 \setminus \{N\} \to \mathbb{R}^2$ and identifying $\mathbb{R}^2$ with the complex plane $\mathbb{C}$, we can identify S^2 with $\mathbb{C} \cup \{\infty\}$, where ∞ is the so-called **point at infinity**. A **Möbius transformation** is a map $f : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$

of the form

$$f(z) = \frac{az+b}{cz+d},$$

where $a,b,c,d \in \mathbb{C}$ satisfy $ad - bc \neq 0$ and ∞ satisfies

$$\frac{\alpha}{\infty} = 0, \qquad \frac{\alpha}{0} = \infty$$

for any $\alpha \in \mathbb{C} \setminus \{0\}$. Show that any Möbius transformation f, seen as a map $f : S^2 \to S^2$, is a diffeomorphism. (**Hint:** Start by showing that any Möbius transformation is a composition of transformations of the form $g(z) = \frac{1}{z}$ and $h(z) = az + b$).

(6) Consider again the two atlases on $\mathbb{R}$ from Example 2.3.2 and Exercise 2.5.4, $\{(\mathbb{R}, \varphi_1)\}$ and $\{(\mathbb{R}, \varphi_2)\}$, where $\varphi_1(x) = x$ and $\varphi_2(x) = x^3$. Show that:
   (a) the identity map $i : (\mathbb{R}, \varphi_1) \to (\mathbb{R}, \varphi_2)$ is not a diffeomorphism;
   (b) the map $f : (\mathbb{R}, \varphi_1) \to (\mathbb{R}, \varphi_2)$ defined by $f(x) = x^3$ is a diffeomorphism (implying that although these two atlases define different differentiable structures, they are diffeomorphic).

4. Tangent Space

Recall from elementary vector calculus that a vector $v \in \mathbb{R}^3$ is said to be **tangent** to a surface $S \subset \mathbb{R}^3$ at a point $p \in S$ if there exists a differentiable curve $c : (-\varepsilon, \varepsilon) \to S \subset \mathbb{R}^3$ such that $c(0) = p$ and $\dot{c}(0) = v$ (cf. Exercise 2.5.5). The set T_pS of all these vectors is a 2-dimensional vector space, called the **tangent space** to S at p, and can be identified with the plane in $\mathbb{R}^3$ wich is tangent to S at p.

To generalize this to an abstract n-dimensional manifold we need to find a description of v which does not involve the ambient Euclidean space $\mathbb{R}^3$. To do so, we notice that the components of v are

$$v^i = \frac{d(x^i \circ c)}{dt}(0),$$

where $x^i : \mathbb{R}^3 \to \mathbb{R}$ is the i-th coordinate function. If we ignore the ambient space, $x^i : S \to \mathbb{R}$ is just a differentiable function, and

$$v^i = v(x^i),$$

where, for any differentiable function $f : S \to \mathbb{R}$, we define

$$v(f) := \frac{d(f \circ c)}{dt}(0).$$

This allows us to see v as an operator $v : C^\infty(S) \to \mathbb{R}$, and it is clear that this operator completely determines v. It is this new interpretation of tangent vector that will be used to define tangent spaces for manifolds.

Definition 4.1. *Let $c : (-\varepsilon, \varepsilon) \to M$ be a differentiable curve on a smooth manifold M. Consider the set $C^\infty(p)$ of all functions $f : M \to \mathbb{R}$ that are differentiable at $c(0) = p$ (i.e., C^∞ on a neighborhood of p). The*

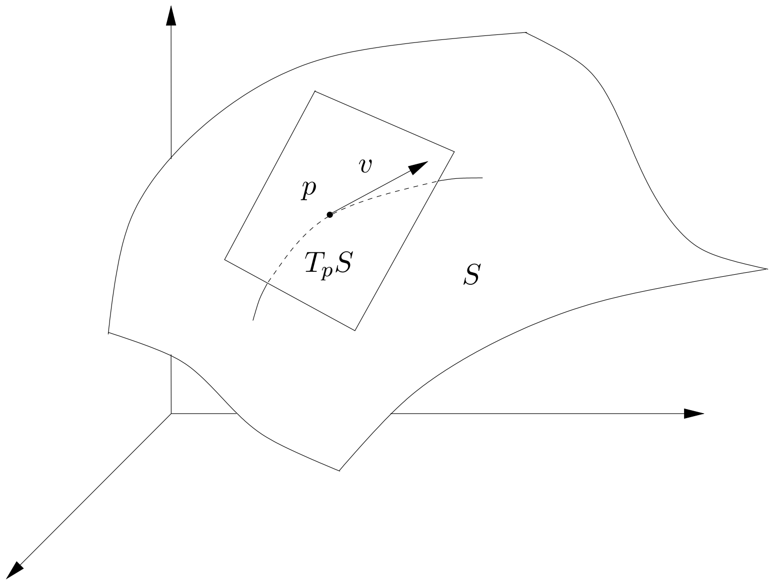

FIGURE 11. Tangent vector to a surface.

tangent vector to the curve c *at* p *is the operator* $\dot{c}(0) : C^\infty(p) \to \mathbb{R}$ *given by*

$$\dot{c}(0)(f) = \frac{d(f \circ c)}{dt}(0).$$

A **tangent vector** *to* M *at* p *is a tangent vector to some differentiable curve* $c : (-\varepsilon, \varepsilon) \to M$ *with* $c(0) = p$*. The* **tangent space** *at* p *is the space* T_pM *of all tangent vectors at* p*.*

Choosing a parametrization $\varphi : U \subset \mathbb{R}^n \to M$ around p, the curve c is given in local coordinates by the curve in U

$$\hat{c}(t) := \left(\varphi^{-1} \circ c\right)(t) = (x^1(t), \dots, x^n(t)),$$

and

$$\begin{aligned}
\dot{c}(0)(f) &= \frac{d(f \circ c)}{dt}(0) = \frac{d}{dt}\left(\overbrace{(f \circ \varphi)}^{\hat{f}} \circ \overbrace{(\varphi^{-1} \circ c)}^{\hat{c}}\right)_{|t=0} = \\
&= \frac{d}{dt}\left(\hat{f}(x^1(t), \dots, x^n(t))\right)_{|t=0} = \sum_{i=1}^{n} \frac{\partial \hat{f}}{\partial x^i}(\hat{c}(0)) \frac{dx^i}{dt}(0) = \\
&= \left(\sum_{i=1}^{n} \dot{x}^i(0) \left(\frac{\partial}{\partial x^i}\right)_{\varphi^{-1}(p)}\right)(\hat{f}).
\end{aligned}$$

Hence we can write

$$\dot{c}(0) = \sum_{i=1}^{n} \dot{x}^i(0) \left(\frac{\partial}{\partial x^i} \right)_p ,$$

where $\left(\frac{\partial}{\partial x^i}\right)_p$ denotes the operator defined by the vector tangent to the curve c_i at p given in local coordinates by

$$\hat{c}_i(t) = (x^1, \dots, x^{i-1}, x^i + t, x^{i+1}, \dots, x^n),$$

with $(x^1, \dots, x^n) = \varphi^{-1}(p)$.

EXAMPLE 4.2. The map $\psi : (0, \pi) \times (-\pi, \pi) \to S^2$ given by

$$\psi(\theta, \varphi) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

parametrizes a neighborhood of the point $(1, 0, 0) = \psi\left(\frac{\pi}{2}, 0\right)$. Consequently, $\left(\frac{\partial}{\partial\theta}\right)_{(1,0,0)} = \dot{c}_\theta(0)$ and $\left(\frac{\partial}{\partial\varphi}\right)_{(1,0,0)} = \dot{c}_\varphi(0)$, where

$$c_\theta(t) = \psi\left(\frac{\pi}{2} + t, 0\right) = (\cos t, 0, -\sin t);$$
$$c_\varphi(t) = \psi\left(\frac{\pi}{2}, t\right) = (\cos t, \sin t, 0).$$

Note that, in the notation above,

$$\hat{c}_\theta(t) = \left(\frac{\pi}{2} + t, 0\right) \quad \text{and} \quad \hat{c}_\varphi(t) = \left(\frac{\pi}{2}, t\right).$$

Moreover, since c_θ and c_φ are curves in $\mathbb{R}^3$, $\left(\frac{\partial}{\partial\theta}\right)_{(1,0,0)}$ and $\left(\frac{\partial}{\partial\varphi}\right)_{(1,0,0)}$ can be identified with the vectors $(0, 0, -1)$ and $(0, 1, 0)$.

PROPOSITION 4.3. *The tangent space to M at p is an n-dimensional vector space.*

PROOF. Consider a parametrization $\varphi : U \subset \mathbb{R}^n \to M$ around p and take the vector space generated by the operators $\left(\frac{\partial}{\partial x^i}\right)_p$,

$$\mathcal{D}_p := \operatorname{span}\left\{ \left(\frac{\partial}{\partial x^1}\right)_p, \dots, \left(\frac{\partial}{\partial x^n}\right)_p \right\}.$$

It is easy to show (cf. Exercise 4.9.1) that these operators are linearly independent. Moreover, each tangent vector to M at p can be represented by a linear combination of these operators, so the tangent space T_pM is a subset of $\mathcal{D}_p$. We will now see that $\mathcal{D}_p \subset T_pM$. Let $v \in \mathcal{D}_p$; then v can be written as

$$v = \sum_{i=1}^{n} v^i \left(\frac{\partial}{\partial x^i}\right)_p .$$

If we consider the curve $c : (-\varepsilon, \varepsilon) \to M$, defined by

$$c(t) = \varphi(x^1 + v^1 t, \dots, x^n + v^n t)$$

(where $(x^1, \dots, x^n) = \varphi^{-1}(p)$), then

$$\hat{c}(t) = (x^1 + v^1 t, \dots, x^n + v^n t)$$

and so $\dot{x}^i(0) = v^i$, implying that $\dot{c}(0) = v$. Therefore $v \in T_pM$. $\square$

REMARK 4.4.

(1) The basis $\left\{ \left(\frac{\partial}{\partial x^i}\right)_p \right\}_{i=1}^n$ determined by the chosen parametrization around p is called the **associated basis** to that parametrization.

(2) Note that the definition of tangent space at p only uses functions that are differentiable on a neighborhood of p. Hence, if U is an open set of M containing p, the tangent space T_pU is naturally identified with T_pM.

If we consider the disjoint union of all tangent spaces T_pM at all points of M, we obtain the space

$$TM = \bigcup_{p \in M} T_pM = \{v \in T_pM \mid p \in M\},$$

which admits a differentiable structure naturally determined by the one on M (cf. Exercise 4.9.8). With this differentiable structure, this space is called the **tangent bundle**. Note that there is a natural projection $\pi : TM \to M$ which takes $v \in T_pM$ to p (cf. Section 10.3).

Now that we have defined tangent space, we can define the **derivative at a point** p of a differentiable map $f : M \to N$ between smooth manifolds. We want this derivative to be a linear transformation

$$(df)_p : T_pM \to T_{f(p)}N$$

of the corresponding tangent spaces, to be the usual derivative (Jacobian) of f when M and N are Euclidean spaces, and to satisfy the chain rule.

DEFINITION 4.5. *Let $f : M \to N$ be a differentiable map between smooth manifolds. For $p \in M$, the derivative of f at p is the map*

$$\begin{aligned} (df)_p : T_pM &\to T_{f(p)}N \\ v &\mapsto \frac{d\,(f \circ c)}{dt}(0), \end{aligned}$$

where $c : (-\varepsilon, \varepsilon) \to M$ is a curve satisfying $c(0) = p$ and $\dot{c}(0) = v$.

PROPOSITION 4.6. *The map $(df)_p : T_pM \to T_{f(p)}N$ defined above is a linear transformation that does not depend on the choice of the curve c.*

PROOF. Let (U, φ) and (V, ψ) be two parametrizations around p and $f(p)$ such that $f(\varphi(U)) \subset \psi(V)$ (cf. Figure 12). Consider a vector $v \in T_pM$ and a curve $c : (-\varepsilon, \varepsilon) \to M$ such that $c(0) = p$ and $\dot{c}(0) = v$. If, in local coordinates, the curve c is given by

$$\hat{c}(t) := (\varphi^{-1} \circ c)(t) = (x^1(t), \dots, x^m(t)),$$

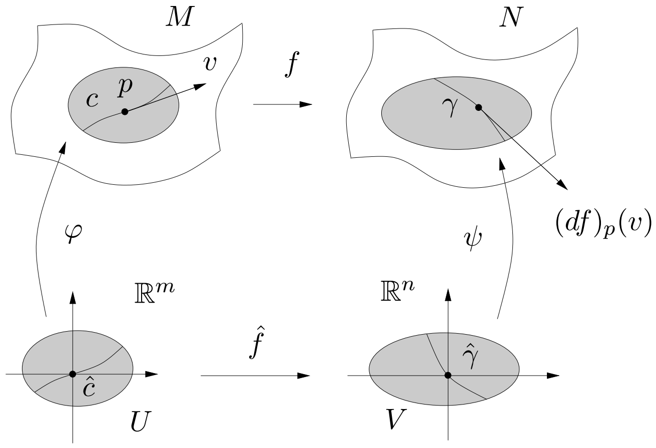

FIGURE 12. Derivative of a differentiable map.

and the curve $\gamma := f \circ c : (-\varepsilon, \varepsilon) \to N$ is given by

$$\begin{aligned}\hat{\gamma}(t) := \left(\psi^{-1} \circ \gamma\right)(t) &= \left(\psi^{-1} \circ f \circ \varphi\right)(x^1(t), \dots, x^m(t)) \\ &= (y^1(x(t)), \dots, y^n(x(t))),\end{aligned}$$

then $\dot{\gamma}(0)$ is the tangent vector in $T_{f(p)}N$ given by

$$\begin{aligned}\dot{\gamma}(0) &= \sum_{i=1}^{n} \frac{d}{dt}\left(y^i(x^1(t), \dots, x^m(t))\right)_{|t=0} \left(\frac{\partial}{\partial y^i}\right)_{f(p)} \\ &= \sum_{i=1}^{n} \left\{\sum_{k=1}^{m} \dot{x}^k(0) \left(\frac{\partial y^i}{\partial x^k}\right)(x(0))\right\} \left(\frac{\partial}{\partial y^i}\right)_{f(p)} \\ &= \sum_{i=1}^{n} \left\{\sum_{k=1}^{m} v^k \left(\frac{\partial y^i}{\partial x^k}\right)(x(0))\right\} \left(\frac{\partial}{\partial y^i}\right)_{f(p)},\end{aligned}$$

where the v^k are the components of v in the basis associated to (U, φ). Hence $\dot{\gamma}(0)$ does not depend on the choice of c, as long as $\dot{c}(0) = v$. Moreover, the components of $w = (df)_p(v)$ in the basis associated to (V, ψ) are

$$w^i = \sum_{j=1}^{n} \frac{\partial y^i}{\partial x^j} v^j,$$

where $\left(\frac{\partial y^i}{\partial x^j}\right)$ is an $n \times m$ matrix (the Jacobian matrix of the local representation of f at $\varphi^{-1}(p)$). Therefore, $(df)_p : T_pM \to T_{f(p)}N$ is the linear transformation which, on the basis associated to the parametrizations φ and ψ, is represented by this matrix. □

REMARK 4.7. The derivative $(df)_p$ is sometimes called **differential** of f at p. Several other notations are often used for df, as for example f_*, Df and f'.

EXAMPLE 4.8. Let $\varphi : U \subset \mathbb{R}^n \to M$ be a parametrization around a point $p \in M$. We can view φ as a differentiable map between two smooth manifolds and we can compute its derivative at $x = \varphi^{-1}(p)$

$$(d\varphi)_x : T_xU \to T_pM.$$

For $v \in T_xU \cong \mathbb{R}^n$, the i-th component of $(d\varphi)_x(v)$ is

$$\sum_{j=1}^{n} \frac{\partial x^i}{\partial x^j} v^j = v^i$$

(where $\left(\frac{\partial x^i}{\partial x^j}\right)$ is the identity matrix). Hence, $(d\varphi)_x(v)$ is the vector in T_pM which, in the basis $\left\{\left(\frac{\partial}{\partial x^i}\right)_p\right\}$ associated to the parametrization φ, is represented by v.

Given a differentiable map $f : M \to N$ we can also define a global derivative df (also called **push-forward** and denoted f_*) between the corresponding tangent bundles:

$$\begin{aligned} df : TM &\to TN \\ T_pM \ni v &\mapsto (df)_p(v) \in T_{f(p)}N. \end{aligned}$$

EXERCISES 4.9.

(1) Show that the operators $\left(\frac{\partial}{\partial x^i}\right)_p$ are linearly independent.

(2) Let M be a smooth manifold, p a point in M and v a vector tangent to M at p. If, for two basis associated to different parametrizations around p, v can be written as $v = \sum_{i=1}^n a^i (\frac{\partial}{\partial x^i})_p$ and $v = \sum_{i=1}^n b^i (\frac{\partial}{\partial y^i})_p$, show that

$$b^j = \sum_{i=1}^{n} \frac{\partial y^j}{\partial x^i} a^i.$$

(3) Let M be an n-dimensional differentiable manifold and $p \in M$. Show that the following sets can be canonically identified with T_pM (and therefore constitute alternative definitions of the tangent space):

(a) $\mathcal{C}_p/\sim$, where $\mathcal{C}_p$ is the set of differentiable curves $c : I \subset \mathbb{R} \to M$ such that $c(0) = p$ and $\sim$ is the equivalence relation defined by

$$c_1 \sim c_2 \Leftrightarrow \frac{d}{dt}(\varphi^{-1} \circ c_1)(0) = \frac{d}{dt}(\varphi^{-1} \circ c_2)(0)$$

for some parametrization $\varphi : U \subset \mathbb{R}^n \to M$ of a neighborhood of p.

(b) $\{(\alpha, v_\alpha) : p \in \varphi_\alpha(U_\alpha) \text{ and } v_\alpha \in \mathbb{R}^n\}/\sim$, where $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ is the differentiable structure and $\sim$ is the equivalence relation defined by

$$(\alpha, v_\alpha) \sim (\beta, v_\beta) \Leftrightarrow v_\beta = d(\varphi_\beta^{-1} \circ \varphi_\alpha)_{\varphi_\alpha^{-1}(p)}(v_\alpha).$$

(4) (*Chain Rule*) Let $f : M \to N$ and $g : N \to P$ be two differentiable maps. Then $g \circ f : M \to P$ is also differentiable (cf. Exercise 3.2.3). Show that for $p \in M$,

$$(d(g \circ f))_p = (dg)_{f(p)} \circ (df)_p.$$

(5) Let $\phi : (0, +\infty) \times (0, \pi) \times (0, 2\pi) \to \mathbb{R}^3$ be the parametrization of $U = \mathbb{R}^3 \setminus \{(x, 0, z) \mid x \geq 0 \text{ and } z \in \mathbb{R}\}$ by spherical coordinates,

$$\phi(r, \theta, \varphi) = (r \sin\theta \cos\varphi, r \sin\theta \sin\varphi, r \cos\theta).$$

Determine the Cartesian components of $\frac{\partial}{\partial r}$, $\frac{\partial}{\partial \theta}$ and $\frac{\partial}{\partial \varphi}$ at each point of U.

(6) Compute the derivative $(df)_N$ of the antipodal map $f : S^n \to S^n$ at the north pole N.

(7) Let W be a coordinate neighborhood on M, let $x : W \to \mathbb{R}^n$ be a coordinate chart and consider a smooth function $f : M \to \mathbb{R}$. Show that for $p \in W$, the derivative $(df)_p$ is given by

$$(df)_p = \frac{\partial \hat{f}}{\partial x^1}(x(p))(dx^1)_p + \cdots + \frac{\partial \hat{f}}{\partial x^n}(x(p))(dx^n)_p,$$

where $\hat{f} := f \circ x^{-1}$.

(8) (*Tangent bundle*) Let $\{(U_\alpha, \varphi_\alpha)\}$ be a differentiable structure on M and consider the maps

$$\begin{array}{rcl} \Phi_\alpha : U_\alpha \times \mathbb{R}^n & \to & TM \\ (x, v) & \mapsto & (d\varphi_\alpha)_x(v) \in T_{\varphi_\alpha(x)}M. \end{array}$$

Show that the family $\{(U_\alpha \times \mathbb{R}^n, \Phi_\alpha)\}$ defines a differentiable structure for TM. Conclude that, with this differentiable structure, TM is a smooth manifold of dimension $2 \times \dim M$.

(9) Let $f : M \to N$ be a differentiable map between smooth manifolds. Show that:
   (a) $df : TM \to TN$ is also differentiable;
   (b) if $f : M \to M$ is the identity map then $df : TM \to TM$ is also the identity;
   (c) if f is a diffeomorphism then $df : TM \to TN$ is also a diffeomorphism and $(df)^{-1} = df^{-1}$.

(10) Let M_1, M_2 be two differentiable manifolds and

$$\begin{array}{rcl} \pi_1 : M_1 \times M_2 & \to & M_1 \\ \pi_2 : M_1 \times M_2 & \to & M_2 \end{array}$$

the corresponding canonical projections.
   (a) Show that $d\pi_1 \times d\pi_2$ is a diffeomorphism between the tangent bundle $T(M_1 \times M_2)$ and the product manifold $TM_1 \times TM_2$.

(b) Show that if N is a smooth manifold and $f_i : N \to M_i$ $(i = 1, 2)$ are differentiable maps, then $d(f_1 \times f_2) = df_1 \times df_2$.

5. Immersions and Embeddings

In this section we will study the local behavior of differentiable maps $f : M \to N$ between smooth manifolds. We have already seen that f is said to be a local diffeomorphism at a point $p \in M$ if $\dim M = \dim N$ and f transforms a neighborhood of p diffeomorphically onto a neighborhood of $f(p)$. In this case, its derivative $(df)_p : T_pM \to T_{f(p)}N$ must necessarily be an isomorphism (cf. Exercise 4.9.9c). Conversely, if $(df)_p$ is an isomorphism then the Inverse Function Theorem implies that f is a local diffeomorphism (cf. Section 10.4). Therefore, to check whether f maps a neighborhood of p diffeomorphically onto a neighborhood of $f(p)$, one just has to check that the determinant of the local representation of $(df)_p$ is nonzero.

When $\dim M < \dim N$, the best we can hope for is that $(df)_p : T_pM \to T_{f(p)}N$ is injective. The map f is then called an **immersion at** p. If f is an immersion at every point in M, it is called an **immersion**. Locally, every immersion is (up to a diffeomorphism) the **canonical immersion** of $\mathbb{R}^m$ into $\mathbb{R}^n$ $(m < n)$ where a point $(x^1, \dots, x^m)$ is mapped to $(x^1, \dots, x^m, 0, \dots, 0)$. This result is known as the **Local Immersion Theorem** .

THEOREM 5.1. *Let $f : M \to N$ be an immersion at $p \in M$. Then there exist local coordinates around p and $f(p)$ on which f is the canonical immersion.*

PROOF. Let (U, φ) and (V, ψ) be parametrizations around p and $q = f(p)$. Let us assume for simplicity that $\varphi(0) = p$ and $\psi(0) = q$. Since f is an immersion, $(d\hat{f})_0 : \mathbb{R}^m \to \mathbb{R}^n$ is injective (where $\hat{f} := \psi^{-1} \circ f \circ \varphi$ is the expression of f in local coordinates). Hence we can assume (changing basis on $\mathbb{R}^n$ if necessary) that this linear transformation is represented by the $n \times m$ matrix

$$\begin{pmatrix} I_{m\times m} \\ --- \\ 0 \end{pmatrix},$$

where $I_{m\times m}$ is the $m \times m$ identity matrix. Therefore, the map

$$\begin{aligned} F : U \times \mathbb{R}^{n-m} \quad &\to \quad \mathbb{R}^n \\ (x^1, \dots, x^n) \quad &\mapsto \quad \hat{f}(x^1, \dots, x^m) + (0, \dots, 0, x^{m+1}, \dots, x^n), \end{aligned}$$

has derivative $(dF)_0 : \mathbb{R}^n \to \mathbb{R}^n$ given by the matrix

$$\left(\begin{array}{ccc} I_{m\times m} & | & 0 \\ --- & + & --- \\ 0 & | & I_{(n-m)\times(n-m)} \end{array} \right) = I_{n\times n}.$$

Applying the Inverse Function Theorem, we conclude that F is a local diffeomorphism at 0. This implies that $\psi \circ F$ is also a local diffeomorphism at 0, and so $\psi \circ F$ is another parametrization of N around q. Denoting the canonical immersion of $\mathbb{R}^m$ into $\mathbb{R}^n$ by j, we have $\hat{f} = F \circ j \Leftrightarrow f = \psi \circ F \circ j \circ \varphi^{-1}$, implying that the following diagram commutes:

$$\begin{array}{ccc} M \supset \varphi(\tilde{U}) & \xrightarrow{f} & (\psi \circ F)(\tilde{V}) \subset N \\ \varphi \uparrow & & \uparrow \psi \circ F \\ \mathbb{R}^m \supset \tilde{U} & \xrightarrow{j} & \tilde{V} \subset \mathbb{R}^n \end{array}$$

(for possibly smaller open sets $\tilde{U} \subset U$ and $\tilde{V} \subset V$). Hence, on these new coordinates, f is the canonical immersion. $\square$

Remark 5.2. As a consequence of the Local Immersion Theorem, any immersion at a point $p \in M$ is an immersion on a neighborhood of p.

When an immersion $f : M \to N$ is also a homeomorphism onto its image $f(M) \subset N$ with its subspace topology, it is called an **embedding**. We leave as an exercise to show that the Local Immersion Theorem implies that, locally, any immersion is an embedding.

Example 5.3.

(1) The map $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(t) = (t^2, t^3)$ is not an immersion at $t = 0$.
(2) The map $f : \mathbb{R} \to \mathbb{R}^2$ defined by $f(t) = (\cos t, \sin 2t)$ is an immersion but it is not an embedding (it is not injective).
(3) Let $g : \mathbb{R} \to \mathbb{R}$ be the function $g(t) = 2 \arctan(t) + \pi/2$. If f is the map from (2), $h := f \circ g$ is an injective immersion which is not an embedding. Indeed, the set $S = h(\mathbb{R})$ in Figure 13 is not the image of an embedding of $\mathbb{R}$ into $\mathbb{R}^2$. The arrows in the figure mean that the line approaches itself arbitrarily close at the origin but never self-intersects. If we consider the usual topologies on $\mathbb{R}$ and on $\mathbb{R}^2$, the image of an open set in $\mathbb{R}$ containing 0 is not an open set in $h(\mathbb{R})$ for the subspace topology, and so h^{-1} is not continuous.

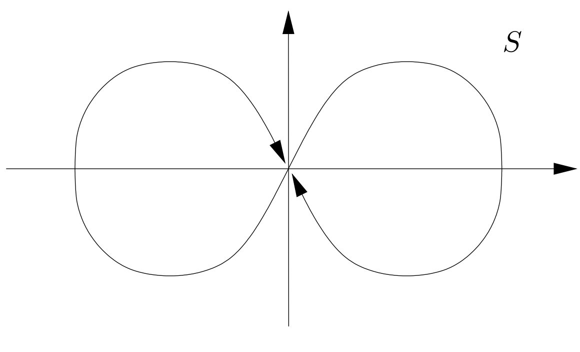

Figure 13

(4) The map $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(t) = (e^t \cos t, e^t \sin t)$ is an embedding of $\mathbb{R}$ into $\mathbb{R}^2$.

If $M \subset N$ and the inclusion map $i : M \hookrightarrow N$ is an embedding, M is said to be a **submanifold** of N. Therefore, an embedding $f : M \to N$ maps M diffeomorphically onto a submanifold of N. Charts on $f(M)$ are just restrictions of appropriately chosen charts on N to $f(M)$ (cf. Exercise 5.9.3).

A differentiable map $f : M \to N$ for which $(df)_p$ is surjective is called a **submersion at** p. Note that, in this case, we necessarily have $m \geq n$. If f is a submersion at every point in M it is called a **submersion**. Locally, every submersion is the standard projection of $\mathbb{R}^m$ onto the first n factors.

THEOREM 5.4. *Let $f : M \to N$ be a submersion at $p \in M$. Then there exist local coordinates around p and $f(p)$ for which f is the standard projection.*

PROOF. Let us consider parametrizations (U, φ) and (V, ψ) around p and $f(p)$, such that $f(\varphi(U)) \subset \psi(V)$, $\varphi(0) = p$ and $\psi(0) = f(p)$. In local coordinates f is given by $\hat{f} := \psi^{-1} \circ f \circ \varphi$ and, as $(df)_p$ is surjective, $(d\hat{f})_0 : \mathbb{R}^m \to \mathbb{R}^n$ is a surjective linear transformation. By a linear change of coordinates on $\mathbb{R}^n$ we may assume that $(d\hat{f})_0 = \left(\begin{array}{c|c} I_{n\times n} & * \end{array} \right)$. As in the proof of the Local Immersion Theorem, we will use an auxiliary map F that will allow us to use the Inverse Function Theorem,

$$\begin{array}{rcl} F : U \subset \mathbb{R}^m & \to & \mathbb{R}^m \\ (x^1, \dots, x^m) & \mapsto & \left(\hat{f}(x^1, \dots, x^m), x^{n+1}, \dots, x^m \right). \end{array}$$

Its derivative at 0 is the linear map given by

$$(dF)_0 = \left(\begin{array}{c|c} I_{n\times n} & * \\ \hline 0 & I_{(m-n)\times(m-n)} \end{array} \right).$$

The Inverse Function Theorem then implies that F is a local diffeomorphism at 0, meaning that it maps some open neighborhood of this point $\tilde{U} \subset U$, diffeomorphically onto an open set W of $\mathbb{R}^m$ containing 0. If $\pi_1 : \mathbb{R}^m \to \mathbb{R}^n$ is the standard projection onto the first n factors, we have $\pi_1 \circ F = \hat{f}$, and hence

$$\hat{f} \circ F^{-1} = \pi_1 : W \to \mathbb{R}^n.$$

Therefore, replacing φ by $\tilde{\varphi} := \varphi \circ F^{-1}$, we obtain coordinates for which f is the standard projection π_1 onto the first n factors:

$$\psi^{-1} \circ f \circ \tilde{\varphi} = \psi^{-1} \circ f \circ \varphi \circ F^{-1} = \hat{f} \circ F^{-1} = \pi_1.$$

$\square$

REMARK 5.5. This result is often stated together with the Local Immersion Theorem in what is known as the **Rank Theorem**.

Let $f : M \to N$ be a differentiable map between smooth manifolds of dimensions m and n, respectively. A point $q \in N$ is called a **regular value** of f if, for every $p \in f^{-1}(q)$, $(df)_p$ is surjective. If $p \in M$ is such that $(df)_p$ is not surjective it is called a **critical point** of f. The corresponding value $f(p)$ is called a **critical value**. Note that if there exists a regular value of f then $m \geq n$. We can obtain differentiable manifolds by taking inverse images of regular values.

THEOREM 5.6. *Let $q \in N$ be a regular value of $f : M \to N$ and assume that the level set $L = f^{-1}(q) = \{p \in M : f(p) = q\}$ is nonempty. Then L is a submanifold of M and $T_pL = \ker(df)_p \subset T_pM$ for all $p \in L$.*

PROOF. For each point $p \in f^{-1}(q)$, we choose parametrizations (U, φ) and (V, ψ) around p and q for which f is the standard projection π_1 onto the first n factors, $\varphi(0) = p$ and $\psi(0) = q$ (cf. Theorem 5.4). We then construct a differentiable structure for $L = f^{-1}(q)$ in the following way: take the sets U from each of these parametrizations of M; since $f \circ \varphi = \psi \circ \pi_1$, we have

$$\begin{aligned} \varphi^{-1}(f^{-1}(q)) &= \pi_1^{-1}(\psi^{-1}(q)) = \pi_1^{-1}(0) \\ &= \{(0, \ldots, 0, x^{n+1}, \ldots, x^m) \mid x^{n+1}, \ldots, x^m \in \mathbb{R}\}, \end{aligned}$$

and so

$$\widetilde{U} := \varphi^{-1}(L) = \{(x^1, \ldots, x^m) \in U : x^1 = \cdots = x^n = 0\};$$

hence, taking $\pi_2 : \mathbb{R}^m \to \mathbb{R}^{m-n}$ the standard projection onto the last $m - n$ factors and $j : \mathbb{R}^{m-n} \to \mathbb{R}^m$ the immersion given by

$$j(x^1, \ldots, x^{m-n}) = (0, \ldots, 0, x^1, \ldots, x^{m-n}),$$

the family $\{(\pi_2(\widetilde{U}), \varphi \circ j)\}$ is an atlas for L.

Moreover, the inclusion map $i : L \to M$ is an embedding. In fact, if A is an open set in L contained in a coordinate neighborhood then

$$A = \varphi\left((\mathbb{R}^n \times (\varphi \circ j)^{-1}(A)) \cap U\right) \cap L$$

is an open set for the subspace topology on L.

We will now show that $T_pL = \ker(df)_p$. For that, for each $v \in T_pL$, we consider a curve c on L such that $c(0) = p$ and $\dot{c}(0) = v$. Then $(f \circ c)(t) = q$ for every t and so

$$\frac{d}{dt}(f \circ c)(0) = 0 \Leftrightarrow (df)_p \, \dot{c}(0) = (df)_p \, v = 0,$$

implying that $v \in \ker(df)_p$. As $\dim T_pL = \dim(\ker(df)_p) = m - n$, the result follows. ☐

Given a differentiable manifold, we can ask ourselves if it can be embedded into $\mathbb{R}^K$ for some $K \in \mathbb{N}$. The following theorem, which was proved by Whitney in [**Whi44a, Whi44b**] answers this question and is known as the **Whitney Embedding Theorem**.

THEOREM 5.7. (Whitney) *Any differentiable manifold M of dimension n can be embedded in $\mathbb{R}^{2n}$ (and, provided that $n > 1$, immersed in $\mathbb{R}^{2n-1}$).*

REMARK 5.8. By the Whitney Embedding Theorem, any smooth manifold M^n is diffeomorphic to a submanifold of $\mathbb{R}^{2n}$.

EXERCISES 5.9.

(1) Show that any parametrization $\varphi : U \subset \mathbb{R}^m \to M$ is an embedding of U into M.
(2) Show that, locally, any immersion is an embedding, i.e., if $f : M \to N$ is an immersion and $p \in M$, then there is an open set $W \subset M$ containing p such that $f|_W$ is an embedding.
(3) Let N be a manifold and $M \subset N$. Show that M is a submanifold of N of dimension m if and only if, for each $p \in M$, there is a coordinate system $x : W \to \mathbb{R}^n$ around p on N, for which $M \cap W$ is defined by the equations $x^{m+1} = \cdots = x^n = 0$.
(4) Consider the sphere
$$S^n = \left\{ x \in \mathbb{R}^{n+1} : (x^1)^2 + \cdots (x^{n+1})^2 = 1 \right\}.$$
Show that S^n is an n-dimensional submanifold of $\mathbb{R}^{n+1}$ and
$$T_x S^n = \left\{ v \in \mathbb{R}^{n+1} : \langle x, v \rangle = 0 \right\},$$
where $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^n$.
(5) Let $f : M \to N$ be a differentiable map between smooth manifolds and let $V \subset M$, $W \subset N$ be submanifolds. If $f(V) \subset W$, show that $f : V \to W$ is also a differentiable map.
(6) Let $f : M \to N$ be an injective immersion. Show that if M is compact then $f(M)$ is a submanifold of N.

6. Vector Fields

A **vector field** on a smooth manifold M is a map that to each point $p \in M$ assigns a vector tangent to M at p:
$$\begin{aligned} X : M &\to TM \\ p &\mapsto X(p) := X_p \in T_pM. \end{aligned}$$
The vector field is said to be **differentiable** if this map is differentiable. The set of all differentiable vector fields on M is denoted by $\mathfrak{X}(M)$. Locally we have:

PROPOSITION 6.1. *Let W be a coordinate neighborhood on M (that is, $W = \varphi(U)$ for some parametrization $\varphi : U \to M$), and let $x := \varphi^{-1} : W \to \mathbb{R}^n$ be the corresponding coordinate chart. Then, a map $X : W \to TW$ is a differentiable vector field on W if and only if,*
$$X_p = X^1(p) \left(\frac{\partial}{\partial x^1} \right)_p + \cdots + X^n(p) \left(\frac{\partial}{\partial x^n} \right)_p$$
for some differentiable functions $X^i : W \to \mathbb{R}$ $(i = 1, \ldots, n)$.

PROOF. Let us consider the coordinate chart $x = (x^1, \dots, x^n)$. As $X_p \in T_pM$, we have

$$X_p = X^1(p)\left(\frac{\partial}{\partial x^1}\right)_p + \cdots + X^n(p)\left(\frac{\partial}{\partial x^n}\right)_p$$

for some functions $X^i : W \to \mathbb{R}$. In the local chart associated with the parametrization $(U \times \mathbb{R}^n, d\varphi)$ of TM, the local representation of the map X is

$$\hat{X}(x^1, \dots, x^n) = (x^1, \dots, x^n, \hat{X}^1(x^1, \dots, x^n), \dots, \hat{X}^n(x^1, \dots, x^n)).$$

Therefore X is differentiable if and only if the functions $\hat{X}^i : U \to \mathbb{R}$ are differentiable, i.e., if and only if the functions $X^i : W \to \mathbb{R}$ are differentiable. □

A vector field X is differentiable if and only if, given any differentiable function $f : M \to \mathbb{R}$, the function

$$\begin{array}{rcl} X \cdot f : M & \to & \mathbb{R} \\ p & \mapsto & X_p \cdot f := X_p(f) \end{array}$$

is also differentiable (cf. Exercise 6.11.1). This function $X \cdot f$ is called the **directional derivative** of f along X. Thus one can view $X \in \mathfrak{X}(M)$ as a linear operator $X : C^\infty(M) \to C^\infty(M)$.

Let us now take two vector fields $X, Y \in \mathfrak{X}(M)$. In general, the operators $X \circ Y$, $Y \circ X$ will involve derivatives of order two, and will not correspond to vector fields. However, the commutator $X \circ Y - Y \circ X$ does define a vector field.

PROPOSITION 6.2. *Given two differentiable vector fields $X, Y \in \mathfrak{X}(M)$ on a smooth manifold M, there exists a unique differentiable vector field $Z \in \mathfrak{X}(M)$ such that*

$$Z \cdot f = (X \circ Y - Y \circ X) \cdot f$$

for every differentiable function $f \in C^\infty(M)$.

PROOF. Considering a coordinate chart $x : W \subset M \to \mathbb{R}^n$, we have

$$X = \sum_{i=1}^n X^i \frac{\partial}{\partial x^i} \quad \text{and} \quad Y = \sum_{i=1}^n Y^i \frac{\partial}{\partial x^i}.$$

Then,

$$\begin{aligned}
&(X \circ Y - Y \circ X) \cdot f \\
&= X \cdot \left(\sum_{i=1}^n Y^i \frac{\partial \hat{f}}{\partial x^i} \right) - Y \cdot \left(\sum_{i=1}^n X^i \frac{\partial \hat{f}}{\partial x^i} \right) \\
&= \sum_{i=1}^n \left((X \cdot Y^i) \frac{\partial \hat{f}}{\partial x^i} - (Y \cdot X^i) \frac{\partial \hat{f}}{\partial x^i} \right) + \sum_{i,j=1}^n \left(X^j Y^i \frac{\partial^2 \hat{f}}{\partial x^j \partial x^i} - Y^j X^i \frac{\partial^2 \hat{f}}{\partial x^j \partial x^i} \right) \\
&= \left(\sum_{i=1}^n \left(X \cdot Y^i - Y \cdot X^i \right) \frac{\partial}{\partial x^i} \right) \cdot f,
\end{aligned}$$

and so, at each point $p \in W$, one has $(X \circ Y - Y \circ X)(f)(p) = Z_p \cdot f$, where

$$Z_p = \sum_{i=1}^n \left(X \cdot Y^i - Y \cdot X^i \right) \left(\frac{\partial}{\partial x^i} \right)_p .$$

Hence, the operator $X \circ Y - Y \circ X$ is a derivation at each point, and consequently defines a vector field. Note that this vector field is differentiable, as $(X \circ Y - Y \circ X) \cdot f$ is smooth for any smooth function $f : M \to \mathbb{R}$. □

The vector field Z is called the **Lie bracket** of X and Y, and is denoted by $[X, Y]$. In local coordinates it is given by

$$[X, Y] = \sum_{i=1}^n \left(X \cdot Y^i - Y \cdot X^i \right) \frac{\partial}{\partial x^i}. \tag{1}$$

We say that two vector fields $X, Y \in \mathfrak{X}(M)$ **commute** if $[X, Y] = 0$. We leave the proof of the following properties of the Lie bracket as an exercise.

PROPOSITION 6.3. *Given $X, Y, Z \in \mathfrak{X}(M)$, we have:*

(i) **Bilinearity:** *for any $\alpha, \beta \in \mathbb{R}$,*

$$\begin{aligned}
[\alpha X + \beta Y, Z] &= \alpha [X, Z] + \beta [Y, Z] \\
[X, \alpha Y + \beta Z] &= \alpha [X, Y] + \beta [X, Z];
\end{aligned}$$

(ii) **Antisymmetry:**

$$[X, Y] = -[Y, X];$$

(iii) **Jacobi identity:**

$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0;$$

(iv) **Leibniz Rule:** *For any $f, g \in C^\infty(M)$,*

$$[f\,X, g\,Y] = fg\,[X, Y] + f(X \cdot g)Y - g(Y \cdot f)X.$$

The space $\mathfrak{X}(M)$ of vector fields on M is a particular case of a **Lie algebra**:

Definition 6.4. *A vector space V equipped with an anti-symmetric bilinear map $[\cdot,\cdot] : V \times V \to V$ (called a* **Lie bracket***) satisfying the Jacobi identity is called a* **Lie algebra**. *A linear map $F : V \to W$ between Lie algebras is called a* **Lie algebra homomorphism** *if $F([v_1, v_2]) = [F(v_1), F(v_2)]$ for all $v_1, v_2 \in V$. If F is bijective then it is called a* **Lie algebra isomorphism**.

Given a vector field $X \in \mathfrak{X}(M)$ and a diffeomorphism $f : M \to N$ between smooth manifolds, we can naturally define a vector field on N using the derivative of f. This vector field, the **push-forward** of X, is denoted by f_*X and is defined in the following way: given $p \in M$,

$$(f_*X)_{f(p)} := (df)_p X_p.$$

This makes the following diagram commute:

$$\begin{array}{ccc} TM & \xrightarrow{df} & TN \\ X \uparrow & & \uparrow f_*X \\ M & \xrightarrow{f} & N \end{array}$$

Let us now turn to the definition of integral curve. If $X \in \mathfrak{X}(M)$ is a smooth vector field, an **integral curve** of X is a smooth curve $c : (-\varepsilon, \varepsilon) \to M$ such that $\dot{c}(t) = X_{c(t)}$. If this curve has initial value $c(0) = p$, we denote it by c_p and we say that c_p is an **integral curve of X at p**.

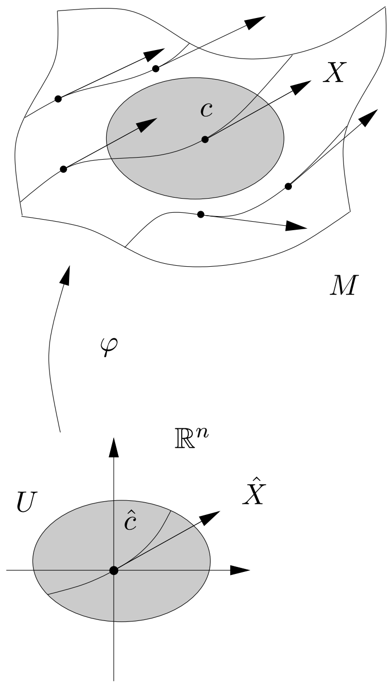

Figure 14. Integral curves of a vector field.

Considering a parametrization $\varphi : U \subset \mathbb{R}^n \to M$ on M, the integral curve c is locally given by $\hat{c} := \varphi^{-1} \circ c$. Applying $(d\varphi^{-1})_{c(t)}$ to both sides of the equation defining c, we obtain

$$\dot{\hat{c}}(t) = \hat{X}(\hat{c}(t)),$$

where $\hat{X} = d\varphi^{-1} \circ X \circ \varphi$ is the local representation of X with respect to the parametrizations (U, φ) and $(TU, d\varphi)$ on M and on TM (cf. Figure 14). This equation is just a system of n ordinary differential equations:

$$\frac{d\hat{c}^i}{dt}(t) = \hat{X}^i(\hat{c}(t)), \text{ for } i = 1, \dots, n. \tag{2}$$

The (local) existence and uniqueness of integral curves is then determined by the Picard-Lindelöf Theorem of ordinary differential equations (see for example [**Arn92**]), and we have

Theorem 6.5. *Let M be a smooth manifold and $X \in \mathfrak{X}(M)$ a smooth vector field on M. Given $p \in M$, there exists an integral curve $c_p : I \to M$ of X at p (that is, $\dot{c}_p(t) = X_{c_p(t)}$ for $t \in I = (-\varepsilon, \varepsilon)$ and $c_p(0) = p$). Moreover, this curve is unique, meaning that any two such curves agree on the intersection of their domains.*

This solution of (2) depends smoothly on the initial point p (see [**Arn92**]).

Theorem 6.6. *Indeed, if $y > 0$ the allowed interval for t decreases as b increases, if $y < 0$ the allowed interval for t decreases as a decreases, and if $y = 0$ this interval decrease as b increases or as a decreases. Let $X \in \mathfrak{X}(M)$. For each $p \in M$ there exists a neighborhood W of p, an interval $I = (-\varepsilon, \varepsilon)$ and a mapping $F : W \times I \to M$ such that:*

(i) for a fixed $q \in W$ the curve $F(q, t)$, $t \in I$, is an integral curve of X at q, that is, $F(q, 0) = q$ and $\frac{\partial F}{\partial t}(q, t) = X_{F(q,t)}$;
(ii) the map F is differentiable.

The map $F : W \times I \to M$ defined above is called the **local flow** of X at p. Let us now fix $t \in I$ and consider the map

$$\begin{array}{rcl} \psi_t : W & \to & M \\ q & \mapsto & F(q, t) = c_q(t). \end{array}$$

defined by the local flow. The following proposition then holds:

Proposition 6.7. *The maps $\psi_t : W \to M$ above are local diffeomorphisms and satisfy*

$$(\psi_t \circ \psi_s)(q) = \psi_{t+s}(q), \tag{3}$$

whenever $t, s, t + s \in I$ and $\psi_s(q) \in W$.

Proof. First we note that

$$\frac{dc_q}{dt}(t) = X_{c_q(t)}$$

and so

$$\frac{d}{dt}(c_q(t+s)) = X_{c_q(t+s)}.$$

Hence, as $c_q(t+s)|_{t=0} = c_q(s)$, the curve $c_{c_q(s)}(t)$ is just $c_q(t+s)$, that is, $\psi_{t+s}(q) = \psi_t(\psi_s(q))$. We can use this formula to extend ψ_t to $\psi_s(W)$ for all $s \in I$ such that $t+s \in I$. In particular, ψ_{-t} is well defined on $\psi_t(W)$, and $(\psi_{-t} \circ \psi_t)(q) = \psi_0(q) = c_q(0) = q$ for all $q \in W$. Thus the map ψ_{-t} is the inverse of ψ_t, which consequently is a local diffeomorphism (it maps W diffeomorphically onto its image). $\square$

A collection of diffeomorphisms $\{\psi_t : M \to M\}_{t\in I}$, where $I = (-\varepsilon, \varepsilon)$, satisfying (3) is called a **local 1-parameter group of diffeomorphisms**. When the interval of definition I of c_q is $\mathbb{R}$, this local 1-parameter group of diffeomorphisms becomes a **group of diffeomorphisms**. A vector field X whose local flow defines a 1-parameter group of diffeomorphisms is said to be **complete**. This happens for instance when the vector field X has compact support.

THEOREM 6.8. *If $X \in \mathfrak{X}(M)$ is a smooth vector field with compact support then it is complete.*

PROOF. For each $p \in M$ we can take a neighborhood W and an interval $I = (-\varepsilon, \varepsilon)$ such that the local flow of X at p, $F(q,t) = c_q(t)$, is defined on $W \times I$. We can therefore cover the support of X (which is compact) by a finite number of such neighborhoods W_k and consider an interval $I_0 = (-\varepsilon_0, \varepsilon_0)$ contained in the intersection of the corresponding intervals I_k. If q is not in $\operatorname{supp}(X)$, then $X_q = 0$ and so $c_q(t)$ is trivially defined on I_0. Hence we can extend the map F to $M \times I_0$. Moreover, condition (3) is true for each $-\varepsilon_0/2 < s, t < \varepsilon_0/2$, and we can again extend the map F, this time to $M \times \mathbb{R}$. In fact, for any $t \in \mathbb{R}$, we can write $t = k\varepsilon_0/2 + s$, where $k \in \mathbb{Z}$ and $0 \leq s < \varepsilon_0/2$, and define $F(q,t) := F^k(F(q,s), \varepsilon_0/2)$. Indeed, if $y > 0$ the allowed interval for t decreases as b increases, if $y < 0$ the allowed interval for t decreases as a decreases, and if $y = 0$ this interval decrease as b increases or as a decreases. $\square$

COROLLARY 6.9. *If M is compact then all smooth vector fields on M are complete.*

We finish this section with an important result, whose proof is left as an exercise (cf. Exercise 6.11.12).

THEOREM 6.10. *Let $X_1, X_2 \in \mathfrak{X}(M)$ be two complete vector fields. Then their flows ψ_1, ψ_2 commute (i.e., $\psi_{1,t} \circ \psi_{2,s} = \psi_{2,s} \circ \psi_{1,t}$ for all $s, t \in \mathbb{R}$) if and only if $[X_1, X_2] = 0$.*

EXERCISES 6.11.

(1) Let $X : M \to TM$ be a differentiable vector field on M and, for a smooth function $f : M \to \mathbb{R}$, consider its directional derivative

along X defined by

$$\begin{aligned} X\cdot f : M &\to \mathbb{R} \\ p &\mapsto X_p\cdot f. \end{aligned}$$

Show that:
(a) $(X\cdot f)(p) = (df)_p X_p$;
(b) the vector field X is smooth if and only if $X\cdot f$ is a differentiable function for any smooth function $f : M \to \mathbb{R}$;
(c) the directional derivative satisfies the following properties: for $f, g \in C^\infty(M)$ and $\alpha \in \mathbb{R}$,
(i) $X\cdot(f+g) = X\cdot f + X\cdot g$;
(ii) $X\cdot(\alpha f) = \alpha X\cdot f$;
(iii) $X\cdot(fg) = fX\cdot g + gX\cdot f$.

(2) Prove Proposition 6.3.
(3) Show that $(\mathbb{R}^3, \times)$ is a Lie algebra, where $\times$ is the cross product on $\mathbb{R}^3$.
(4) Let $X_1, X_2, X_3 \in \mathfrak{X}(\mathbb{R}^3)$ be the vector fields defined by

$$X_1 = y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}, \quad X_2 = z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}, \quad X_3 = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x},$$

where (x, y, z) are the usual Cartesian coordinates.
(a) Compute the Lie brackets $[X_i, X_j]$ for $i, j = 1, 2, 3$.
(b) Show that $\operatorname{span}\{X_1, X_2, X_3\}$ is a Lie subalgebra of $\mathfrak{X}(\mathbb{R}^3)$, isomorphic to $(\mathbb{R}^3, \times)$.
(c) Compute the flows $\psi_{1,t}, \psi_{2,t}, \psi_{3,t}$ of X_1, X_2, X_3.
(d) Show that $\psi_{i,\frac{\pi}{2}} \circ \psi_{j,\frac{\pi}{2}} \neq \psi_{j,\frac{\pi}{2}} \circ \psi_{i,\frac{\pi}{2}}$ for $i \neq j$.

(5) Give an example of a non complete vector field.
(6) Let N be a differentiable manifold, $M \subset N$ a submanifold and $X, Y \in \mathfrak{X}(N)$ vector fields tangent to M, i.e., such that $X_p, Y_p \in T_pM$ for all $p \in M$. Show that $[X, Y]$ is also tangent to M.
(7) Let $f : M \to N$ be a smooth map between manifolds. Two vector fields $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$ are said to be ***f*-related** (and we write $Y = f_*X$) if, for each $q \in N$ and $p \in f^{-1}(q) \subset M$, we have $(df)_p X_p = Y_q$. Show that:
(a) The vector field X is f-related to Y if and only if, for any differentiable function g defined on some open subset W of N, $(Y\cdot g)\circ f = X\cdot(g\circ f)$ on the inverse image $f^{-1}(W)$ of the domain of g;
(b) For differentiable maps $f : M \to N$ and $g : N \to P$ between smooth manifolds and vector fields $X \in \mathfrak{X}(M), Y \in \mathfrak{X}(N)$ and $Z \in \mathfrak{X}(P)$, if X is f-related to Y and Y is g-related to Z, then X is $(g\circ f)$-related to Z.

(8) Let $f : M \to N$ be a diffeomorphism between smooth manifolds. Show that $f_*[X, Y] = [f_*X, f_*Y]$ for every $X, Y \in \mathfrak{X}(M)$. Therefore, f_* induces a Lie algebra isomorphism between $\mathfrak{X}(M)$ and $\mathfrak{X}(N)$.

(9) Let $f : M \to N$ be a differentiable map between smooth manifolds and consider two vector fields $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$. Show that:
   (a) if the vector field Y is f-related to X then any integral curve of X is mapped by f into an integral curve of Y;
   (b) the vector field Y is f-related to X if and only if the local flows F_X and F_Y satisfy $f(F_X(p, t)) = F_Y(f(p), t)$ for all (t, p) for which both sides are defined.

(10) (*Lie derivative of a function*) Given a vector field $X \in \mathfrak{X}(M)$, we define the **Lie derivative** of a smooth function $f : M \to \mathbb{R}$ in the direction of X as

$$L_X f(p) := \frac{d}{dt}((f \circ \psi_t)(p))_{|t=0},$$

where $\psi_t = F(\cdot, t)$, for F the local flow of X at p. Show that $L_X f = X \cdot f$, meaning that the Lie derivative of f in the direction of X is just the directional derivative of f along X.

(11) (*Lie derivative of a vector field*) For two vector fields $X, Y \in \mathfrak{X}(M)$ we define the **Lie derivative** of Y in the direction of X as,

$$L_X Y := \frac{d}{dt}((\psi_{-t})_* Y)_{|t=0},$$

where $\{\psi_t\}_{t \in I}$ is the local flow of X. Show that:
   (a) $L_X Y = [X, Y]$;
   (b) $L_X[Y, Z] = [L_X Y, Z] + [Y, L_X Z]$, for $X, Y, Z \in \mathfrak{X}(M)$;
   (c) $L_X \circ L_Y - L_Y \circ L_X = L_{[X,Y]}$.

(12) Let $X, Y \in \mathfrak{X}(M)$ be two complete vector fields with flows ψ, ϕ. Show that:
   (a) given a diffeomorphism $f : M \to M$, we have $f_* X = X$ if and only if $f \circ \psi_t = \psi_t \circ f$ for all $t \in \mathbb{R}$;
   (b) $\psi_t \circ \phi_s = \phi_s \circ \psi_t$ for all $s, t \in \mathbb{R}$ if and only if $[X, Y] = 0$.

7. Lie Groups

A **Lie group** G is a smooth manifold which is at the same time a group, in such a way that the group operations

$$\begin{array}{ccc} G \times G & \to & G \\ (g, h) & \mapsto & gh \end{array} \quad \text{and} \quad \begin{array}{ccc} G & \to & G \\ g & \mapsto & g^{-1} \end{array}$$

are differentiable maps (where we consider the standard differentiable structure of the product on $G \times G$).

EXAMPLE 7.1.

(1) $(\mathbb{R}^n, +)$ is trivially an abelian Lie group
(2) The **general linear group**

$$GL(n) = \{n \times n \text{ invertible real matrices}\}$$

is the most basic example of a nontrivial Lie group. We have seen in Example 2.3.4 that it is a smooth manifold of dimension n^2. Moreover, the group multiplication is just the restriction to

$$GL(n) \times GL(n)$$

of the usual multiplication of $n \times n$ matrices, whose coordinate functions are quadratic polynomials; the inversion is just the restriction to $GL(n)$ of the usual inversion of nonsingular matrices which, by Cramer's rule, is a map with rational coordinate functions with nonzero denominators (only the determinant appears on the denominator).
(3) The **orthogonal group**

$$O(n) = \{A \in \mathcal{M}_{n\times n} \mid A^t A = I\}$$

of orthogonal transformations of $\mathbb{R}^n$ is also a Lie group. We can show this by considering the map $f : A \mapsto A^t A$ from $\mathcal{M}_{n\times n} \cong \mathbb{R}^{n^2}$ to the space $\mathcal{S}_{n\times n} \cong \mathbb{R}^{\frac{1}{2}n(n+1)}$ of symmetric $n \times n$ matrices. Its derivative at a point $A \in O(n)$, $(df)_A$, is a surjective map from $T_A\mathcal{M}_{n\times n} \cong \mathcal{M}_{n\times n}$ onto $T_{f(A)}\mathcal{S}_{n\times n} \cong \mathcal{S}_{n\times n}$. Indeed,

$$\begin{aligned}
(df)_A(B) &= \lim_{h\to 0} \frac{f(A+hB) - f(A)}{h} \\
&= \lim_{h\to 0} \frac{(A+hB)^t(A+hB) - A^tA}{h} \\
&= B^tA + A^tB,
\end{aligned}$$

and any symmetric matrix S can be written as $B^tA + A^tB$ with $B = \frac{1}{2}(A^{-1})^tS = \frac{1}{2}AS$. In particular, the identity I is a regular value of f and so, by Theorem 5.6, we have that $O(n) = f^{-1}(I)$ is a submanifold of $\mathcal{M}_{n\times n}$ of dimension $\frac{1}{2}n(n-1)$. Moreover, it is also a Lie group as the group multiplication and inversion are restrictions of the same operations on $GL(n)$ to $O(n)$ (a submanifold) and have values on $O(n)$ (cf. Exercise 5.9.5).
(4) The map $f : GL(n) \to \mathbb{R}$ given by $f(A) = \det A$ is differentiable, and the level set $f^{-1}(1)$ is

$$SL(n) = \{A \in \mathcal{M}_{n\times n} \mid \det A = 1\},$$

the **special linear group**. Again, the derivative of f is surjective at a point $A \in GL(n)$, making $SL(n)$ into a Lie group. Indeed, it is easy to see that

$$(df)_I(B) = \lim_{h\to 0} \frac{\det(I+hB) - \det I}{h} = \operatorname{tr} B$$

implying that

$$\begin{aligned}(df)_A(B) &= \lim_{h\to 0}\frac{\det(A+hB)-\det A}{h}\\ &= \lim_{h\to 0}\frac{(\det A)\det(I+hA^{-1}B)-\det A}{h}\\ &= (\det A)\lim_{h\to 0}\frac{\det(I+hA^{-1}B)-1}{h}\\ &= (\det A)\,(df)_I(A^{-1}B)=(\det A)\operatorname{tr}(A^{-1}B).\end{aligned}$$

Since $\det(A)=1$, for any $k\in\mathbb{R}$, we can take the matrix $B=\frac{k}{n}A$ to obtain $(df)_A(B)=\operatorname{tr}\left(\frac{k}{n}I\right)=k$. Therefore, $(df)_A$ is surjective for every $A\in SL(n)$, and so 1 is a regular value of f. Consequently, $SL(n)$ is a submanifold of $GL(n)$. As in the preceding example, the group multiplication and inversion are differentiable, and so $SL(n)$ is a Lie group.

(5) The map $A\mapsto \det A$ is a differentiable map from $O(n)$ to $\{-1,1\}$, and the level set $f^{-1}(1)$ is

$$SO(n)=\{A\in O(n)\mid \det A=1\},$$

the **special orthogonal group** or the **rotation group** in $\mathbb{R}^n$, which is then an open subset of $O(n)$, and therefore a Lie group of the same dimension.

(6) We can also consider the space $\mathcal{M}_{n\times n}(\mathbb{C})$ of complex $n\times n$ matrices, and the space $GL(n,\mathbb{C})$ of complex $n\times n$ invertible matrices. This is a Lie group of real dimension $2n^2$. Moreover, similarly to what was done above for $O(n)$, we can take the group of unitary transformations on $\mathbb{C}^n$,

$$U(n)=\{A\in\mathcal{M}_{n\times n}(\mathbb{C})\mid A^*A=I\},$$

where A^* is the adjoint of A. This group is a submanifold of $\mathcal{M}_{n\times n}(\mathbb{C})\cong\mathbb{C}^{n^2}\cong\mathbb{R}^{2n^2}$, and a Lie group, called the **unitary group**. This can be seen from the fact that I is a regular value of the map $f:A\mapsto A^*A$ from $\mathcal{M}_{n\times n}(\mathbb{C})$ to the space of selfadjoint matrices. As any element of $\mathcal{M}_{n\times n}(\mathbb{C})$ can be uniquely written as a sum of a selfadjoint with an anti-selfadjoint matrix, and the map $A\to iA$ is an isomorphism from the space of selfadjoint matrices to the space of anti-selfadjoint matrices, we conclude that these two spaces have real dimension $\frac{1}{2}\dim_{\mathbb{R}}\mathcal{M}_{n\times n}(\mathbb{C})=n^2$. Hence, $\dim U(n)=n^2$.

(7) The **special unitary group**

$$SU(n)=\{A\in U(n)\mid \det A=1\},$$

is also a Lie group now of dimension n^2-1 (note that $A\mapsto\det(A)$ is now a differentiable map from $U(n)$ to S^1).

As a Lie group G is, by definition, a manifold, we can consider the tangent space at one of its points. In particular, the tangent space at the identity e is usually denoted by

$$\mathfrak{g} := T_e G.$$

For $g \in G$, we have the maps

$$\begin{array}{rcl} L_g : G & \to & G \\ h & \mapsto & g \cdot h \end{array} \quad \text{and} \quad \begin{array}{rcl} R_g : G & \to & G \\ h & \mapsto & h \cdot g \end{array}$$

which correspond to **left multiplication** and **right multiplication**.

A vector field on G is called **left invariant** if $(L_g)_* X = X$ for every $g \in G$, that is,

$$((L_g)_* X)_{gh} = X_{gh} \text{ or } (dL_g)_h X_h = X_{gh},$$

for every $g, h \in G$. There is, of course, a vector space isomorphism between $\mathfrak{g}$ and the space of left invariant vector fields on G that, to each $V \in \mathfrak{g}$, assigns the vector field X^V defined by

$$X^V_g := (dL_g)_e V,$$

for any $g \in G$. This vector field is left invariant as

$$(dL_g)_h X^V_h = (dL_g)_h (dL_h)_e V = (d(L_g \circ L_h))_e V = (dL_{gh})_e V = X^V_{gh}.$$

Note that, given a left invariant vector field X, the corresponding element of $\mathfrak{g}$ is X_e. As the space $\mathfrak{X}_L(G)$ of left invariant vector fields is closed under the Lie bracket of vector fields (because, from Exercise 6.11.8, $(L_g)_*[X,Y] = [(L_g)_* X, (L_g)_* Y]$), it has a structure of Lie subalgebra of the Lie algebra of vector fields (see Definition 6.4). The isomorphism $\mathfrak{X}_L(G) \cong \mathfrak{g}$ then determines a Lie algebra structure on $\mathfrak{g}$. We call $\mathfrak{g}$ the **Lie algebra** of the Lie group G.

Example 7.2.

(1) If $G = GL(n)$, then $\mathfrak{gl}(n) = T_I GL(n) = \mathcal{M}_{n \times n}$ is the space of $n \times n$ matrices with real coefficients, and the Lie bracket on $\mathfrak{gl}(n)$ is the commutator of matrices

$$[A, B] = AB - BA.$$

In fact, if $A, B \in \mathfrak{gl}(n)$ are two $n \times n$ matrices, the corresponding left invariant vector fields are given by

$$\begin{aligned} X^A_g &= (dL_g)_I (A) = \sum_{i,k,j} x^{ik} a^{kj} \frac{\partial}{\partial x^{ij}} \\ X^B_g &= (dL_g)_I (B) = \sum_{i,k,j} x^{ik} b^{kj} \frac{\partial}{\partial x^{ij}}, \end{aligned}$$

where $g \in GL(n)$ is a matrix with components x^{ij}. The ij-component of $[X^A, X^B]_g$ is given by $X^A_g \cdot (X^B)^{ij} - X^B_g \cdot (X^A)^{ij}$, i.e.

$$\begin{aligned}
[X^A, X^B]^{ij}(g) &= \left(\sum_{l,m,p} x^{lp}a^{pm}\frac{\partial}{\partial x^{lm}}\right)\left(\sum_k x^{ik}b^{kj}\right) - \\
&\quad - \left(\sum_{l,m,p} x^{lp}b^{pm}\frac{\partial}{\partial x^{lm}}\right)\left(\sum_k x^{ik}a^{kj}\right) \\
&= \sum_{k,l,m,p} x^{lp}a^{pm}\delta_{il}\delta_{km}b^{kj} - \sum_{k,l,m,p} x^{lp}b^{pm}\delta_{il}\delta_{km}a^{kj} \\
&= \sum_{m,p} x^{ip}(a^{pm}b^{mj} - b^{pm}a^{mj}) \\
&= \sum_p x^{ip}(AB - BA)^{pj}
\end{aligned}$$

(where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$ is the **Kronecker symbol**). Making $g = I$, we obtain $[A, B] = [X^A, X^B]_I = AB - BA$. This will always be the case when G is a matrix group, that is, when G is a subgroup of $GL(n)$ for some n.

(2) If $G = O(n)$ then its Lie algebra is

$$\mathfrak{o}(n) = \{A \in \mathcal{M}_{n\times n} \mid A^t + A = 0\}.$$

In fact, we have seen in Example 7.1.3 that $O(n) = f^{-1}(I)$ where the identity I is a regular value of the map

$$\begin{aligned}
f : \mathcal{M}_{n\times n} &\to \mathcal{S}_{n\times n} \\
A &\mapsto A^tA.
\end{aligned}$$

Hence, $\mathfrak{o}(n) = T_IG = \ker(df)_I = \{A \in \mathcal{M}_{n\times n} \mid A^t + A = 0\}$ is the space of skew-symmetric matrices.

(3) If $G = SO(n) = \{A \in O(n) \mid \det A = 1\}$, then its Lie algebra is

$$\mathfrak{so}(n) = T_ISO(n) = T_IO(n) = \mathfrak{o}(n).$$

(4) Similarly to Example 7.2.2, the Lie algebra of $U(n)$ is

$$\mathfrak{u}(n) = \{A \in \mathcal{M}_{n\times n}(\mathbb{C}) \mid A^* + A = 0\},$$

the space of skew-hermitian matrices. To determine the Lie algebra of $SU(n)$, we see that $SU(n)$ is the level set $f^{-1}(1)$, where $f(A) = \det A$, and so

$$\mathfrak{su}(n) = \ker(df)_I = \{A \in \mathfrak{u}(n) \mid \operatorname{tr}(A) = 0\}.$$

We now study the flow of a left invariant vector field.

Proposition 7.3. *Let F be the local flow of a left invariant vector field X at a point $h \in G$. Then the map ψ_t defined by F (that is, $\psi_t(g) = F(g, t)$) satisfies $\psi_t = R_{\psi_t(e)}$. Moreover, the flow of X is globally defined for all $t \in \mathbb{R}$.*

PROOF. For $g \in G$, $R_{\psi_t(e)}(g) = g \cdot \psi_t(e) = L_g(\psi_t(e))$. Hence,

$$R_{\psi_0(e)}(g) = g \cdot e = g$$

and

$$\begin{aligned}
\frac{d}{dt}\left(R_{\psi_t(e)}(g)\right) &= \frac{d}{dt}\left(L_g(\psi_t(e))\right) = (dL_g)_{\psi_t(e)}\left(\frac{d}{dt}(\psi_t(e))\right) \\
&= (dL_g)_{\psi_t(e)}\left(X_{\psi_t(e)}\right) = X_{g\cdot\psi_t(e)} \\
&= X_{R_{\psi_t(e)}(g)},
\end{aligned}$$

implying that $R_{\psi_t(e)}(g) = c_g(t) = \psi_t(g)$ is the integral curve of X at g. Consequently, if $\psi_t(e)$ is defined for $t \in (-\varepsilon, \varepsilon)$, then $\psi_t(g)$ is defined for $t \in (-\varepsilon, \varepsilon)$ and $g \in G$. Moreover, condition (3) in Section 6 is true for each $-\varepsilon/2 < s, t < \varepsilon/2$ and we can extend the map F to $G \times \mathbb{R}$ as before: for any $t \in \mathbb{R}$, we write $t = k\varepsilon/2 + s$ where $k \in \mathbb{Z}$ and $0 \leq s < \varepsilon/2$, and define $F(g,t) := F^k(F(g,s), \varepsilon/2) = gF(e,s)F^k(e,\varepsilon/2)$. □

REMARK 7.4. A homomorphism $F : G_1 \to G_2$ between Lie groups is called a **Lie group homomorphism** if, besides being a group homomorphism, it is also a differentiable map. Since

$$\psi_{t+s}(e) = \psi_s(\psi_t(e)) = R_{\psi_s(e)}\psi_t(e) = \psi_t(e) \cdot \psi_s(e),$$

the integral curve $t \mapsto \psi_t(e)$ defines a group homomorphism between $(\mathbb{R}, +)$ and $(G, \cdot)$.

DEFINITION 7.5. *The* **exponential map** $\exp : \mathfrak{g} \to G$ *is the map that, to each* $V \in \mathfrak{g}$*, assigns the value* $\psi_1(e)$*, where* ψ_t *is the flow of the left-invariant vector field* X^V.

REMARK 7.6. If $c_g(t)$ is the integral curve of X at g and $s \in \mathbb{R}$, it is easy to check that $c_g(st)$ is the integral curve of sX at g. On the other hand, for $V \in \mathfrak{g}$ one has $X^{sV} = sX^V$. Consequently,

$$\psi_t(e) = c_e(t) = c_e(t \cdot 1) = F(e, 1) = \exp(tV),$$

where F is the flow of $tX^V = X^{tV}$.

EXAMPLE 7.7. If G is a group of matrices, then for $A \in \mathfrak{g}$,

$$\exp A = e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

In fact, this series converges for any matrix A and the map $h(t) = e^{At}$ satisfies

$$\begin{aligned}
h(0) &= e^0 = I \\
\frac{dh}{dt}(t) &= e^{At}A = h(t)A.
\end{aligned}$$

Hence, h is the flow of X^A at the identity (that is, $h(t) = \psi_t(e)$), and so $\exp A = \psi_1(e) = e^A$.

Let now G be any group and M be any set. We say that G **acts on** M if there is a homomorphism ϕ from G to the group of bijective mappings from M to M, or, equivalently, writing

$$\phi(g)(p) = A(g, p),$$

if there is a mapping $A : G \times M \to M$ satisfying the following conditions:

(i) if e is the identity in G, then $A(e, p) = p$, $\forall p \in M$;
(ii) if $g, h \in G$, then $A(g, A(h, p)) = A(gh, p)$, $\forall p \in M$.

Usually we denote $A(g, p)$ by $g \cdot p$.

EXAMPLE 7.8.

(1) Let G be a group and $H \subset G$ a subgroup. Then H acts on G by left multiplication: $A(h, g) = h \cdot g$ for $h \in H$, $g \in G$.
(2) $GL(n)$ acts on $\mathbb{R}^n$ through $A \cdot x = Ax$ for $A \in GL(n)$ and $x \in \mathbb{R}^n$. The same is true for any subgroup $G \subset GL(n)$.

For each $p \in M$ we can define the **orbit of p** as the set $G \cdot p := \{g \cdot p \mid g \in G\}$. If $G \cdot p = \{p\}$ then p is called a **fixed point** of G. If there is a point $p \in M$ whose orbit is all of M (i.e. $G \cdot p = M$), then the action is said to be **transitive**. Note that when this happens, there is only one orbit and, for every $p, q \in M$ with $p \neq q$, there is always an element of the group $g \in G$ such that $q = g \cdot p$. The manifold M is then called a **homogeneous space** of G. The **stabilizer** (or **isotropy subgroup**) of a point $p \in M$ is the group

$$G_p = \{g \in G \mid g \cdot p = p\}.$$

The action is called **free** if all the stabilizers are trivial.

If G is a Lie group and M is a smooth manifold, we say that the action is **smooth** if the map $A : G \times M \to M$ is differentiable. In this case, the map $p \mapsto g \cdot p$ is a diffeomorphism. We will always assume the action of a Lie group on a differentiable manifold to be smooth. A smooth action is said to be **proper** if the map

$$\begin{array}{ccc} G \times M & \to & M \times M \\ (g, p) & \mapsto & (g \cdot p, p) \end{array}$$

is proper (recall that a map is called proper if the preimage of any compact set is compact).

REMARK 7.9. Note that a smooth action is proper if and only if, given two convergent sequences $\{p_n\}$ and $\{g_n \cdot p_n\}$ in M, there exists a convergent subsequence $\{g_{n_k}\}$ in G. If G is compact this condition is always satisfied.

PROPOSITION 7.10. *If the action of a Lie group G on a differentiable manifold M is proper, then the orbit space $M/\sim$ is a Hausdorff space.*

PROOF. The relation $p \sim q \Leftrightarrow q \in G \cdot p$ is an open equivalence relation. Indeed, since $p \mapsto g \cdot p$ is a homeomorphism, the set $[U] = \{g \cdot p \mid p \in$

U and $g \in G\} = \bigcup_{g \in G} g \cdot U$ is an open subset of M for any open set U in M. Therefore we just have to show that the set

$$R = \{(p, q) \in M \times M \mid p \sim q\}$$

is closed (cf. Proposition 10.2). This follows from the fact that R is the image of the map

$$\begin{array}{rcl} G \times M & \to & M \times M \\ (g, p) & \mapsto & (g \cdot p, p) \end{array}$$

which is continuous and proper, hence closed. $\square$

Under certain conditions the orbit space M/G is naturally a differentiable manifold.

THEOREM 7.11. *Let M be a differentiable manifold equipped with a free proper action of a Lie group G. Then the orbit space M/G is naturally a differentiable manifold of dimension* $\dim M - \dim G$, *and the quotient map $\pi : M \to M/G$ is a submersion.*

PROOF. By the previous proposition, the quotient M/G is Hausdorff. Moreover, this quotient satisfies the second countability axiom because M does so and the equivalence relation defined by G is open. It remains to be shown that M/G has a natural differentiable structure for which the quotient map is a submersion. We do this only in the case of a discrete Lie group.

In this case, we just have to prove that for each point $p \in M$ there exists a neighborhood $U \ni p$ such that $g \cdot U \cap h \cdot U = \varnothing$ for $g \neq h$. This guarantees that each point $[p] \in M/G$ has a neighborhood $[U]$ homeomorphic to U, which we can assume to be a coordinate neighborhood. Since G acts by diffeomorphisms, the differentiable structure defined in this way does not depend on the choice of $p \in [p]$. Since the charts of M/G are obtained from charts of M, the overlap maps are smooth. Therefore M/G has a natural differentiable structure for which $\pi : M \to M/G$ is a local diffeomorphism (as the coordinate expression of $\pi|_U : U \to [U]$ is the identity map).

Showing that $g \cdot U \cap h \cdot U = \varnothing$ for $g \neq h$ is equivalent to showing that $g \cdot U \cap U = \varnothing$ for $g \neq e$. Assume that this did not happen for any neighborhood $U \ni p$. Then there would exist a sequence of open sets $U_n \ni p$ with $U_{n+1} \subset U_n$, $\bigcap_{n=1}^{+\infty} U_n = \{p\}$ and a sequence $g_n \in G \setminus \{e\}$ asuch that $g_n \cdot U_n \cap U_n \neq \varnothing$. Choose $p_n \in g_n \cdot U_n$. Then $p_n = g_n \cdot q_n$ for some $q_n \in U_n$. We have $p_n \to p$ and $q_n \to p$. Since the action is proper, g_n admits a convergent subsequence g_{n_k}. Let g be its limit. Making $k \to +\infty$ in $q_{n_k} = g_{n_k} \cdot p_{n_k}$ yields $g \cdot p = p$, implying that $g = e$ (the action is free). Because G is discrete, we would then have $g_{n_k} = e$ for sufficiently large k, which is a contradiction. $\square$

EXAMPLE 7.12.

(1) Let $S^n = \{x \in \mathbb{R}^{n+1} \mid \sum_{i=1}^n (x^i)^2 = 1\}$ be equipped with the action of $G = \mathbb{Z}_2 = \{-I, I\}$ given by $-I \cdot x = -x$ (antipodal map). This action is proper and free, and so the orbit space S^n/G is an n-dimensional manifold. This space is the real projective space $\mathbb{R}P^n$ (cf. Exercise 2.5.8).
(2) The group $G = \mathbb{R} \setminus \{0\}$ acts on $M = \mathbb{R}^{n+1} \setminus \{0\}$ by multiplication: $t \cdot x = tx$. This action is proper and free, and so M/G is a differentiable manifold of dimension n (which is again $\mathbb{R}P^n$).
(3) Consider $M = \mathbb{R}^n$ equipped with an action of $G = \mathbb{Z}^n$ defined by:
$$(k^1, \dots, k^n) \cdot (x^1, \dots, x^n) = (x^1 + k^1, \dots, x^n + k^n).$$
This action is proper and free, and so the quotient M/G is a manifold of dimension n. This space with the quotient differentiable structure defined in Theorem 7.11 is called the n-**torus** and is denoted by $\mathbb{T}^n$. It is diffeomorphic to the product manifold $S^1 \times \cdots \times S^1$ and, when $n = 2$, is diffeomorphic to the torus of revolution in $\mathbb{R}^3$.

Quotients by dicrete group actions determine **coverings** of manifolds.

Definition 7.13. *A smooth* **covering** *of a differentiable manifold B is a pair (M, π), where M is a connected differentiable manifold, $\pi : M \to B$ is a surjective local diffeomorphism, and, for each $p \in B$, there exists a connected neighborhood U of p in B such that $\pi^{-1}(U)$ is the union of disjoint open sets $U_\alpha \subset M$ (called* **slices***), and the restrictions π_α of π to U_α are diffeomorphisms onto U. The map π is called a* **covering map** *and M is called a* **covering manifold**.

Remark 7.14.
(1) It is clear that we must have $\dim M = \dim B$.
(2) Note that the collection of mutually disjoint open sets $\{U_\alpha\}$ must be countable (M has a countable basis).
(3) The **fibers** $\pi^{-1}(p) \subset M$ have the discrete topology. Indeed, as each slice U_α is open and intersects $\pi^{-1}(p)$ in exactly one point, this point is open in the subspace topology.

Example 7.15.
(1) The map $\pi : \mathbb{R} \to S^1$ given by
$$\pi(t) = (\cos(2\pi t), \sin(2\pi t))$$
is a smooth covering of S^1. However the restriction of this map to $(0, +\infty)$ is a surjective local diffeomorphism which is not a covering map.
(2) The product of covering maps is clearly a covering map. Thus we can generalize the above example and obtain a covering of $\mathbb{T}^n \cong S^1 \times \cdots \times S^1$ by $\mathbb{R}^n$.
(3) In Example 7.12.1 we have a covering of $\mathbb{R}P^n$ by S^n.

A diffeomorphism $h : M \to M$, where M is a covering manifold, is called a **deck transformation** (or **covering transformation**) if $\pi \circ h = \pi$, or, equivalently, if each set $\pi^{-1}(p)$ is carried to itself by h. It can be shown that the group G of all covering transformations is a discrete Lie group whose action on M is free and proper.

If the covering manifold M is **simply connected** (cf. Section 10.5), the covering is said to be a **universal covering**. In this case, B is diffeomorphic to M/G. Moreover, G is isomorphic to the **fundamental group** $\pi_1(B)$ of B (cf. Section 10.5).

Lie's Theorem states that for a given Lie algebra $\mathfrak{g}$ there exists a unique simply connected Lie group $\widetilde{G}$ whose Lie algebra is $\mathfrak{g}$. If a Lie group G also has $\mathfrak{g}$ as its Lie algebra, then there exists a unique Lie group homomorphism $\pi : \widetilde{G} \to G$ which is a covering map. The group of deck transformations is, in this case, simply $\ker(\pi)$, and hence G is diffeomorphic to $\widetilde{G}/\ker(\pi)$. In fact, G is also isomorphic to $\widetilde{G}/\ker(\pi)$, which has a natural group structure ($\ker(\pi)$ is a normal subgroup).

EXAMPLE 7.16.

(1) In the universal covering of S^1 of Example 7.15.1 the deck transformations are translations $h_k : t \mapsto t + k$ by an integer k, and so the fundamental group of S^1 is $\mathbb{Z}$.
(2) Similarly, the deck transformations of the universal covering of $\mathbb{T}^n$ are translations by integer vectors (cf. Example 7.15.2), and so the fundamental group of $\mathbb{T}^n$ is $\mathbb{Z}^n$.
(3) In the universal covering of $\mathbb{R}P^n$ from Example 7.15.3, the only deck transformations are the identity and the antipodal map, and so the fundamental group of $\mathbb{R}P^n$ is $\mathbb{Z}_2$.

EXERCISES 7.17.

(1) (a) Given two Lie groups G_1, G_2, show that $G_1 \times G_2$ (the direct product of the two groups) is a Lie group with the standard differentiable structure on the product.
    (b) The circle S^1 can be identified with the subset of complex numbers of absolute value 1. Show that S^1 is a Lie group and conclude that the n-torus $T^n \cong S^1 \times \ldots \times S^1$ is also a Lie group.
(2) (a) Show that $(\mathbb{R}^n, +)$ is a Lie group, determine its Lie algebra and write an expression for the exponential map.
    (b) Prove that, if G is an abelian Lie group, then $[V, W] = 0$ for all $V, W \in \mathfrak{g}$.
(3) We can identify each point in
$$H = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$$
with an invertible affine map $h : \mathbb{R} \to \mathbb{R}$ given by $h(t) = yt + x$. The set of all such maps is a group under composition; consequently, our identification induces a group structure on H.

(a) Show that the induced group operation is given by

$$(x,y)\cdot(z,w)=(yz+x,yw),$$

and that H, with this group operation, is a Lie group.

(b) Show that the derivative of the left translation map $L_{(x,y)}: H\to H$ at point $(z,w)\in H$ is represented in the above coordinates by the matrix

$$\left(dL_{(x,y)}\right)_{(z,w)}=\begin{pmatrix} y & 0\\ 0 & y\end{pmatrix}.$$

Conclude that the left-invariant vector field $X^V\in\mathfrak{X}(H)$ determined by the vector

$$V=\xi\frac{\partial}{\partial x}+\eta\frac{\partial}{\partial y}\in\mathfrak{h}\equiv T_{(0,1)}H \qquad (\xi,\eta\in\mathbb{R})$$

is given by

$$X^V_{(x,y)}=\xi y\frac{\partial}{\partial x}+\eta y\frac{\partial}{\partial y}.$$

(c) Given $V,W\in\mathfrak{h}$, compute $[V,W]$.

(d) Determine the flow of the vector field X^V, and give an expression for the exponential map $\exp:\mathfrak{h}\to H$.

(e) Confirm your results by first showing that H is the subgroup of $GL(2)$ formed by matrices

$$\begin{pmatrix} y & x\\ 0 & 1\end{pmatrix}$$

with $y>0$.

(4) Consider the group

$$SL(2)=\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix} : ad-bc=1\right\},$$

which we already know to be a 3-manifold. Making

$$a=p+q,\quad d=p-q,\quad b=r+s,\quad c=r-s,$$

show that $SL(2)$ is diffeomorphic to $S^1\times\mathbb{R}^2$.

(5) For $A\in\mathfrak{gl}(n)$, consider the differentiable map

$$\begin{array}{rcl} h:\mathbb{R} & \to & \mathbb{R}\backslash\{0\}\\ t & \mapsto & \det e^{At}\end{array}$$

and show that:

(a) this map is a group homomorphism between $(\mathbb{R},+)$ and $(\mathbb{R}\backslash\{0\},\cdot)$;

(b) $h'(0)=\operatorname{tr}A$;

(c) $\det(e^A)=e^{\operatorname{tr}A}$.

(6) (a) If $A \in \mathfrak{sl}(2)$, show that there is a $\lambda \in \mathbb{R} \cup i\mathbb{R}$ such that

$$e^A = \cosh\lambda \, I + \frac{\sinh\lambda}{\lambda} A.$$

(b) Show that $\exp : \mathfrak{sl}(2) \to SL(2)$ is not surjective.

(7) Consider the vector field $X \in \mathfrak{X}(\mathbb{R}^2)$ defined by

$$X = \sqrt{x^2 + y^2}\,\frac{\partial}{\partial x}.$$

(a) Show that the flow of X defines a free action of $\mathbb{R}$ on $M = \mathbb{R}^2 \setminus \{0\}$.

(b) Describe the topological quotient space $M/\mathbb{R}$. Is the action above proper?

(8) Let $M = S^2 \times S^2$ and consider the diagonal S^1 action on M given by

$$e^{i\theta} \cdot (u, v) = (e^{i\theta} \cdot u, e^{2i\theta} \cdot v),$$

where, for $u \in S^2 \subset \mathbb{R}^3$ and $e^{i\beta} \in S^1$, $e^{i\beta} \cdot u$ denotes the rotation of u by an angle β around the z-axis.

(a) Determine the fixed points for this action.

(b) What are the possible nontrivial stabilizers?

(9) Let G be a Lie group and H a closed Lie subgroup, i.e. a subgroup of G which is also a closed submanifold of G. Show that the action of H in G defined by $A(h, g) = h \cdot g$ is free and proper.

(10) (*Grassmannian*) Consider the set $H \subset GL(n)$ of invertible matrices of the form

$$\begin{pmatrix} A & 0 \\ C & B \end{pmatrix},$$

where $A \in GL(k)$, $B \in GL(n-k)$ and $C \in \mathcal{M}_{(n-k)\times k}$.

(a) Show that H is a Lie subgroup of $GL(n)$. Therefore H acts freely and properly on $GL(n)$ (cf. Exercise 7.17.9).

(b) Show that the points of the quotient manifold

$$Gr(n, k) = GL(n)/H$$

can be identified with the set of k-dimensional subspaces of $\mathbb{R}^n$ (in particular $Gr(n, 1)$ is just the projective space $\mathbb{R}P^{n-1}$).

(c) The manifold $Gr(n, k)$ is called the **Grassmannian** of k-planes in $\mathbb{R}^n$. What is its dimension?

(11) Let G and H be Lie groups and $F : G \to H$ a Lie group homomorphism. Show that:

(a) $(dF)_e : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism;

(b) if $(dF)_e$ is an isomorphism then F is a local diffeomorphism;

(c) if F is a surjective local diffeomorphism then F is a covering map.

(12) (a) Show that $\mathbb{R} \cdot SU(2)$ is a four dimensional real linear subspace of $\mathcal{M}_{2\times 2}(\mathbb{C})$, closed under matrix multiplication, with basis

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

$$j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

satisfying $i^2 = j^2 = k^2 = ijk = -1$. Therefore this space can be identified with the **quaternions**. Show that $SU(2)$ can be identified with the quaternions of Euclidean length equal to 1, and is therefore diffeomorphic to S^3.

(b) Let us identify $\mathbb{R}^3$ with the quaternions of zero real part. Show that if $n \in \mathbb{R}^3$ is a unit vector then

$$\exp\left(\frac{n\theta}{2}\right) = 1\cos\left(\frac{\theta}{2}\right) + n\sin\left(\frac{\theta}{2}\right)$$

is also a unit quaternion.

(c) Show that the map

$$\begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R}^3 \\ v & \mapsto & \exp\left(\frac{n\theta}{2}\right) \cdot v \cdot \exp\left(-\frac{n\theta}{2}\right) \end{array}$$

is a rotation by an angle θ about the axis defined by n.

(d) Show that there exists a surjective homomorphism $F : SU(2) \to SO(3)$, and use this to conclude that $SU(2)$ is the universal covering of $SO(3)$.

(e) What is the fundamental group of $SO(3)$?

8. Orientability

Let V be a finite dimensional vector space and consider two ordered bases $\beta = \{b_1, \ldots, b_n\}$ and $\beta' = \{b'_1, \ldots, b'_n\}$. There is a unique linear transformation $S : V \to V$ such that $b'_i = S\, b_i$ for every $i = 1, \ldots, n$. We say that the two bases are **equivalent** if $\det S > 0$. This defines an equivalence relation that divides the set of all ordered basis of V into two equivalence classes. An **orientation** for V is an assignment of a positive sign to the elements of one equivalence class and a negative sign to the elements of the other. The sign assigned to a basis is called its **orientation** and the basis is said to be **positively oriented** or **negatively oriented** according to its sign. It is clear that there are exactly two possible orientations for V.

Remark 8.1.

(1) The ordering of the basis is very important. If we interchange the positions of two basis vectors we obtain a different ordered basis with the opposite orientation.

(2) An orientation for a zero-dimensional vector space is just an assignment of a sign $+1$ or -1.
(3) We call the **standard orientation** of $\mathbb{R}^n$ to the orientation that assigns a positive sign to the standard ordered basis.

An isomorphism $A : V \to W$ between two oriented vector spaces carries two ordered bases of V in the same equivalence class to equivalent ordered bases of W. Hence, for any ordered basis β, the sign of the image $A\beta$ is either always the same as the sign of β or always the opposite. In the first case, the isomorphism A is said to be **orientation preserving**, and in the latter it is called **orientation reversing**.

An **orientation** of a smooth manifold consists on a choice of orientations for all tangent spaces T_pM. If $\dim M = n \geq 1$, these orientations have to fit together smoothly, meaning that for each point $p \in M$ there exists a parametrization (U, φ) around p such that

$$(d\varphi)_x : \mathbb{R}^n \to T_{\varphi(x)}M$$

preserves the standard orientation of $\mathbb{R}^n$ at each point $x \in U$.

Remark 8.2. If the dimension of M is zero, an orientation is just an assignment of a sign ($+1$ or -1), called **orientation number**, to each point $p \in M$.

Definition 8.3. *A smooth manifold M is said to be* **orientable** *if it admits an orientation.*

Proposition 8.4. *If a smooth manifold M is connected and orientable then it admits precisely two orientations.*

Proof. We will show that the set of points where two orientations agree and the set of points where they disagree are both open. Hence, one of them has to be M and the other the empty set. Let p be a point in M and let $(U_\alpha, \varphi_\alpha)$, (U_β, φ_β) be two parametrizations centered at p such that $d\varphi_\alpha$ is orientation preserving for the first orientation and $d\varphi_\beta$ is orientation preserving for the second. The map $\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right)_0 : \mathbb{R}^n \to \mathbb{R}^n$ is either orientation preserving (if the two orientations agree at p) or reversing. In the first case, it has positive determinant at 0, and so, by continuity, $\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right)_x$ has positive determinant for x on a neighborhood of 0, implying that the two orientations agree on a neighborhood of p. Similarly, if $\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right)_0$ is orientation reversing, the determinant of $\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right)_x$ is negative on a neighborhood of 0, and so the two orientations disagree on a neighborhood of p.

Let O be an orientation for M (i.e. a smooth choice of an orientation O_p of T_pM for each $p \in M$), and $-O$ the opposite orientation (corresponding to taking the opposite orientation $-O_p$ at each tangent space T_pM). If O' is another orientation for M, then, for a given point $p \in M$, we know that

O'_p agrees either with O_p or with $-O_p$ (because a vector space has just two possible orientations). Consequently, O' agrees with either O or $-O$ on M. □

An alternative characterization of orientability is given by the following proposition, whose proof is left as an exercise.

PROPOSITION 8.5. *A smooth manifold M is orientable if and only if there exists an atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ for which all the overlap maps $\varphi_\beta^{-1} \circ \varphi_\alpha$ are orientation-preserving.*

An **oriented** manifold is an orientable manifold together with a choice of an orientation.

EXERCISES 8.6.

(1) Prove that the relation of "being equivalent" between ordered basis of a finite dimensional vector space described above is an equivalence relation.
(2) Show that a differentiable manifold M is orientable *iff* there exists an atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ for which all the overlap maps $\varphi_\beta^{-1} \circ \varphi_\alpha$ are orientation-preserving.
(3) (a) Show that if a manifold M is covered by two coordinate neighborhoods V_1 and V_2 such that $V_1 \cap V_2$ is connected, then M is orientable.
    (b) Show that S^n is orientable.
(4) Let M be an oriented n-dimensional manifold and $c : I \to M$ a differentiable curve. A **smooth vector field along** c is a differentiable map $V : I \to TM$ such that $V(t) \in T_{c(t)}M$ for all $t \in I$ (cf. Section 2 in Chapter 3). Show that if $V_1, \dots, V_n : I \to M$ are smooth vector fields along c such that $\{V_1(t), \dots, V_n(t)\}$ is a basis of $T_{c(t)}M$ for all $t \in I$ then all these basis have the same orientation.
(5) The **Möbius band** is the 2-dimensional submanifold of $\mathbb{R}^3$ given by the image of the immersion $g : (-1, 1) \times \mathbb{R} \to \mathbb{R}^3$ defined by
$$g(t, \varphi) = \left(\left(1 + t\cos\left(\frac{\varphi}{2}\right)\right)\cos\varphi, \left(1 + t\cos\left(\frac{\varphi}{2}\right)\right)\sin\varphi, t\sin\left(\frac{\varphi}{2}\right)\right).$$
Show that the Möbius band is not orientable.
(6) Let $f : M \to N$ be a diffeomorphism between two smooth manifolds. Show that M is orientable if and only if N is orientable. If, in addition, both manifolds are connected and oriented, and $(df)_p : T_pM \to T_{f(p)}N$ preserves orientation at one point $p \in M$, show that it is orientation preserving at all points. The map f is said to be **orientation preserving** in this case, and **orientation reversing** otherwise.
(7) Let M and N be two oriented manifolds. We define an orientation on the product manifold $M \times N$ (called **product orientation**) in the following way: If $\alpha = \{a_1, \dots, a_m\}$ and $\beta = \{b_1, \dots, b_n\}$

are ordered bases of T_pM and T_qN, we consider the ordered basis $\{(a_1,0),\ldots,(a_m,0),(0,b_1),\ldots,(0,b_n)\}$ of $T_{(p,q)}(M\times N)\cong T_pM\times T_qN$. We then define an orientation on this space by setting the sign of this basis equal to the product of the signs of α and β. Show that this orientation does not depend on the choice of α and β.

(8) Show that the tangent bundle TM is always orientable, even if M is not.

(9) (*Orientable double covering*) Let M be a non-orientable n-dimensional manifold. For each point $p\in M$ we consider the set $\mathcal{O}_p$ of the (two) equivalence classes of bases of T_pM. Let $\overline{M}$ be the set

$$\overline{M}=\{(p,O_p)\mid p\in M, O_p\in\mathcal{O}_p\}.$$

Given a parametrization (U,φ) of M consider the maps $\overline{\varphi}:U\to\overline{M}$ defined by

$$\overline{\varphi}(x^1,\ldots,x^n)=\left(\varphi(x^1,\ldots,x^n),\left[\left(\frac{\partial}{\partial x^1}\right)_{\varphi(x)},\ldots,\left(\frac{\partial}{\partial x^n}\right)_{\varphi(x)}\right]\right),$$

where $x=(x^1,\ldots,x^n)\in U$ and $\left[\left(\frac{\partial}{\partial x^1}\right)_{\varphi(x)},\ldots,\left(\frac{\partial}{\partial x^n}\right)_{\varphi(x)}\right]$ represents the equivalence class of the basis $\left\{\left(\frac{\partial}{\partial x^1}\right)_{\varphi(x)},\ldots,\left(\frac{\partial}{\partial x^n}\right)_{\varphi(x)}\right\}$ of $T_{\varphi(x)}M$.

(a) Show that these maps determine the structure of an **orientable** differentiable manifold of dimension n on $\overline{M}$.

(b) Consider the map $\pi:\overline{M}\to M$ defined by $\pi(p,O_p)=p$. Show that π is differentiable and surjective. Moreover, show that, for each $p\in M$, there exists a neighborhood V of p with $\pi^{-1}(V)=W_1\cup W_2$, where W_1 e W_2 are two disjoint open subsets of $\overline{M}$, such that π restricted to W_i $(i=1,2)$ is a diffeomorphism onto V.

(c) Show that M is connected ($\overline{M}$ is therefore called the **orientable double covering** of M).

(d) Let $\sigma:\overline{M}\to\overline{M}$ be the map defined by $\sigma(p,O_p)=(p,-O_p)$, where $-O_p$ represents the orientation of T_pM opposite to O_p. Show that σ is a diffeomorphism which reverses orientations satisfying $\pi\circ\sigma=\pi$ and $\sigma\circ\sigma=\mathrm{id}$.

(e) Show that any simply connected manifold is orientable.

9. Manifolds with Boundary

Let us consider again the **closed half space**

$$\mathbb{H}^n=\{(x^1,\ldots,x^n)\in\mathbb{R}^n\mid x^n\geq 0\}$$

with the topology induced by the usual topology of $\mathbb{R}^n$. Recall that a map $f:U\to\mathbb{R}^m$ defined on an open set $U\subset\mathbb{H}^n$ is said to be **differentiable** if it is the restriction to U of a differentiable map $\tilde{f}$ defined on an open subset of $\mathbb{R}^n$ containing U (cf. Section 10.2). In this case, the derivative $(df)_p$ is defined to be $\left(d\tilde{f}\right)_p$. Note that this derivative is independent of the extension used since any two extensions have to agree on U.

DEFINITION 9.1. *A smooth n-**manifold with boundary** is a topological manifold with boundary of dimension n and a family of parametrizations $\varphi_\alpha : U_\alpha \subset \mathbb{H}^n \to M$ (that is, homeomorphisms of open sets U_α of $\mathbb{H}^n$ onto open sets of M), such that:*

(i) the coordinate neighborhoods cover M, meaning that $\bigcup_\alpha \varphi_\alpha(U_\alpha) = M$;

(ii) for each pair of indices α, β such that

$$W := \varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta) \neq \varnothing,$$

the overlap maps

$$\begin{aligned} \varphi_\beta^{-1} \circ \varphi_\alpha : \varphi_\alpha^{-1}(W) &\to \varphi_\beta^{-1}(W) \\ \varphi_\alpha^{-1} \circ \varphi_\beta : \varphi_\beta^{-1}(W) &\to \varphi_\alpha^{-1}(W) \end{aligned}$$

are smooth;

(iii) the family $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ is maximal with respect to (i) and (ii), meaning that, if $\varphi_0 : U_0 \to M$ is a parametrization such that $\varphi_0 \circ \varphi^{-1}$ and $\varphi^{-1} \circ \varphi_0$ are C^∞ for all φ in $\mathcal{A}$, then φ_0 is in $\mathcal{A}$.

Recall that a point in M is said to be a **boundary point** if it is on the image of $\partial\mathbb{H}^n$ under some parametrization (that is, if there is a parametrization $\varphi : U \subset \mathbb{H}^n \to M$ such that $\varphi(x^1, \ldots, x^{n-1}, 0) = p$ for some $(x^1, \ldots, x^{n-1}) \in \mathbb{R}^{n-1}$), and that the set ∂M of all such points is called the **boundary** of M.

PROPOSITION 9.2. *The boundary of a smooth n-manifold with boundary is a differentiable manifold of dimension $n-1$.*

PROOF. Suppose that p is a boundary point of M (an n-manifold with boundary) and choose a parametrization $\varphi_\alpha : U_\alpha \subset \mathbb{H}^n \to M$ around p. Letting $V_\alpha := \varphi_\alpha(U_\alpha)$, we claim that $\varphi_\alpha(\partial U_\alpha) = \partial V_\alpha$, where $\partial U_\alpha = U_\alpha \cap \partial\mathbb{H}^n$ and $\partial V_\alpha = V_\alpha \cap \partial M$. By definition of boundary, we already know that $\varphi_\alpha(\partial U_\alpha) \subset \partial V_\alpha$, so we just have to show that $\partial V_\alpha \subset \varphi_\alpha(\partial U_\alpha)$. Let $q \in \partial V_\alpha$ and consider a parametrization $\varphi_\beta : U_\beta \to V_\alpha$ around q, mapping an open subset of $\mathbb{H}^n$ to an open subset of M and such that $q \in \varphi_\beta(\partial U_\beta)$. If we show that $\varphi_\beta(\partial U_\beta) \subset \varphi_\alpha(\partial U_\alpha)$ we are done. For that, we prove that $\left(\varphi_\alpha^{-1} \circ \varphi_\beta\right)(\partial U_\beta) \subset \partial U_\alpha$. Indeed, suppose that this map $\varphi_\alpha^{-1} \circ \varphi_\beta$ takes a point $x \in \partial U_\beta$ to an interior point (in $\mathbb{R}^n$) of U_α. As this map is a diffeomorphism, x would be an interior point (in $\mathbb{R}^n$) of U_β. This, of course, contradicts the assumption that $x \in \partial U_\beta$. Hence, $\left(\varphi_\alpha^{-1} \circ \varphi_\beta\right)(\partial U_\beta) \subset \partial U_\alpha$ and so $\varphi_\beta(\partial U_\beta) \subset \varphi_\alpha(\partial U_\alpha)$.

The map φ_α then restricts to a diffeomorphism from ∂U_α onto ∂V_α, where we identify ∂U_α with an open subset of $\mathbb{R}^{n-1}$. We obtain in this way a parametrization around p in ∂M. ∎

Remark 9.3. In the above proof we saw that the definition of a boundary point does not depend on the parametrization chosen, meaning that, if there exists a parametrization around p such that p is an image of a point in $\partial\mathbb{H}^n$, then any parametrization around p maps a boundary point of $\mathbb{H}^n$ to p.

It is easy to see that if M is orientable then so is ∂M.

Proposition 9.4. *Let M be an orientable manifold with boundary. Then ∂M is also orientable.*

Proof. If M is orientable we can choose an atlas $\{(U_\alpha, \varphi_\alpha)\}$ on M for which the determinants of the derivatives of all overlap maps are positive. With this atlas we can obtain an atlas $\{(\partial U_\alpha, \tilde{\varphi}_\alpha)\}$ for ∂M in the way described in the proof of Proposition 9.2. For any overlap map

$$(\varphi_\beta^{-1} \circ \varphi_\alpha)(x^1, \ldots, x^n) = (y^1(x^1, \ldots, x^n), \ldots, y^n(x^1, \ldots, x^n))$$

we have

$$(\varphi_\beta^{-1} \circ \varphi_\alpha)(x^1, \ldots, x^{n-1}, 0) = (y^1(x^1, \ldots, x^{n-1}, 0), \ldots, y^{n-1}(x^1, \ldots, x^{n-1}, 0), 0)$$

and

$$(\tilde{\varphi}_\beta^{-1} \circ \tilde{\varphi}_\alpha)(x^1, \ldots, x^{n-1}) = (y^1(x^1, \ldots, x^{n-1}, 0), \ldots, y^{n-1}(x^1, \ldots, x^{n-1}, 0)).$$

Consequently, denoting $(x^1, \ldots, x^{n-1}, 0)$ by $(\tilde{x}, 0)$,

$$(d(\varphi_\beta^{-1} \circ \varphi_\alpha))_{(\tilde{x},0)} = \begin{pmatrix} (d(\tilde{\varphi}_\beta^{-1} \circ \tilde{\varphi}_\alpha))_{\tilde{x}} & | & * \\ -- & + & -- \\ 0 & | & \frac{\partial y^n}{\partial x^n}(\tilde{x}, 0) \end{pmatrix}$$

and so

$$\det\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right)_{(\tilde{x},0)} = \frac{\partial y^n}{\partial x^n}(\tilde{x}, 0) \, \det\left(d(\tilde{\varphi}_\beta^{-1} \circ \tilde{\varphi}_\alpha)\right)_{\tilde{x}}.$$

However, fixing $x^1, \cdots, x^{n-1}$, we have that y^n is positive for positive values of x^n and is zero for $x^n = 0$. Consequently, $\frac{\partial y^n}{\partial x^n}(\tilde{x}, 0) > 0$, and so

$$\det\left(d(\tilde{\varphi}_\beta^{-1} \circ \tilde{\varphi}_\alpha)\right)_{\tilde{x}} > 0.$$

$\square$

Hence, choosing an orientation on a manifold with boundary M induces an orientation on the boundary ∂M. The convenient choice, called the **induced orientation**, can be obtained in the following way. For $p \in \partial M$ the tangent space $T_p(\partial M)$ is a subspace of T_pM of codimension 1. As we have seen above, considering a coordinate system $x : W \to \mathbb{R}^n$ around p, we have $x^n(p) = 0$ and $(x^1, \ldots, x^{n-1})$ is a coordinate system around p in ∂M. Setting $n_p := -\left(\frac{\partial}{\partial x^n}\right)_p$ (called an **outward pointing vector** at p), the induced orientation on ∂M is defined by assigning a positive sign to an ordered basis β of $T_p(\partial M)$ whenever the ordered basis $\{n_p, \beta\}$ of T_pM is positive, and negative otherwise. Note that, since $\frac{\partial y^n}{\partial x^n}(\varphi^{-1}(p)) > 0$ (in the above notation), the sign of the last component of n_p does not depend on the choice of coordinate system. In general, the induced orientation is **not** the one obtained from the charts of M by simply dropping the last coordinate (in fact, it is $(-1)^n$ times this orientation).

Exercises 9.5.

(1) Show that there is no diffeomorphism between a neighborhood of 0 in $\mathbb{R}^n$ and a neighborhood of 0 in $\mathbb{H}^n$.
(2) Show with an example that the product of two manifolds with boundary is not always a manifold with boundary.
(3) Let M be a manifold without boundary and N a manifold with boundary. Show that the product $M \times N$ is a manifold with boundary. What is $\partial(M \times N)$?
(4) Show that a diffeomorphism between two manifolds with boundary M and N maps the boundary ∂M diffeomorphically onto ∂N.

10. Notes on Chapter 1

10.1. Section 1. We begin by briefly reviewing the main concepts and results from general topology that we will need (see [**Mun00**] for a detailed exposition).

(1) A **topology** on a set M is a collection $\mathcal{T}$ of subsets of M having the following properties:
   (i) the sets $\varnothing$ and M are in $\mathcal{T}$;
   (ii) the union of the elements of any sub-collection of $\mathcal{T}$ is in $\mathcal{T}$;
   (iii) the intersection of the elements of any finite sub-collection of $\mathcal{T}$ is in $\mathcal{T}$.

   A set M equipped with a topology $\mathcal{T}$ is called a **topological space**. We say that a subset $U \subset M$ is an **open set** of M if it belongs to the topology $\mathcal{T}$. A **neighborhood** of a point $p \in M$ is simply an open set $U \in \mathcal{T}$ containing p. A **closed set** $F \subset M$ is a set whose complement $M \setminus F$ is open. The **interior** $\operatorname{int}A$ of a subset $A \subset M$ is the largest open set contained in A, and its **closure** $\overline{A}$ is the smallest closed set containing A. Finally, the **subspace topology** on $A \subset M$ is $\{U \cap A\}_{U \in \mathcal{T}}$.
(2) A topological space $(M, \mathcal{T})$ is said to be **Hausdorff** if, for each pair of distinct points $p_1, p_2 \in M$, there exist neighborhoods U_1, U_2 of p_1 and p_2 such that $U_1 \cap U_2 = \varnothing$.
(3) A **basis** for a topology $\mathcal{T}$ on M is a collection $\mathcal{B} \subset \mathcal{T}$ such that, for each point $p \in M$ and each open set U containing p, there exists a basis element $B \in \mathcal{B}$ for which $p \in B \subset U$. If $\mathcal{B}$ is a basis for a topology $\mathcal{T}$ then any element of $\mathcal{T}$ is a union of elements of $\mathcal{B}$. A topological space $(M, \mathcal{T})$ is said to satisfy the **second countability axiom** if $\mathcal{T}$ has a countable base.
(4) A map $f : M \to N$ between two topological spaces is said to be **continuous** if, for each open set $U \subset N$, the preimage $f^{-1}(U)$ is

an open subset of M. A bijection f is called a **homeomorphism** if both f and its inverse f^{-1} are continuous.

(5) An **open cover** for a topological space $(M, \mathcal{T})$ is a collection $\{U_\alpha\} \subset \mathcal{T}$ such that $\bigcup_\alpha U_\alpha = M$. A **subcover** is a sub-collection $\{V_\beta\} \subset \{U_\alpha\}$ which is still an open cover. A topological space is said to be **compact** if every open cover admits a finite subcover. A subset $A \subset M$ is said to be a **compact subset** if it is a compact topological space for the subspace topology. Continuous maps carry compact sets to compact sets.

(6) A topological space is said to be **connected** if the only subsets of M which are simultaneously open and closed are $\varnothing$ and M. A subset $A \subset M$ is said to be a **connected subset** if it is a connected topological space for the subspace topology. Continuous maps carry connected sets to connected sets.

(7) Let $(M, \mathcal{T})$ be a topological space. A sequence $\{p_n\}$ in M is said to **converge** to $p \in M$ if, for each neighborhood V of p, there exists an $N \in \mathbb{N}$ for which $p_n \in V$ for $n > N$. If $(M, \mathcal{T})$ is Hausdorff, then a convergent sequence has a unique limit. If in addition $(M, \mathcal{T})$ is second countable, then $F \subset M$ is closed if and only if every convergent sequence in F has limit in F, and $K \subset M$ is compact if and only if every sequence in K has a sublimit in K.

(8) If M and N are topological spaces, the set of all Cartesian products of open subsets of M by open subsets of N is a basis for a topology on $M \times N$, called the **product topology**. Note that with this topology the canonical projections are continuous maps.

(9) An **equivalence relation** $\sim$ on a set M is a relation with the following properties:
   (i) *reflexivity*: $p \sim p$ for every $p \in M$;
   (ii) *symmetry*: if $p \sim q$ then $q \sim p$;
   (iii) *transitivity*: if $p \sim q$ and $q \sim r$ then $p \sim r$.

   Given a point $p \in M$, we define the **equivalence class** of p as the set

$$[p] = \{q \in M \mid q \sim p\}.$$

   Note that $p \in [p]$ by reflexivity. Whenever we have an equivalence relation $\sim$ on a set M, the corresponding set of equivalence classes is called the **quotient space**, and is denoted by $M/\sim$. There is a canonical projection $\pi : M \to M/\sim$, which maps each element of M to its equivalence class. If M is a topological space, we can define a topology on the quotient space (called the **quotient topology**) by letting a subset $V \subset M/\sim$ be open if and only if the set $\pi^{-1}(V)$ is open in M. The map π is then continuous for this topology. We will be interested in knowing whether some quotient spaces are Hausdorff. For that, the following definition will be helpful.

Definition 10.1. *An equivalence relation $\sim$ on a topological space M is called* **open** *if the map $\pi : M \to M/\!\sim$ is open, i.e., if for every open set $U \subset M$, the set $[U] = \pi(U)$ is open.*

Proposition 10.2. *Let $\sim$ be an open equivalence relation on M and let $R = \{(p,q) \in M \times M \mid p \sim q\}$. Then the quotient space is Hausdorff if and only if R is closed in $M \times M$.*

Proof. Assume that R is closed. Let $[p],[q] \in M/\sim$ with $[p] \neq [q]$. Then $p \nsim q$, and $(p,q) \notin R$. As R is closed, there are open sets U, V containing p, q, respectively, such that $(U \times V) \cap R = \varnothing$. This implies that $[U] \cap [V] = \varnothing$. In fact, if there were a point $[r] \in [U] \cap [V]$, then r would be equivalent to points $p' \in U$ and $q' \in V$ (that is $p' \sim r$ and $r \sim q'$). Therefore we would have $p' \sim q'$ (implying that $(p',q') \in R$), and so $(U \times V) \cap R$ would not be empty. Since $[U]$ and $[V]$ are open (as $\sim$ is an open equivalence relation), we conclude that $M/\!\sim$ is Hausdorff.

Conversely, let us assume that $M/\!\sim$ is Hausdorff. If $(p,q) \notin R$, then $p \nsim q$ and $[p] \neq [q]$, implying the existence of open sets $\widetilde{U}, \widetilde{V} \subset M/\!\sim$ containing $[p]$ and $[q]$, such that $\widetilde{U} \cap \widetilde{V} = \varnothing$. The sets $U := \pi^{-1}(\widetilde{U})$ and $V := \pi^{-1}(\widetilde{V})$ are open in M and $(U \times V) \cap R = \varnothing$. In fact, if that was not so, there would exist points $p' \in U$ and $q' \in V$ such that $p' \sim q'$. Then we would have $[p'] = [q']$, contradicting the fact that $\widetilde{U} \cap \widetilde{V} = \varnothing$ (as $[p'] \in \pi(U) = \widetilde{U}$ and $[q'] \in \pi(V) = \widetilde{V}$). Since $(p,q) \in U \times V \subset (M \times M) \setminus R$ and $U \times V$ is open, we conclude that $(M \times M) \setminus R$ is open, and hence R is closed. ☐

10.2. Section 2.

(1) Let us begin by reviewing some facts about differentiability of maps on $\mathbb{R}^n$. A function $f : U \to \mathbb{R}$ defined on an open subset U of $\mathbb{R}^n$ is said to be **continuously differentiable** on U if all partial derivatives $\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}$ exist and are continuous on U. In this book, the words *differentiable* and *smooth* will be used to mean **infinitely differentiable**, that is, all partial derivatives $\frac{\partial^k f}{\partial x^{i_1} \cdots \partial x^{i_k}}$ exist and are continuous on U. Similarly, a map $F : U \to \mathbb{R}^m$, defined on an open subset of $\mathbb{R}^n$, is said to be **differentiable** or **smooth** if all coordinate functions f^i have the same property, that is, if they all possess continuous partial derivatives of all orders. If the map F is differentiable on U, its derivative at each point of U is the linear map $DF : \mathbb{R}^n \to \mathbb{R}^m$ represented in the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ by the **Jacobian matrix**

$$DF = \begin{bmatrix} \frac{\partial f^1}{\partial x^1} & \cdots & \frac{\partial f^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial f^m}{\partial x^1} & \cdots & \frac{\partial f^m}{\partial x^n} \end{bmatrix}.$$

A map $F : A \to \mathbb{R}^m$ defined on an arbitrary set $A \subset \mathbb{R}^n$ (not necessarily open) is said to be **differentiable** on A is there exists an open set $U \supset A$ and a differentiable map $\widetilde{F} : U \to \mathbb{R}^m$ such that $F = \widetilde{F}|_A$.

10.3. Section 4.

(1) Let E, B and F be smooth manifolds and $\pi : E \to B$ a differentiable map. Then, $\pi : E \to B$ is called a **fiber bundle** of **basis** B, **total space** E and **fiber** F if
  (i) the map π is surjective;
  (ii) there is a covering of B by open sets $\{U_\alpha\}$ and diffeomorphisms $\psi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times F$ such that for every $b \in U_\alpha$ we have $\psi_\alpha(\pi^{-1}(b)) = \{b\} \times F$.

10.4. Section 5.

(1) (*The Inverse Function Theorem*) Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be a smooth function and $p \in U$ such that $(df)_p$ is a linear isomorphism. Then there exists an open subset $V \subset U$ containing p such that $f|_V : V \to f(V)$ is a diffeomorphism.

10.5. Section 7.

(1) A **group** is a set G equipped with a binary operation $\cdot : G \times G \to G$ satisfying:
  (i) **Associativity:** $g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$ for all $g_1, g_2, g_3 \in G$;
  (ii) **Existence of identity:** There exists an element $e \in G$ such that $e \cdot g = g \cdot e = g$ for all $g \in G$;
  (iii) **Existence of inverses:** For all $g \in G$ there exists $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

  If the group operation is commutative, meaning that $g_1 \cdot g_2 = g_2 \cdot g_1$ for all $g_1, g_2 \in G$, the group is said to be **abelian**. A subset $H \subset G$ is said to be a **subgroup** of G if the restriction of $\cdot$ to $H \times H$ is a binary operation on H, and H, with this operation, is a group. A subgroup $H \subset G$ is said to be **normal** if $ghg^{-1} \in H$ for all $g \in G, h \in H$. A map $f : G \to H$ between two groups G and H is said to be a **group homomorphism** if $f(g_1 \cdot g_2) = f(g_1) \cdot f(g_2)$ for all $g_1, g_2 \in G$. An **isomorphism** is a bijective homomorphism. The **kernel** of a group homomorphism $f : G \to H$ is the subset $\ker(f) = \{g \in G \mid f(g) = e\}$, and is easily seen to be a normal subgroup of G.

(2) Let $f, g : X \to Y$ be two continuous maps between topological spaces and let $I = [0, 1]$. We say that f is **homotopic** to g if there exists a continuous map $H : I \times X \to Y$ such that $H(0, x) = f(x)$ and $H(1, x) = g(x)$ for every $x \in X$. This map is called a **homotopy**.

Homotopy of maps forms an equivalence relation in the set of continuous maps between X and Y. As an application, let us fix a **base point** p on a manifold M and consider the homotopy classes of continuous maps $f : I \to M$ such that $f(0) = f(1) = p$ (these maps are called **loops** based at p), with the additional restriction that $H(t, 0) = H(t, 1) = p$ for all $t \in I$. This set of homotopy classes is called the **fundamental group** of M relative to the base point p, and is denoted by $\pi_1(M, p)$. Among its elements there is the class of the **constant loop based at** p, given by $f(t) = p$ for every $t \in I$. Note that the set $\pi_1(M, p)$ is indeed a group with operation $*$ (**composition** of loops) defined by $[f]*[g] := [h]$, where $h : I \to M$ is given by

$$h(t) = \begin{cases} f(2t) & \text{if } t \in [0, \frac{1}{2}] \\ g(2t-1) & \text{if } t \in [\frac{1}{2}, 1] \end{cases} .$$

The identity element of this group is the equivalence class of the constant loop based at p.

If M is connected and this is the only class in $\pi_1(M, p)$, M is said to be **simply connected**. This means that every loop through p can be continuously deformed to the constant loop. This property does not depend on the choice of point p, and is equivalent to the condition that any closed path may be continuously deformed to a constant loop in M.

10.6. Bibliographical notes. The material in this chapter is completely standard, and can be found in almost any book on differential geometry (e.g. [**Boo03, dC93, GHL04**]). Immersions and embeddings are the starting point of **differential topology**, which is studied on [**GP73, Mil97**]. Lie groups and Lie algebras are a huge field of Mathematics, to which we could not do justice. See for instance [**BtD03, DK99, War83**]. More details on the fundamental group and covering spaces can be found in [**Mun00**].

CHAPTER 2

Differential Forms

This chapter deals with **differential forms**, a fundamental tool in differential geometry.

Section 1 reviews the notions of **tensors** and **tensor product**, and introduces **alternating tensors** and their **exterior product**.

Tensor fields, which are natural generalizations of vector fields, are discussed in Section 2, where a new operation, the **pull-back** of a covariant tensor field by a smooth map, is defined. Section 3 studies fields of alternating tensors, or **differential forms**, and their **exterior derivative**. Important ideas such as the **Poincaré Lemma** and **de Rham cohomology**, which will not be needed in the remainder of this book, are discussed in the exercises.

In Section 4 we define the **integral** of a differential form on a smooth manifold. To do so we make use of another fundamental tool in differential geometry, namely the existence of **partitions of unity**.

The far-reaching **Stokes Theorem** is proved in Section 5, and some of its consequences are explored in the exercises.

Finally, in Section 6 we study the relation between orientability and the existence of special differential forms, called **volume forms**.

1. Tensors

Let V be an n-dimensional vector space. A **k-tensor** on V is a real multilinear function (meaning linear in each variable) defined on the product $V \times \cdots \times V$ of k copies of V. The set of all k-tensors is itself a vector space and is usually denoted by $\mathcal{T}^k(V^*)$.

EXAMPLE 1.1.
(1) The space of 1-tensors $\mathcal{T}^1(V^*)$ is equal to V^*, the **dual space** of V, that is, the space of real-valued linear functions on V.
(2) The usual inner product on $\mathbb{R}^n$ is an example of a 2-tensor.
(3) The determinant is an n-tensor on $\mathbb{R}^n$.

Given a k-tensor T and an m-tensor S, we define their **tensor product** as the $(k+m)$-tensor $T \otimes S$ given by

$$T \otimes S(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+m}) := T(v_1, \ldots, v_k) \cdot S(v_{k+1}, \ldots, v_{k+m}).$$

This operation is bilinear and associative, but not commutative (cf. Exercise 1.14.1).

Proposition 1.2. *If* $\{T_1, \dots, T_n\}$ *is a basis for* $\mathcal{T}^1(V^*) = V^*$ *(the dual space of* V*), then the set* $\{T_{i_1} \otimes \cdots \otimes T_{i_k} \mid 1 \leq i_1, \dots, i_k \leq n\}$ *is a basis of* $\mathcal{T}^k(V^*)$*, and therefore* $\dim \mathcal{T}^k(V^*) = n^k$.

Proof. We will first show that the elements of this set are linearly independent. If

$$T := \sum_{i_1, \cdots, i_k} a_{i_1 \cdots i_k} T_{i_1} \otimes \cdots \otimes T_{i_k} = 0,$$

then, taking the basis $\{v_1, \dots, v_n\}$ of V dual to $\{T_1, \dots, T_n\}$, meaning that $T_i(v_j) = \delta_{ij}$ (cf. Section 7.1), we have $T(v_{j_1}, \dots, v_{j_k}) = a_{j_1 \cdots j_k} = 0$ for every $1 \leq j_1, \dots, j_k \leq n$.

To show that $\{T_{i_1} \otimes \cdots \otimes T_{i_k} \mid 1 \leq i_1, \dots, i_k \leq n\}$ spans $\mathcal{T}^k(V^*)$, we take any element $T \in \mathcal{T}^k(V^*)$ and consider the k-tensor S defined by

$$S := \sum_{i_1, \cdots, i_k} T(v_{i_1}, \dots, v_{i_k}) T_{i_1} \otimes \cdots \otimes T_{i_k}.$$

Clearly, $S(v_{i_1}, \dots, v_{i_k}) = T(v_{i_1}, \dots, v_{i_k})$ for every $1 \leq i_1, \dots, i_k \leq n$, and so, by linearity, $S = T$. □

If we consider k-tensors on V^*, instead of V, we obtain the space $\mathcal{T}^k(V)$ (note that $(V^*)^* = V$, as is shown in Section 7.1). These tensors are called **contravariant tensors** on V, while the elements of $\mathcal{T}^k(V^*)$ are called **covariant tensors** on V. Note that the contravariant tensors on V are the covariant tensors on V^*. The words covariant and contravariant are related to the transformation behavior of the tensor components under a change of basis in V, as explained in Section 7.1.

We can also consider **mixed** (k, m)-tensors on V, that is, multilinear functions defined on the product $V \times \cdots \times V \times V^* \times \cdots \times V^*$ of k copies of V and m copies of V^*. A (k, m)-tensor is then k times covariant and m times contravariant on V. The space of all (k, m)-tensors on V is denoted by $\mathcal{T}^{k,m}(V^*, V)$.

Remark 1.3.

(1) We can identify the space $\mathcal{T}^{1,1}(V^*, V)$ with the space of linear maps from V to V. Indeed, for each element $T \in \mathcal{T}^{1,1}(V^*, V)$, we define the linear map from V to V, given by $v \mapsto T(v, \cdot)$. Note that $T(v, \cdot) : V^* \to \mathbb{R}$ is a linear function on V^*, that is, an element of $(V^*)^* = V$.

(2) Generalizing the above definition of tensor product to tensors defined on different vector spaces, we can define the spaces $\mathcal{T}^k(V^*) \otimes \mathcal{T}^m(W^*)$ generated by the tensor products of elements of $\mathcal{T}^k(V^*)$ by elements of $\mathcal{T}^m(W^*)$. Note that $\mathcal{T}^{k,m}(V^*, V) = \mathcal{T}^k(V^*) \otimes \mathcal{T}^m(V)$. We leave it as an exercise to find a basis for this space.

A tensor is called **alternating** if, like the determinant, it changes sign every time two of its variables are interchanged, that is, if

$$T(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -T(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$

The space of all alternating k-tensors is a vector subspace $\Lambda^k(V^*)$ of $\mathcal{T}^k(V^*)$. Note that, for any alternating k-tensor T, we have $T(v_1, \dots, v_k) = 0$ if $v_i = v_j$ for some $i \neq j$.

EXAMPLE 1.4.

(1) All 1-tensors are trivially alternating, that is, $\Lambda^1(V^*) = \mathcal{T}^1(V^*) = V^*$.
(2) The determinant is an alternating n-tensor on $\mathbb{R}^n$.

Consider now S_k, the group of all possible permutations of $\{1, \dots, k\}$. If $\sigma \in S_k$, we set $\sigma(v_1, \dots, v_k) = (v_{\sigma(1)}, \dots, v_{\sigma(k)})$. Given a k-tensor $T \in \mathcal{T}^k(V^*)$ we can define a new alternating k-tensor, called $\operatorname{Alt}(T)$, in the following way:

$$\operatorname{Alt}(T) := \frac{1}{k!} \sum_{\sigma \in S_k} (\operatorname{sgn} \sigma)\, (T \circ \sigma)$$

where $\operatorname{sgn}\sigma$ is $+1$ or -1 according to whether σ is an even or an odd permutation. We leave it as an exercise to show that $\operatorname{Alt}(T)$ is in fact alternating.

EXAMPLE 1.5. If $T \in \mathcal{T}^3(V^*)$,

$$\begin{aligned} \operatorname{Alt}(T)(v_1, v_2, v_3) &= \tfrac{1}{6}\left(T(v_1, v_2, v_3) + T(v_3, v_1, v_2) + T(v_2, v_3, v_1)\right. \\ &\quad \left. -T(v_1, v_3, v_2) - T(v_2, v_1, v_3) - T(v_3, v_2, v_1)\right). \end{aligned}$$

We will now define the **wedge product** between alternating tensors: if $T \in \Lambda^k(V^*)$ and $S \in \Lambda^m(V^*)$, then $T \wedge S \in \Lambda^{k+m}(V^*)$ is given by

$$T \wedge S := \frac{(k+m)!}{k!\, m!} \operatorname{Alt}(T \otimes S).$$

EXAMPLE 1.6. If $T, S \in \Lambda^1(V^*) = V^*$, then

$$T \wedge S = 2 \operatorname{Alt}(T \otimes S) = T \otimes S - S \otimes T,$$

implying that $T \wedge S = -S \wedge T$ and $T \wedge T = 0$.

It is easy to verify that this product is bilinear. To prove associativity we need the following proposition

PROPOSITION 1.7.

(i) Let $T \in \mathcal{T}^k(V^*)$ *and* $S \in \mathcal{T}^m(V^*)$. *If* $\operatorname{Alt}(T) = 0$ *then*

$$\operatorname{Alt}(T \otimes S) = \operatorname{Alt}(S \otimes T) = 0;$$

(ii) $\operatorname{Alt}(\operatorname{Alt}(T \otimes S) \otimes R) = \operatorname{Alt}(T \otimes S \otimes R) = \operatorname{Alt}(T \otimes \operatorname{Alt}(S \otimes R))$.

PROOF.

(i) Let us consider

$$(k+m)!\ \mathrm{Alt}(T\otimes S)(v_1,\dots,v_{k+m}) = \sum_{\sigma\in S_{k+m}} (\operatorname{sgn}\sigma)\, T(v_{\sigma(1)},\dots,v_{\sigma(k)})S(v_{\sigma(k+1)},\dots,v_{\sigma(k+m)}).$$

Taking the subgroup G of S_{k+m} formed by the permutations that leave $k+1,\dots,k+m$ fixed, we have

$$\begin{aligned}\sum_{\sigma\in G}(\operatorname{sgn}\sigma)T(v_{\sigma(1)},\dots,v_{\sigma(k)})S(v_{\sigma(k+1)},\dots,v_{\sigma(k+m)}) =&\\ = \sum_{\sigma\in G}(\operatorname{sgn}\sigma)T(v_{\sigma(1)},\dots,v_{\sigma(k)})S(v_{k+1},\dots,v_{k+m})&\\ = k!\ \left(\mathrm{Alt}(T)\otimes S\right)(v_1,\dots,v_{k+m}) = 0.&\end{aligned}$$

Then, since G decomposes S_{k+m} into disjoint right cosets $G\cdot\widetilde{\sigma} = \{\sigma\widetilde{\sigma} \mid \sigma\in G\}$, and for each coset

$$\begin{aligned}\sum_{\sigma\in G\cdot\widetilde{\sigma}}(\operatorname{sgn}\sigma)(T\otimes S)(v_{\sigma(1)},\dots,v_{\sigma(k+m)}) =&\\ = (\operatorname{sgn}\widetilde{\sigma})\sum_{\sigma\in G}(\operatorname{sgn}\sigma)\,(T\otimes S)(v_{\sigma(\widetilde{\sigma}(1))},\dots,v_{\sigma(\widetilde{\sigma}(k+m))})&\\ = (\operatorname{sgn}\widetilde{\sigma})k!\ (\mathrm{Alt}(T)\otimes S)(v_{\widetilde{\sigma}(1)},\dots,v_{\widetilde{\sigma}(k+m)}) = 0,&\end{aligned}$$

we have that $\mathrm{Alt}(T\otimes S)=0$. Similarly, we prove that $\mathrm{Alt}(S\otimes T)=0$.

(ii) By linearity of the operator Alt and the fact that $\mathrm{Alt}\circ\mathrm{Alt}=\mathrm{Alt}$ (cf. Exercise 1.14.3), we have

$$\mathrm{Alt}(\mathrm{Alt}(S\otimes R) - S\otimes R) = 0.$$

Hence, by (i),

$$\begin{aligned}0 &= \mathrm{Alt}(T\otimes(\mathrm{Alt}(S\otimes R) - S\otimes R))\\ &= \mathrm{Alt}(T\otimes\mathrm{Alt}(S\otimes R)) - \mathrm{Alt}(T\otimes S\otimes R),\end{aligned}$$

and the result follows.

□

Using these properties we can show that

Proposition 1.8. *$(T\wedge S)\wedge R = T\wedge(S\wedge R)$.*

Proof. By Proposition 1.7, for $T\in\Lambda^k(V^*)$, $S\in\Lambda^m(V^*)$ and $R\in\Lambda^l(V^*)$, we have

$$\begin{aligned}(T\wedge S)\wedge R &= \frac{(k+m+l)!}{(k+m)!\,l!}\mathrm{Alt}((T\wedge S)\otimes R)\\ &= \frac{(k+m+l)!}{k!\,m!\,l!}\mathrm{Alt}(T\otimes S\otimes R)\end{aligned}$$

and

$$\begin{aligned} T \wedge (S \wedge R) &= \frac{(k+m+l)!}{k!\,(m+l)!} \mathrm{Alt}(T \otimes (S \wedge R)) \\ &= \frac{(k+m+l)!}{k!\,m!\,l!} \mathrm{Alt}(T \otimes S \otimes R). \end{aligned}$$

□

We can now prove the following theorem.

THEOREM 1.9. *If $\{T_1, \dots, T_n\}$ is a basis for V^*, then the set*

$$\{T_{i_1} \wedge \cdots \wedge T_{i_k} \mid 1 \le i_1 < \ldots < i_k \le n\}$$

is a basis for $\Lambda^k(V^*)$, and

$$\dim \Lambda^k(V^*) = \begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}.$$

PROOF. Let $T \in \Lambda^k(V^*) \subset \mathcal{T}^k(V^*)$. By Proposition 1.2,

$$T = \sum_{i_1,\dots,i_k} a_{i_1 \cdots i_k} T_{i_1} \otimes \cdots \otimes T_{i_k}$$

and, since T is alternating,

$$T = \mathrm{Alt}(T) = \sum_{i_1,\dots,i_k} a_{i_1 \cdots i_k} \mathrm{Alt}(T_{i_1} \otimes \cdots \otimes T_{i_k}).$$

We can show by induction that $\mathrm{Alt}(T_{i_1} \otimes \cdots \otimes T_{i_k}) = \frac{1}{k!} T_{i_1} \wedge T_{i_2} \wedge \cdots \wedge T_{i_k}$. Indeed, for $k = 1$, the result is trivially true, and, assuming it is true for k basis tensors, we have, by Proposition 1.7, that

$$\begin{aligned} \mathrm{Alt}(T_{i_1} \otimes \cdots \otimes T_{i_{k+1}}) &= \mathrm{Alt}(\mathrm{Alt}(T_{i_1} \otimes \cdots \otimes T_{i_k}) \otimes T_{i_{k+1}}) \\ &= \frac{k!}{(k+1)!} \mathrm{Alt}(T_{i_1} \otimes \cdots \otimes T_{i_k}) \wedge T_{i_{k+1}} \\ &= \frac{1}{(k+1)!} T_{i_1} \wedge T_{i_2} \wedge \cdots \wedge T_{i_{k+1}}. \end{aligned}$$

Hence,

$$T = \frac{1}{k!} \sum_{i_1,\dots,i_k} a_{i_1 \cdots i_k} T_{i_1} \wedge T_{i_2} \wedge \cdots \wedge T_{i_k}.$$

However, the tensors $T_{i_1} \wedge \cdots \wedge T_{i_k}$ are not linearly independent. Indeed, due to anticommutativity, if two sequences $(i_1, \dots i_k)$ and $(j_1, \dots j_k)$ differ only in their orderings, then $T_{i_1} \wedge \cdots \wedge T_{i_k} = \pm T_{j_1} \wedge \cdots \wedge T_{j_k}$. In addition, if any two of the indices are equal, then $T_{i_1} \wedge \cdots \wedge T_{i_k} = 0$. Hence, we can avoid repeating terms by considering only increasing index sequences:

$$T = \sum_{i_1 < \cdots < i_k} b_{i_1 \cdots i_k} T_{i_1} \wedge \cdots \wedge T_{i_k}$$

and so the set $\{T_{i_1} \wedge \cdots \wedge T_{i_k} \mid 1 \leq i_1 < \ldots < i_k \leq n\}$ spans $\Lambda^k(V^*)$. Moreover, the elements of this set are linearly independent. Indeed, if

$$0 = T = \sum_{i_1 < \cdots < i_k} b_{i_1 \cdots i_k} T_{i_1} \wedge \cdots \wedge T_{i_k},$$

then, taking a basis $\{v_1, \ldots, v_n\}$ of V dual to $\{T_1, \ldots, T_n\}$ and an increasing index sequence $(j_1, \ldots, j_k)$, we have

$$\begin{aligned} 0 = T(v_{j_1}, \ldots, v_{j_k}) &= k! \sum_{i_1 < \cdots < i_k} b_{i_1 \cdots i_k} \operatorname{Alt}(T_{i_1} \otimes \cdots \otimes T_{i_k})(v_{j_1}, \ldots, v_{j_k}) \\ &= \sum_{i_1 < \cdots < i_k} b_{i_1 \cdots i_k} \sum_{\sigma \in S_k} (\operatorname{sgn} \sigma)\, T_{i_1}(v_{j_{\sigma(1)}}) \cdots T_{i_k}(v_{j_{\sigma(k)}}). \end{aligned}$$

Since $(i_1, \ldots, i_k)$ and $(j_1, \ldots, j_k)$ are both increasing, the only term of the second sum that may be different from zero is the one for which $\sigma = \mathrm{id}$. Consequently,

$$0 = T(v_{j_1}, \ldots, v_{j_k}) = b_{j_1 \cdots j_k}.$$

□

The following result is clear from the anticommutativity shown in Example 1.6.

Proposition 1.10. *If $T \in \Lambda^k(V^*)$ and $S \in \Lambda^m(V^*)$, then*

$$T \wedge S = (-1)^{km} S \wedge T.$$

Remark 1.11.

(1) Another consequence of Theorem 1.9 is that $\dim(\Lambda^n(V^*)) = 1$. Hence, if $V = \mathbb{R}^n$, any alternating n-tensor in $\mathbb{R}^n$ is a multiple of the determinant.
(2) It is also clear that $\Lambda^k(V^*) = 0$ if $k > n$. Moreover, the set $\Lambda^0(V^*)$ is defined to be equal to $\mathbb{R}$ (identified with the set of constant functions on V).

A linear transformation $F : V \to W$ induces a linear transformation $F^* : \mathcal{T}^k(W^*) \to \mathcal{T}^k(V^*)$ defined by

$$(F^*T)(v_1, \ldots, v_k) = T(F(v_1), \ldots, F(v_k)).$$

If $T \in \Lambda^k(W^*)$, the tensor F^*T is an alternating tensor on V. It is easy to check that

$$F^*(T \otimes S) = (F^*T) \otimes (F^*S)$$

for $T \in \mathcal{T}^k(W^*)$ and $S \in \mathcal{T}^m(W^*)$. One can then easily show that if T and S are alternating, then

$$F^*(T \wedge S) = (F^*T) \wedge (F^*S).$$

Another important fact about alternating tensors is the following.

Theorem 1.12. *Let $F : V \to V$ be a linear map and let $T \in \Lambda^n(V^*)$. Then $F^*T = (\det A)T$, where A is any matrix representing F.*

PROOF. As $\Lambda^n(V^*)$ is 1-dimensional and F is a linear map, F^* is just multiplication by some constant C. Let us consider an isomorphism H between V and $\mathbb{R}^n$. Then, $H^* \det$ is an alternating n-tensor in V, and so $F^*H^* \det = CH^* \det$. Hence, by Exercise 1.14.4,

$$(H^{-1})^* F^* H^* \det = C \det \Leftrightarrow (H \circ F \circ H^{-1})^* \det = C \det \Leftrightarrow A^* \det = C \det,$$

where A is the matrix representation of F induced by H. Taking the standard basis in $\mathbb{R}^n$, $\{e_1, \dots, e_n\}$, we have

$$A^* \det(e_1, \dots, e_n) = C \det(e_1, \dots, e_n) = C,$$

and so

$$\det(Ae_1, \dots, Ae_n) = C,$$

implying that $C = \det A$. □

REMARK 1.13. By the above theorem, if $T \in \Lambda^n(V^*)$ and $T \neq 0$, then two ordered basis $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_n\}$ are equivalently oriented if and only if $T(v_1, \dots, v_n)$ and $T(w_1, \dots, w_n)$ have the same sign.

EXERCISES 1.14.

(1) Show that the tensor product is bilinear and associative but not commutative.
(2) Find a basis for the space $\mathcal{T}^{k,m}(V^*, V)$ of mixed (k,m)-tensors.
(3) If $T \in \mathcal{T}^k(V^*)$, show that
  (a) $\mathrm{Alt}(T)$ is an alternating tensor;
  (b) if T is alternating then $\mathrm{Alt}(T) = T$;
  (c) $\mathrm{Alt}(\mathrm{Alt}(T)) = \mathrm{Alt}(T)$.
(4) Let $F : V_1 \to V_2$, and $H : V_2 \to V_3$ be two linear maps between vector spaces. Show that:
  (a) $(H \circ F)^* = F^* \circ H^*$;
  (b) for $T \in \Lambda^k(V_2^*)$ and $S \in \Lambda^m(V_2^*)$, $F^*(T \wedge S) = F^*T \wedge F^*S$.
(5) Prove Proposition 1.10.
(6) Let $T_1, \dots, T_k \in V^*$. Show that

$$(T_1 \wedge \cdots \wedge T_k)(v_1, \dots, v_k) = \det[T_i(v_j)].$$

(7) Let $T_1, \dots, T_k \in \Lambda^1(V^*) = V^*$. Show that they are linearly independent if and only if $T_1 \wedge \cdots \wedge T_k \neq 0$.
(8) Let $T \in \Lambda^k(V^*)$ and let $v \in V$. We define **contraction** of T by v, $\iota(v)T$, as the $(k-1)$-tensor given by

$$(\iota(v)T)(v_1, \dots, v_{k-1}) = T(v, v_1, \dots, v_{k-1}).$$

Show that:
  (a) $\iota(v_1)(\iota(v_2)T) = -\iota(v_2)(\iota(v_1)T)$;
  (b) if $T \in \Lambda^k(V^*)$ and $S \in \Lambda^m(V^*)$ then

$$\iota(v)(T \wedge S) = (\iota(v)T) \wedge S + (-1)^k T \wedge (\iota(v)S).$$

2. Tensor Fields

The definition of vector field can be generalized to tensor fields of general type. For that, we denote by T_p^*M the dual of the tangent space T_pM at a point p in M (usually called the **cotangent space** to M at p).

DEFINITION 2.1. *A* (k,m)-**tensor field** *is a map that to each point* $p \in M$ *assigns a tensor* $T \in \mathcal{T}^{k,m}(T_p^*M, T_pM)$.

EXAMPLE 2.2. A vector field is a $(0,1)$-tensor field (or a 1-contravariant tensor field), that is, a map that to each point $p \in M$ assigns the 1-contravariant tensor $X_p \in T_pM$.

EXAMPLE 2.3. Let $f : M \to \mathbb{R}$ be a differentiable function. We can define a $(1,0)$-tensor field df which carries each point $p \in M$ to $(df)_p$, where

$$(df)_p : T_pM \to \mathbb{R}$$

is the derivative of f at p. This tensor field is called the **differential** of f. For any $v \in T_pM$ we have $(df)_p(v) = v \cdot f$ (the directional derivative of f at p along the vector v). Considering a coordinate system $x : W \to \mathbb{R}^n$, we can write $v = \sum_{i=1}^n v^i \left(\frac{\partial}{\partial x^i}\right)_p$, and so

$$(df)_p(v) = \sum_i v^i \frac{\partial \hat{f}}{\partial x^i}(x(p)),$$

where $\hat{f} = f \circ x^{-1}$. Taking the projections $x^i : W \to \mathbb{R}$, we can obtain 1-forms dx^i defined on W. These satisfy

$$(dx^i)_p \left(\left(\frac{\partial}{\partial x^j} \right)_p \right) = \delta_{ij}$$

and so they form a basis of each cotangent space T_p^*M, dual to the coordinate basis $\left\{ \left(\frac{\partial}{\partial x^1}\right)_p, \cdots, \left(\frac{\partial}{\partial x^n}\right)_p \right\}$ of T_pM. Hence, any $(1,0)$-tensor field on W can be written as $\omega = \sum_i \omega_i dx^i$, where $\omega_i : W \to \mathbb{R}$ is such that $\omega_i(p) = \omega_p(\left(\frac{\partial}{\partial x^i}\right)_p)$. In particular, df can be written in the usual way

$$(df)_p = \sum_{i=1}^n \frac{\partial \hat{f}}{\partial x^i}(x(p))(dx^i)_p.$$

REMARK 2.4. Similarly to what was done for the tangent bundle, we can consider the disjoint union of all cotangent spaces and obtain the manifold

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

called the **cotangent bundle** of M. Note that a $(1,0)$-tensor field is just a map from M to T^*M defined by

$$p \mapsto \omega_p \in T_p^*M.$$

This construction can be easily generalized for arbitrary tensor fields.

The space of (k,m)-tensor fields is clearly a vector space since linear combinations of (k,m)-tensors are still (k,m)-tensors. If W is a coordinate neighborhood of M, we know that $\left\{(dx^i)_p\right\}$ is a basis for T_p^*M and that $\left\{\left(\frac{\partial}{\partial x^i}\right)_p\right\}$ is a basis for T_pM. Hence, the value of a (k,m)-tensor field T at a point $p \in W$ can be written as the tensor

$$T_p = \sum a_{i_1\cdots i_k}^{j_1\cdots j_m}(p)(dx^{i_1})_p \otimes \cdots \otimes (dx^{i_k})_p \otimes \left(\frac{\partial}{\partial x^{j_1}}\right)_p \otimes \cdots \otimes \left(\frac{\partial}{\partial x^{j_m}}\right)_p$$

where the $a_{i_1\cdots i_k}^{j_1\cdots j_m} : W \to \mathbb{R}$ are functions which at each $p \in W$ give us the components of T_p relative to these bases of T_p^*M and T_pM. Just as we did with vector fileds, we say that a tensor field is **differentiable** if all these functions are differentiable for all coordinate sytems of the maximal atlas. Again, we only need to consider the coordinate sytems of an atlas, since all overlap maps are differentiable (cf. Exercise 2.8.1).

EXAMPLE 2.5. The differential of a smooth function $f : M \to \mathbb{R}$ is clearly a differentiable $(1,0)$-tensor field, since its components $\frac{\partial \hat{f}}{\partial x^i} \circ x$ on a given coordinate system $x : W \to \mathbb{R}^n$ are smooth.

An important operation on covariant tensors is the pullback by a smooth map.

DEFINITION 2.6. *Let $f : M \to N$ be a differentiable map between smooth manifolds. Then, each differentiable k-covariant tensor field T on N defines a k-covariant tensor field f^*T on M in the following way:*

$$(f^*T)_p(v_1,\ldots,v_k) = T_{f(p)}((df)_p v_1,\ldots,(df)_p v_k),$$

for $v_1,\ldots,v_k \in T_pM$.

REMARK 2.7. Notice that $(f^*T)_p$ is just the image of $T_{f(p)}$ by the linear map $(df)_p^* : \mathcal{T}^k(T_{f(p)}^*N) \to \mathcal{T}^k(T_p^*M)$ induced by $(df)_p : T_pM \to T_{f(p)}N$ (cf. Section 1). Therefore the properties $f^*(\alpha T + \beta S) = \alpha(f^*T) + \beta(f^*S)$ and $f^*(T \otimes S) = (f^*T) \otimes (f^*S)$ hold for all $\alpha, \beta \in \mathbb{R}$ and all appropriate covariant tensor fields T, S. We will see in Exercise 2.8.2 that the pull-back of a differentiable covariant tensor field is still a differentiable covariant tensor field.

EXERCISES 2.8.

(1) Find the relation between coordinate functions of a tensor field in two overlapping coordinate systems.
(2) Show that the pull-back of a differentiable covariant tensor field is still a differentiable covariant tensor field.
(3) (*Lie derivative of a tensor field*) Given a vector field $X \in \mathfrak{X}(M)$, we define the **Lie derivative of a k-covariant tensor field** T along X as

$$L_X T := \frac{d}{dt}({\psi_t}^* T)_{|t=0},$$

where $\psi_t = F(\cdot, t)$ with F the local flow of X at p.

(a) Show that

$$L_X\left(T(Y_1,\ldots,Y_k)\right) = L_X T(Y_1,\ldots,Y_k)$$
$$+ T(L_X Y_1,\ldots,Y_k) + \ldots + T(Y_1,\ldots,L_X Y_k),$$

i.e., show that

$$X \cdot (T(Y_1,\ldots,Y_k)) = L_X T(Y_1,\ldots,Y_k)$$
$$+ T([X,Y_1],\ldots,Y_k) + \ldots + T(Y_1,\ldots,[X,Y_k]),$$

for all vector fields $Y_1,\ldots,Y_k$ (cf. Exercises 6.11.10 and 6.11.11 in Chapter 1).

(b) How would you define the Lie derivative of a (k,m)-tensor field?

3. Differential Forms

Fields of alternating tensors are very important objects called **forms**.

DEFINITION 3.1. *Let M be a smooth manifold. A* **form of degree** *k (or k-***form***) on M is a field of alternating k-tensors defined on M, that is, a map ω that, to each point $p \in M$, assigns an element $\omega_p \in \Lambda^k(T_p^*M)$.*

The space of k-forms on M is clearly a vector space. By Theorem 1.9, given a coordinate system $x : W \to \mathbb{R}^n$, any k-form on W can be written as

$$\omega = \sum_I \omega_I dx^I$$

where $I = (i_1,\ldots,i_k)$ denotes any increasing index sequence of integers in $\{1,\ldots,n\}$, dx^I is the form $dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, and the ω_I's are functions defined on W. It is easy to check that the components of ω in the basis $\{dx^{i_1} \otimes \cdots \otimes dx^{i_k}\}$ are $\pm\omega_I$. Therefore ω is a differentiable $(k,0)$-tensor (in which case it is called a **differential form**) if the functions ω_I are smooth for all coordinate systems of the maximal atlas. The set of differential k-forms on M is represented by $\Omega^k(M)$. From now on we will use the word "form" to mean a differential form.

Given a smooth map $f : M \to N$ between differentiable manifolds, we can induce forms on M from forms on N using the pull-back operation (cf. Definition 2.6), since the pull-back of a field of alternating tensors is still a field of alternating tensors.

REMARK 3.2. If $g : N \to \mathbb{R}$ is a 0-form, that is, a function, the pullback is defined as $f^*g = g \circ f$.

It is easy to verify that the pullback of forms satisfies the following properties, the proof of which we leave as an exercise:

PROPOSITION 3.3. *Let $f : M \to N$ be a differentiable map and α, β forms on N. Then,*

(i) $f^*(\alpha+\beta) = f^*\alpha + f^*\beta$*;*
(ii) $f^*(g\alpha) = (g\circ f)f^*\alpha = (f^*g)(f^*\alpha)$ *for any function* $g : N \to \mathbb{R}$*;*
(iii) $f^*(\alpha\wedge\beta) = f^*\alpha \wedge f^*\beta$*;*
(iv) $g^*f^*\alpha = (f\circ g)^*\alpha$ *for any differentiable map* $g : L \to M$*.*

Example 3.4. If $f : M \to N$ is differentiable and we consider coordinate systems $x : V \to \mathbb{R}^m$, $y : W \to \mathbb{R}^n$ respectively on M and N, we have $y^i = \hat{f}^i(x^1, \dots, x^m)$ for $i = 1, \dots, n$ and $\hat{f} = y \circ f \circ x^{-1}$ the local representation of f. If $\omega = \sum_I \omega_I dy^I$ is a k-form on W, then by Proposition 3.3,

$$f^*\omega = f^*\left(\sum_I \omega_I dy^I\right) = \sum_I (f^*\omega_I)(f^*dy^I) = \sum_I (\omega_I \circ f) f^*dy^{i_1} \wedge \cdots \wedge f^*dy^{i_k}.$$

Moreover, for $v \in T_pM$,

$$(f^*(dy^i))_p(v) = (dy^i)_{f(p)}((df)_p v) = \big(d(y^i\circ f)\big)_p (v),$$

that is, $f^*(dy^i) = d(y^i \circ f)$. Hence,

$$\begin{aligned} f^*\omega &= \sum_I (\omega_I \circ f)\, d(y^{i_1}\circ f) \wedge \cdots \wedge d(y^{i_k} \circ f) \\ &= \sum_I (\omega_I \circ f)\, d(\hat{f}^{i_1}\circ x) \wedge \cdots \wedge d(\hat{f}^{i_k} \circ x). \end{aligned}$$

If $k = \dim M = \dim N = n$, then the pullback $f^*\omega$ can easily be computed from Theorem 1.12, according to which

$$(f^*(dy^1 \wedge \cdots \wedge dy^n))_p = \det (d\hat{f})_{x(p)} (dx^1 \wedge \cdots \wedge dx^n)_p. \tag{4}$$

Given any form ω on M and a parametrization $\varphi : U \to M$, we can consider the pullback of ω by φ and obtain a form defined on the open set U, called the **representation** of ω on that parametrization.

Example 3.5. Let $x : W \to \mathbb{R}^n$ be a coordinate system on a smooth manifold M and consider the 1-form dx^i defined on W. The pullback φ^*dx^i by the corresponding parametrization $\varphi := x^{-1}$ is a 1-form on an open subset U of $\mathbb{R}^n$ satisfying

$$\begin{aligned} (\varphi^*dx^i)_x(v) &= (\varphi^*dx^i)_x \left(\sum_{j=1}^n v^j \left(\frac{\partial}{\partial x^j}\right)_x\right) = (dx^i)_p \left(\sum_{j=1}^n v^j (d\varphi)_x \left(\frac{\partial}{\partial x^j}\right)_x\right) \\ &= (dx^i)_p \left(\sum_{j=1}^n v^j \left(\frac{\partial}{\partial x^j}\right)_p\right) = v^i = (dx^i)_x(v), \end{aligned}$$

for $x \in U$, $p = \varphi(x)$ and $v = \sum_{j=1}^n v^j \left(\frac{\partial}{\partial x^j}\right)_x \in T_xU$. Hence, just as we had $\left(\frac{\partial}{\partial x^i}\right)_p = (d\varphi)_x \left(\frac{\partial}{\partial x^i}\right)_x$, we now have $(dx^i)_x = \varphi^*(dx^i)_p$, and so $(dx^i)_p$ is the 1-form in W whose representation on U is $(dx^i)_x$.

If $\omega = \sum_I \omega_I dx^I$ is a k-form defined on an open subset of $\mathbb{R}^n$, we define a $(k+1)$-form called **exterior derivative** of ω as

$$d\omega := \sum_I d\omega_I \wedge dx^I.$$

EXAMPLE 3.6. Consider the form $\omega = -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy$ defined on $\mathbb{R}^2\backslash\{0\}$. Then,

$$\begin{aligned} d\omega &= d\left(-\frac{y}{x^2+y^2}\right) \wedge dx + d\left(\frac{x}{x^2+y^2}\right) \wedge dy \\ &= \frac{y^2-x^2}{(x^2+y^2)^2}\,dy \wedge dx + \frac{y^2-x^2}{(x^2+y^2)^2}\,dx \wedge dy = 0. \end{aligned}$$

The exterior derivative satisfies the following properties:

PROPOSITION 3.7. *If $\alpha, \omega, \omega_1, \omega_2$ are forms on $\mathbb{R}^n$, then*

(i) $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$;
(ii) if ω is k-form, $d(\omega \wedge \alpha) = d\omega \wedge \alpha + (-1)^k \omega \wedge d\alpha$;
(iii) $d(d\omega) = 0$;
(iv) if $f : \mathbb{R}^m \to \mathbb{R}^n$ is smooth, $d(f^*\omega) = f^*(d\omega)$.

PROOF. Property (i) is obvious. Using (i), it is enough to prove (ii) for $\omega = a_I dx^I$ and $\alpha = b_J dx^J$:

$$\begin{aligned} d(\omega \wedge \alpha) &= d(a_I b_J\, dx^I \wedge dx^J) = d(a_I b_J) \wedge dx^I \wedge dx^J \\ &= (b_J\, da_I + a_I\, db_J) \wedge dx^I \wedge dx^J \\ &= b_J\, da_I \wedge dx^I \wedge dx^J + a_I\, db_J \wedge dx^I \wedge dx^J \\ &= d\omega \wedge \alpha + (-1)^k a_I dx^I \wedge db_J \wedge dx^J \\ &= d\omega \wedge \alpha + (-1)^k \omega \wedge d\alpha. \end{aligned}$$

Again, to prove (iii), it is enough to consider forms $\omega = a_I dx^I$. Since

$$d\omega = da_I \wedge dx^I = \sum_{i=1}^n \frac{\partial a_I}{\partial x^i}\, dx^i \wedge dx^I,$$

we have

$$\begin{aligned} d(d\omega) &= \sum_{j=1}^n \sum_{i=1}^n \frac{\partial^2 a_I}{\partial x^j \partial x^i} dx^j \wedge dx^i \wedge dx^I \\ &= \sum_{i=1}^n \sum_{j<i} \left(\frac{\partial^2 a_I}{\partial x^j \partial x^i} - \frac{\partial^2 a_I}{\partial x^i \partial x^j}\right) dx^j \wedge dx^i \wedge dx^I = 0. \end{aligned}$$

To prove (iv), we first consider a 0-form g:

$$\begin{aligned} f^*(dg) &= f^*\left(\sum_{i=1}^n \frac{\partial g}{\partial x^i} dx^i\right) = \sum_{i=1}^n \left(\frac{\partial g}{\partial x^i} \circ f\right) df^i = \sum_{i,j=1}^n \left(\left(\frac{\partial g}{\partial x^i} \circ f\right) \frac{\partial f^i}{\partial x^j}\right) dx^j \\ &= \sum_{j=1}^n \frac{\partial (g \circ f)}{\partial x^j} dx^j = d(g \circ f) = d(f^* g). \end{aligned}$$

Then, if $\omega = a_I dx^I$, we have

$$\begin{aligned} d(f^*\omega) &= d(a_I \circ f) \wedge df^I + (a_I \circ f) d(df^I) = d(a_I \circ f) \wedge df^I = d(f^* a_I) \wedge df^I \\ &= (f^* da_I) \wedge df^I = f^*(da_I \wedge dx^I) = f^*(d\omega) \end{aligned}$$

(where df^I denotes the form $df^{i_1} \wedge \cdots \wedge df^{i_k}$), and the result follows. □

If we consider two parametrizations $\varphi_\alpha : U_\alpha \to M$, $\varphi_\beta : U_\beta \to M$ such that $\varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta) = W \neq \varnothing$, and take the corresponding representations $\omega_\alpha := \varphi_\alpha^* \omega$ and $\omega_\beta := \varphi_\beta^* \omega$ of a k-form ω, it is easy to verify that

$$(\varphi_\alpha^{-1} \circ \varphi_\beta)^* \omega_\alpha = \omega_\beta.$$

Suppose now that ω is a differential k-form on a smooth manifold M. We define the $(k+1)$-form $d\omega$ as the smooth form that is locally represented by $d\omega_\alpha$, that is, for each parametrization $\varphi_\alpha : U_\alpha \to M$, the form $d\omega$ is defined on $\varphi_\alpha(U)$, as $(\varphi_\alpha^{-1})^*(d\omega_\alpha)$. Given another parametrization $\varphi_\beta : U_\beta \to M$ such that $\varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta) = W \neq \varnothing$, then, setting f equal to $\varphi_\alpha^{-1} \circ \varphi_\beta$, we have

$$f^*(d\omega_\alpha) = d(f^*\omega_\alpha) = d\omega_\beta.$$

Consequently,

$$\begin{aligned} (\varphi_\beta^{-1})^* d\omega_\beta &= (\varphi_\beta^{-1})^* f^*(d\omega_\alpha) \\ &= (f \circ \varphi_\beta^{-1})^*(d\omega_\alpha) \\ &= (\varphi_\alpha^{-1})^*(d\omega_\alpha), \end{aligned}$$

and so the two definitions agree on the overlapping set W. Therefore $d\omega$ is well defined. We leave it as an exercise to show that the exterior derivative defined for forms on smooth manifolds also satisfies the properties of Proposition 3.7.

Exercises 3.8.

(1) Prove Proposition 3.3.
(2) (*Exterior derivative*) Let M be a smooth manifold. Given a k-form ω in M we can define its exterior derivative $d\omega$ without using local coordinates: given $k+1$ vector fields $X_1, \ldots, X_{k+1} \in \chi(M)$,

$$\begin{aligned} d\omega(X_1, \ldots, X_{k+1}) &:= \sum_{i=1}^{k+1} (-1)^{i-1} X_i \cdot \omega(X_1, \ldots, \hat{X}_i, \ldots, X_{k+1}) + \\ &\sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], X_1, \ldots, \hat{X}_i, \ldots, \hat{X}_j, \ldots, X_{k+1}), \end{aligned}$$

where the hat indicates an omitted variable.

(a) Show that $d\omega$ defined above is in fact a $(k+1)$-form in M, that is,
   (i) $d\omega(X_1, \ldots, X_i + Y_i, \ldots, X_{k+1}) =$
   $d\omega(X_1, \ldots, X_i, \ldots, X_{k+1}) + d\omega(X_1, \ldots, Y_i, \ldots, X_{k+1})$;

(ii) $d\omega(X_1, \dots, fX_j, \dots, X_{k+1}) = f d\omega(X_1, \dots, X_{k+1})$ for any differentiable function f;

(iii) $d\omega$ is alternating.

(b) Let $x : W \to \mathbb{R}^n$ be a coordinate system of M and let $\omega = \sum_I a_I dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ be the expression of ω in these coordinates (where the a_I's are smooth functions). Show that the local expression of $d\omega$ is the same as the one used in the local definition of exterior derivative, that is,

$$d\omega = \sum_I da_I \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

(3) Show that the exterior derivative defined for forms on smooth manifolds satisfies the properties of Proposition 3.7.

(4) Show that:

(a) if $\omega = f^1 dx + f^2 dy + f^3 dz$ is a 1-form on $\mathbb{R}^3$ then

$$d\omega = g^1 dy \wedge dz + g^2 dz \wedge dx + g^3 dx \wedge dy,$$

where $(g^1, g^2, g^3) = \operatorname{curl}(f^1, f^2, f^3)$;

(b) if $\omega = f^1 dy \wedge dz + f^2 dz \wedge dx + f^3 dx \wedge dy$ is a 2-form on $\mathbb{R}^3$, then

$$d\omega = \operatorname{div}(f^1, f^2, f^3)\, dx \wedge dy \wedge dz.$$

(5) (*De Rham cohomology*) A k-form ω is called **closed** if $d\omega = 0$. If it exists a $(k-1)$-form β such that $\omega = d\beta$ then ω is called **exact**. Note that every exact form is closed. Let Z^k be the set of all closed k-forms on M and define a relation between forms on Z^k as follows: $\alpha \sim \beta$ if and only if they differ by an exact form, that is, if $\beta - \alpha = d\theta$ for some $(k-1)$-form θ.

(a) Show that this relation is an equivalence relation.

(b) Let $H^k(M)$ be the corresponding set of equivalence classes (called the k-dimensional **de Rham cohomology space** of M). Show that addition and scalar multiplication of forms define indeed a vector space structure on $H^k(M)$.

(c) Let $f : M \to N$ be a smooth map. Show that:

(i) the pullback f^* carries closed forms to closed forms and exact forms to exact forms;

(ii) if $\alpha \sim \beta$ on N then $f^*\alpha \sim f^*\beta$ on M;

(iii) f^* induces a linear map on cohomology $f^\sharp : H^k(N) \to H^k(M)$ naturally defined by $f^\sharp[\omega] = [f^*\omega]$;

(iv) if $g : L \to M$ is another smooth map, then $(f \circ g)^\sharp = g^\sharp \circ f^\sharp$.

(d) Show that the dimension of $H^0(M)$ is equal to the number of connected components of M.

(e) Show that $H^k(M) = 0$ for every $k > \dim M$.

(6) Let M be a manifold of dimension n, let U be an open subset of $\mathbb{R}^n$ and let ω be a k-form on $\mathbb{R} \times U$. Writing ω as

$$\omega = dt \wedge \sum_I a_I dx^I + \sum_J b_J dx^J,$$

where $I = (i_1, \ldots, i_{k-1})$ and $J = (j_1, \ldots, j_k)$ are increasing index sequences, $(x^1, \ldots, x^n)$ are coordinates in U and t is the coordinate in $\mathbb{R}$, consider the operator $\mathcal{Q}$ defined by

$$\mathcal{Q}(\omega)_{(t,x)} = \sum_I \left(\int_{t_0}^t a_I ds \right) dx^I,$$

which transforms k-forms ω in $\mathbb{R} \times U$ into $(k-1)$-forms.

(a) Let $f : V \to U$ be a diffeomorphism between open subsets of $\mathbb{R}^n$. Show that the induced diffeomorphism $\tilde{f} := \mathrm{id} \times f : \mathbb{R} \times V \to \mathbb{R} \times U$ satisfies

$$\tilde{f}^* \circ \mathcal{Q} = \mathcal{Q} \circ \tilde{f}^*.$$

(b) Using (a), construct an operator $\mathcal{Q}$ which carries k-forms on $\mathbb{R} \times M$ into $(k-1)$-forms and, for any diffeomorphism $f : M \to N$, the induced diffeomorphism $\tilde{f} := \mathrm{id} \times f : \mathbb{R} \times M \to \mathbb{R} \times N$ satisfies $\tilde{f}^* \circ \mathcal{Q} = \mathcal{Q} \circ \tilde{f}^*$. Show that this operator is linear.

(c) Considering the operator $\mathcal{Q}$ defined in (b) and the inclusion $i_{t_0} : M \to \mathbb{R} \times M$ of M at the "level" t_0, defined by $i_{t_0}(p) = (t_0, p)$, show that $\omega - \pi^* i_{t_0}^* \omega = d\mathcal{Q}\omega + \mathcal{Q}d\omega$, where $\pi : \mathbb{R} \times M \to M$ is the projection on M.

(d) Show that the maps $\pi^\sharp : H^k(M) \to H^k(\mathbb{R} \times M)$ and $i_{t_0}^\sharp : H^k(\mathbb{R} \times M) \to H(M)$ are inverses of each other (and so $H^k(M)$ is isomorphic to $H^k(\mathbb{R} \times M)$).

(e) Use (d) to show that, for $k > 0$ and $n > 0$, every closed k-form in $\mathbb{R}^n$ is exact, that is, $H^k(\mathbb{R}^n) = 0$ if $k > 0$.

(f) Use (d) to show that, if $f, g : M \to N$ are two **smoothly homotopic maps** between smooth manifolds (meaning that there exists a smooth map $H : \mathbb{R} \times M \to N$ such that $H(t_0, p) = f(p)$ and $H(t_1, p) = g(p)$ for some fixed $t_0, t_1 \in \mathbb{R}$), then $f^\sharp = g^\sharp$.

(g) We say that M is **contractible** if the identity map $\mathrm{id} : M \to M$ is smoothly homotopic to a constant map. Show that $\mathbb{R}^n$ is contractible.

(h) (*Poincaré Lemma*) Let M be a contractible smooth manifold. Show that every closed form on M is exact, that is, $H^k(M) = 0$ for all $k > 0$.

(7) (*Symplectic manifold*) A **symplectic manifold** (M, ω), is a manifold M equiped with a closed non-degenerate 2-form ω. Here

non-degenerate means that the map that to each tangent vector $X_p \in T_pM$ associates the 1-tensor in T_pM defined by $\iota(X_p)\omega_p$ is a bijection (cf. Exercise 1.14.8).

(a) Show that $\dim M$ is necessarily even.

(b) Consider coordinates $(x^1, \dots, x^n, y^1, \dots, y^n)$ in $\mathbb{R}^{2n}$, and the differential form $\omega_0 = \sum_{i=1}^n dx^i \wedge dy^i$. Show that $(\mathbb{R}^{2n}, \omega_0)$ is a symplectic manifold and compute the wedge product ω_0^n, of n copies of ω_0. (**Remark:** The form ω_0 is called the **standard symplectic form**. This example gives us a local model for **all** symplectic manifolds – **Darboux Theorem**).

(8) (*Lie derivative of a differential form*) Given a vector field $X \in \mathfrak{X}(M)$, we define the **Lie derivative of a form** ω along X as

$$L_X\omega := \frac{d}{dt}({\psi_t}^*\omega)_{|t=0},$$

where $\psi_t = F(\cdot, t)$ with F the local flow of X at p (cf. Exercise 2.8.3). Show that the Lie derivative satisfies the following properties:

(a) $L_X(\omega_1 \wedge \omega_2) = (L_X\omega_1) \wedge \omega_2 + \omega_1 \wedge (L_X\omega_2)$;

(b) $d(L_X\omega) = L_X(d\omega)$;

(c) **Cartan formula:** $L_X\omega = \iota(X)d\omega + d(\iota(X)\omega)$;

(d) $L_X(\iota(Y)\omega) = \iota(L_XY)\omega + \iota(Y)L_X\omega$

(cf. Exercise 6.11.11 on Chapter 1 and Exercise 1.14.8).

4. Integration on Manifolds

Before we see how to integrate differential forms on manifolds, we will start by studying the $\mathbb{R}^n$ case. For that let us consider an n-form ω defined on an open subset U of $\mathbb{R}^n$. We already know that ω can be written as

$$\omega_x = a(x)\, dx^1 \wedge \cdots \wedge dx^n,$$

where $a : U \to \mathbb{R}$ is a smooth function. The **support** of ω is, by definition, the closure of the set where $\omega \neq 0$ that is,

$$\operatorname{supp}\omega = \overline{\{x \in \mathbb{R}^n : \omega_x \neq 0\}}.$$

We will assume that this set is compact (in which case ω is said to be **compactly supported**) and is a subset of U. We define

$$\int_U \omega = \int_U a(x)\, dx^1 \wedge \cdots \wedge dx^n := \int_U a(x)\, dx^1 \cdots dx^n,$$

where the integral on the right is a multiple integral on a subset of $\mathbb{R}^n$. This definition is almost well-behaved with respect to changes of variables in $\mathbb{R}^n$. Indeed, if $f : V \to U$ is a diffeomorphism of open sets of $\mathbb{R}^n$, we have from (4) that

$$f^*\omega = (a \circ f)(\det df) dy^1 \wedge \cdots \wedge dy^n,$$

and so

$$\int_V f^*\omega = \int_V (a \circ f)(\det df) dy^1 \cdots dy^n.$$

If f is orientation preserving, then $\det(df) > 0$, and the integral on the right is, by the Change of Variables Theorem for multiple integrals in $\mathbb{R}^n$ (cf. Section 7.2), equal to $\int_U \omega$. For this reason, we will only consider orientable manifolds when integrating forms on manifolds. Moreover, we will also assume that $\operatorname{supp}\omega$ is always compact to avoid convergence problems.

Let M be an oriented manifold, and let $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ be an atlas whose parametrizations are orientation-preserving. Suppose that $\operatorname{supp}\omega$ is contained in some coordinate neighborhood $W_\alpha = \varphi_\alpha(U_\alpha)$. Then we define

$$\int_M \omega := \int_{U_\alpha} \varphi_\alpha^*\omega = \int_{U_\alpha} \omega_\alpha.$$

Note that this does not depend on the choice of coordinate neighborhood: if $\operatorname{supp}\omega$ is contained in some other coordinate neighborhood $W_\beta = \varphi_\beta(U_\beta)$, then $\omega_\beta = f^*\omega_\alpha$, where $f = \varphi_\alpha^{-1} \circ \varphi_\beta$, and hence

$$\int_{U_\beta} \omega_\beta = \int_{U_\beta} f^*\omega_\alpha = \int_{U_\alpha} \omega_\alpha.$$

To define the integral in the general case we use a **partition of unity** (cf. Section 7.2) subordinate to the cover $\{W_\alpha\}$ of M, i.e., a family of differentiable functions on M, $\{\rho_i\}_{i\in I}$, such that:

(i) for every point $p \in M$, there exists a neighborhood V of p such that $V \cap \operatorname{supp}\rho_i = \varnothing$ except for a finite number of ρ_i's;
(ii) for every point $p \in M$, $\sum_{i\in I} \rho_i(p) = 1$;
(iii) $0 \le \rho_i \le 1$ and $\operatorname{supp}\rho_i \subset W_{\alpha_i}$ for some element W_{α_i} of the cover.

Because of property (i), $\operatorname{supp}\omega$ (being compact) intersects the supports of only finitely many ρ_i's. Hence we can assume that I is finite, and then

$$\omega = \left(\sum_{i\in I} \rho_i\right)\omega = \sum_{i\in I} \rho_i\omega = \sum_{i\in I} \omega_i$$

with $\omega_i = \rho_i\omega$ and $\operatorname{supp}\omega_i \subset W_{\alpha_i}$. Consequently we define:

$$\int_M \omega := \sum_{i\in I} \int_M \omega_i = \sum_{i\in I} \int_{U_{\alpha_i}} \varphi_{\alpha_i}^*\omega_i.$$

REMARK 4.1.

(1) When $\operatorname{supp}\omega$ is contained in one coordinate neighborhood W, the two definitions above agree. Indeed,

$$\begin{aligned}\int_M \omega &= \int_W \omega = \int_W \sum_{i\in I} \omega_i = \int_U \varphi^*\left(\sum_{i\in I} \omega_i\right) \\ &= \int_U \sum_{i\in I} \varphi^*\omega_i = \sum_{i\in I} \int_U \varphi^*\omega_i = \sum_{i\in I} \int_M \omega_i,\end{aligned}$$

where we used the linearity of the pullback and of integration on $\mathbb{R}^n$.

(2) The definition of integral is independent of the choice of partition of unity and the choice of cover. Indeed, if $\{\tilde{\rho}_j\}_{j\in J}$ is another partition of unity subordinate to another cover $\{\widetilde{W_\beta}\}$ compatible with the same orientation, we have by (1)

$$\sum_{i\in I}\int_M \rho_i\omega = \sum_{i\in I}\sum_{j\in J}\int_M \tilde{\rho}_j\rho_i\omega$$

and

$$\sum_{j\in J}\int_M \tilde{\rho}_j\omega = \sum_{j\in J}\sum_{i\in I}\int_M \rho_i\tilde{\rho}_j\omega.$$

(3) It is also easy to verify the linearity of the integral, that is,

$$\int_M a\omega_1 + b\omega_2 = a\int_M \omega_1 + b\int_M \omega_2.$$

for $a, b \in \mathbb{R}$ and ω_1, ω_2 n-forms on M.

Exercises 4.2.

(1) Let M be an n-dimensional differentiable manifold. A subset $N \subset M$ is said to have **zero measure** if the sets $\varphi_\alpha^{-1}(N) \subset U_\alpha$ have zero measure for every parametrization $\varphi_\alpha : U_\alpha \to M$ in the maximal atlas.
 (a) Prove that in order to show that $N \subset M$ has zero measure it suffices to check that the sets $\varphi_\alpha^{-1}(N) \subset U_\alpha$ have zero measure for the parametrizations in an arbitrary atlas.
 (b) Suppose that M is oriented. Let $\omega \in \Omega^n(M)$ be compactly supported and let $W = \varphi(U)$ be a coordinate neighborhood such that $M \setminus W$ has zero measure. Show that

$$\int_M \omega = \int_U \varphi^*\omega,$$

 where the integral on the right-hand side is defined as above and always exists.

(2) Let x, y, z be the restrictions of the Cartesian coordinate functions in $\mathbb{R}^3$ to S^2, oriented so that $\{(1,0,0);(0,1,0)\}$ is a positively oriented basis of $T_{(0,0,1)}S^2$, and consider the 2-form

$$\omega = xdy \wedge dz + ydz \wedge dx + zdx \wedge dy \in \Omega^2(S^2).$$

Compute the integral

$$\int_{S^2}\omega$$

using the parametrizations corresponding to
 (a) spherical coordinates;
 (b) stereographic projection.

(3) Let M, N be n-dimensional manifolds, $f : M \to N$ an orientation preserving diffeomorphism and $\omega \in \Omega^n(N)$ a compactly supported form. Prove that

$$\int_N \omega = \int_M f^*\omega.$$

5. Stokes Theorem

In this section we will prove a very important theorem.

THEOREM 5.1. (Stokes) *Let M be an n-dimensional oriented smooth manifold with boundary, let ω be a $(n-1)$-differential form on M with compact support, and let $i : \partial M \to M$ be the inclusion of the boundary ∂M in M. Then*

$$\int_{\partial M} i^*\omega = \int_M d\omega,$$

where we consider ∂M with the the induced orientation (cf. Section 9 in Chapter 1).

PROOF. Let us take a partition of unity $\{\rho_i\}_{i\in I}$ subordinate to an open cover of M by coordinate neighborhoods compatible with the orientation. Then $\omega = \sum_{i\in I} \rho_i\omega$, where we can assume I to be finite (ω is compactly supported), and hence

$$d\omega = d\sum_{i\in I} \rho_i\omega = \sum_{i\in I} d(\rho_i\omega).$$

By linearity of the integral we then have,

$$\int_M d\omega = \sum_{i\in I} \int_M d(\rho_i\omega) \quad \text{and} \quad \int_{\partial M} i^*\omega = \sum_{i\in I} \int_{\partial M} i^*(\rho_i\omega).$$

Hence, to prove this theorem, it is enough to consider the case where $\operatorname{supp}\omega$ is contained inside one coordinate neighborhood of the cover. Let us then consider a $(n-1)$-form ω with compact support contained in a coordinate neighborhood W. Let $\varphi : U \to W$ be the corresponding parametrization, where we can assume U to be bounded ($\operatorname{supp}\varphi^*\omega$ is compact). Then, the representation of ω on U can be written as

$$\varphi^*\omega = \sum_{j=1}^n a_j\, dx^1 \wedge \cdots \wedge dx^{j-1} \wedge dx^{j+1} \wedge \cdots \wedge dx^n,$$

(where each $a_j : U \to \mathbb{R}$ is a C^∞-function), and

$$\varphi^* d\omega = d\varphi^*\omega = \sum_{j=1}^n (-1)^{j-1} \frac{\partial a_j}{\partial x^j}\, dx^1 \wedge \cdots \wedge dx^n.$$

The functions a_j can be extended to C^∞-functions on $\mathbb{H}^n$ by letting

$$a_j(x^1, \cdots, x^n) = \begin{cases} a_j(x^1, \cdots, x^n) & \text{if } (x^1, \ldots, x^n) \in U \\ 0 & \text{if } (x^1, \ldots, x^n) \in \mathbb{H}^n \backslash U. \end{cases}$$

If $W \cap \partial M = \varnothing$, then $i^*\omega = 0$. Moreover, if we consider a rectangle I in $\mathbb{H}$ containing U defined by equations $b_j \leq x^j \leq c_j$ ($j = 1, \ldots, n$), we have

$$\begin{aligned}
\int_M d\omega = \int_U \left(\sum_{j=1}^n (-1)^{j-1} \frac{\partial a_j}{\partial x^j} \right) dx^1 \cdots dx^n = \sum_{j=1}^n (-1)^{j-1} \int_I \frac{\partial a_j}{\partial x^j}\, dx^1 \cdots dx^n \\
= \sum_{j=1}^n (-1)^{j-1} \int_{\mathbb{R}^{n-1}} \left(\int_{b_j}^{c_j} \frac{\partial a_j}{\partial x^j} dx^j \right) dx^1 \cdots dx^{j-1} dx^{j+1} \cdots dx^n \\
= \sum_{j=1}^n (-1)^{j-1} \int_{\mathbb{R}^{n-1}} \big(a_j(x^1, \ldots, x^{j-1}, c_j, x^{j+1}, \ldots, x^n) - \\
- a_j(x^1, \ldots, x^{j-1}, b_j, x^{j+1}, \ldots, x^n) \big)\, dx^1 \cdots dx^{j-1} dx^{j+1} \cdots dx^n = 0,
\end{aligned}$$

where we used Fubini Theorem, the Fundamental Theorem of Calculus and the fact that the a_j's are zero outside U. We conclude that, in this case, $\int_{\partial M} i^*\omega = \int_M d\omega = 0$.

If, on the other hand, $W \cap \partial M \neq \varnothing$ we take a rectangle I containing U now defined by the equations $b_j \leq x^j \leq c_j$ for $j = 1, \ldots, n-1$, and $0 \leq x^n \leq c_n$. Then, as in the preceding case, we have

$$\begin{aligned}
\int_M d\omega = \int_U \left(\sum_{j=1}^n (-1)^{j-1} \frac{\partial a_j}{\partial x^j} \right) dx^1 \cdots dx^n = \sum_{j=1}^n (-1)^{j-1} \int_I \frac{\partial a_j}{\partial x^j}\, dx^1 \cdots dx^n \\
= 0 + (-1)^{n-1} \int_{\mathbb{R}^{n-1}} \left(\int_0^{c_n} \frac{\partial a_n}{\partial x^n} dx^n \right) dx^1 \cdots dx^{n-1} \\
= (-1)^{n-1} \int_{\mathbb{R}^{n-1}} \big(a_n(x^1, \ldots, x^{n-1}, c_n) - a_n(x^1, \ldots, x^{n-1}, 0) \big)\, dx^1 \cdots dx^{n-1} \\
= (-1)^n \int_{\mathbb{R}^{n-1}} a_n(x^1, \ldots, x^{n-1}, 0)\, dx^1 \cdots dx^{n-1}.
\end{aligned}$$

To compute $\int_{\partial M} i^*\omega$ we need to consider a parametrization $\tilde{\varphi}$ of ∂M defined on an open subset of $\mathbb{R}^{n-1}$ which preserves the standard orientation on $\mathbb{R}^{n-1}$ when we consider the induced orientation on ∂M. For that, we can for instance consider the set

$$\widetilde{U} = \{ (x^1, \ldots, x^{n-1}) \in \mathbb{R}^{n-1} \mid ((-1)^n x^1, x^2, \ldots, x^{n-1}, 0) \in U \}$$

and the parametrization $\tilde{\varphi} : \widetilde{U} :\to \partial M$ given by

$$\tilde{\varphi}(x^1, \ldots, x^{n-1}) = \varphi\left((-1)^n x^1, x^2, \ldots, x^{n-1}, 0 \right).$$

Recall that the orientation on ∂M obtained from φ by just dropping the last coordinate is $(-1)^n$ times the induced orientation on ∂M (cf. Section 9 in Chapter 1). Therefore $\tilde{\varphi}$ gives the correct orientation. The local expression of $i : \partial M \to M$ on these coordinates ($\hat{i} : \widetilde{U} \to U$ such that $\hat{i} = \varphi^{-1} \circ i \circ \tilde{\varphi}$) is given by

$$\hat{i}(x^1, \ldots, x^{n-1}) = \left((-1)^n x^1, x^2, \ldots, x^{n-1}, 0 \right).$$

Hence,

$$\int_{\partial M} i^*\omega = \int_{\widetilde{U}} \tilde{\varphi}^* i^*\omega = \int_{\widetilde{U}} (i\circ\tilde{\varphi})^*\omega = \int_{\widetilde{U}} (\varphi\circ\hat{\imath})^*\omega = \int_{\widetilde{U}} \hat{\imath}^*\varphi^*\omega.$$

Moreover,

$$\begin{aligned}
\hat{\imath}^*\varphi^*\omega &= \hat{\imath}^*\sum_{j=1}^n a_j\, dx^1\wedge\cdots\wedge dx^{j-1}\wedge dx^{j+1}\wedge\cdots\wedge dx^n \\
&= \sum_{j=1}^n (a_j\circ\hat{\imath})\, d\hat{\imath}^1\wedge\cdots\wedge d\hat{\imath}^{j-1}\wedge d\hat{\imath}^{j+1}\wedge\cdots\wedge d\hat{\imath}^n \\
&= (-1)^n(a_n\circ\hat{\imath})\, dx^1\wedge\cdots\wedge dx^{n-1},
\end{aligned}$$

since $d\hat{\imath}^1 = (-1)^n dx^1$, $d\hat{\imath}^n = 0$ and $d\hat{\imath}^j = dx^j$, for $j\neq 1$ and $j\neq n$. Consequently,

$$\begin{aligned}
\int_{\partial M} i^*\omega &= (-1)^n\int_{\widetilde{U}} (a_n\circ\hat{\imath})\, dx^1\cdots dx^{n-1} \\
&= (-1)^n\int_{\widetilde{U}} a_n\left((-1)^n x^1, x^2,\ldots,x^{n-1},0\right)\, dx^1\cdots dx^{n-1} \\
&= (-1)^n\int_{\mathbb{R}^{n-1}} a_n(x^1,x^2,\ldots,x^{n-1},0)\, dx^1\cdots dx^{n-1} = \int_M d\omega
\end{aligned}$$

(where we have used the Change of Variables Theorem). $\square$

EXERCISES 5.2.

(1) Consider the manifolds

$$S^3 = \left\{(x,y,z,w)\in\mathbb{R}^4 : x^2+y^2+z^2+w^2 = 2\right\};$$
$$T^2 = \left\{(x,y,z,w)\in\mathbb{R}^4 : x^2+y^2 = z^2+w^2 = 1\right\}.$$

The submanifold $T^2\subset S^3$ splits S^3 into two connected components. Let M be one of these components and let ω be the 3-form

$$\omega = z dx\wedge dy\wedge dw - x dy\wedge dz\wedge dw.$$

Compute the two possible values of $\int_M\omega$.

(2) (*Homotopy invariance of the integral*) Recall that two maps $f_0, f_1 : M\to N$ are said to be smoothly homotopic if there exists a differentiable map $H:\mathbb{R}\times M\to N$ such that $H(0,p) = f_0(p)$ and $H(1,p) = f_1(p)$ (cf. Exercise 3.8.6). If M is a compact oriented manifold of dimension n and ω is a closed n-form on N, show that

$$\int_M f_0^*\omega = \int_M f_1^*\omega.$$

(3) (a) Let $X\in\mathfrak{X}(S^n)$ be a vector field with no zeros. Show that

$$H(t,p) = \cos(\pi t)p + \sin(\pi t)\frac{X_p}{\|X_p\|}$$

is a smooth homotopy between the identity map and the antipodal map, where we make use of the identification

$$X_p \in T_pS^n \subset T_p\mathbb{R}^{n+1} \cong \mathbb{R}^{n+1}.$$

(b) Using the Stokes Theorem, show that

$$\int_{S^n} \omega > 0,$$

where

$$\omega = \sum_{i=1}^{n+1} (-1)^{i+1} x^i dx^1 \wedge \cdots \wedge dx^{i-1} \wedge dx^{i+1} \wedge \cdots \wedge dx^{n+1}$$

and $S^n = \partial\{x \in \mathbb{R}^{n+1} : \|x\| \leq 1\}$ has the orientation induced by the standard orientation of $\mathbb{R}^{n+1}$.

(c) Show that if n is even then X cannot exist. What happens when n is odd?

6. Orientation and Volume Forms

In this section we will study the relation between orientation and differential forms.

Definition 6.1. *A* **volume form** *(or* **volume element***) on a manifold M of dimension n is an n-form ω such that $\omega_p \neq 0$ for all $p \in M$.*

The existence of a volume form determines an orientation on M:

Proposition 6.2. *A manifold M of dimension n is orientable if and only if there exists a volume form on M.*

Proof. Let ω be a volume form on M, and consider an atlas $\{(U_\alpha, \varphi_\alpha)\}$. We can assume without loss of generality that the open sets U_α are connected. We will construct a new atlas from this one whose overlap maps have derivatives with positive determinant. Indeed, considering the representation of ω on one of these open sets $U_\alpha \subset \mathbb{R}^n$, we have

$$\varphi_\alpha^* \omega = a_\alpha dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n,$$

where the function a_α cannot vanish, and hence must have a fixed sign. If a_α is positive, we keep the corresponding parametrization. If not, we construct a new parametrization by composing φ_α with $(x^1, \ldots, x^n) \mapsto (-x^1, x^2, \ldots, x^n)$. Clearly, in these new coordinates, the new function a_α is positive. Repeating this for all coordinate neighborhoods we obtain a new atlas for which all the functions a_α are positive, which we will also denote by $\{(U_\alpha, \varphi_\alpha)\}$. Moreover, whenever $W := \varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta) \neq \varnothing$, we have

$$(\varphi_\alpha^{-1})^* \omega_\alpha = (\varphi_\beta^{-1})^* \omega_\beta$$

and so $\omega_\alpha = (\varphi_\beta^{-1} \circ \varphi_\alpha)^* \omega_\beta$. Hence,

$$\begin{aligned} a_\alpha dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n &= (\varphi_\beta^{-1} \circ \varphi_\alpha)^* a_\beta \, dx_\beta^1 \wedge \cdots dx_\beta^n \\ &= (a_\beta \circ \varphi_\beta^{-1} \circ \varphi_\alpha) \det\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right) dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n \end{aligned}$$

and so $\det\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right) > 0$. We conclude that M is orientable.

Conversely, if M is orientable, we consider an atlas $\{(U_\alpha, \varphi_\alpha)\}$ for which the overlap maps $\varphi_\beta^{-1} \circ \varphi_\alpha$ are such that $\det\left(d(\varphi_\beta^{-1} \circ \varphi_\alpha)\right) > 0$. Taking a partition of unity $\{\rho_i\}_{i \in I}$ subordinate to the cover of M by the corresponding coordinate neighborhoods, we may define the forms

$$\omega_i := \rho_i dx_i^1 \wedge \cdots \wedge dx_i^n$$

with $\operatorname{supp} \omega_i = \operatorname{supp} \rho_i \subset \varphi_{\alpha_i}(U_{\alpha_i})$. Extending these forms to M by making them zero outside $\operatorname{supp} \rho_i$, we may define the form $\omega := \sum_{i \in I} \omega_i$. Clearly ω is a well defined n-form on M so we just need to show that $\omega_p \neq 0$ for all $p \in M$. Let p be a point in M. Hence there is an $i \in I$ such that $\rho_i(p) > 0$ and $\operatorname{supp} \rho_i \subset \varphi_{\alpha_i}(U_{\alpha_i})$. Then, there are linearly independent vectors $v_1, \ldots, v_n \in T_pM$ such that $(\omega_i)_p(v_1, \ldots, v_n) > 0$. Moreover, for all other $j \in I \backslash \{i\}$, $(\omega_j)_p(v_1, \ldots, v_n) \geq 0$. Indeed, if $p \notin \varphi_{\alpha_j}(U_{\alpha_j})$, then $(\omega_j)_p(v_1, \ldots, v_n) = 0$. On the other hand, if $p \in \varphi_{\alpha_j}(U_{\alpha_j})$, then equation (4) yields

$$dx_j^1 \wedge \cdots \wedge dx_j^n = \det\left(d(\varphi_{\alpha_j}^{-1} \circ \varphi_{\alpha_i})\right) dx_i^1 \wedge \cdots \wedge dx_i^n$$

and hence

$$(\omega_j)_p(v_1, \ldots, v_n) = \frac{\rho_j(p)}{\rho_i(p)} \det\left(d(\varphi_{\alpha_j}^{-1} \circ \varphi_{\alpha_i})\right)(\omega_i)_p(v_1, \ldots, v_n) \geq 0.$$

Consequently, $\omega_p(v_1, \ldots, v_n) > 0$, and so ω is a volume form. $\square$

Remark 6.3. Sometimes we call a volume form an orientation. In this case the orientation on M is the one for which a basis $\{v_1, \ldots, v_n\}$ of T_pM is positive if and only if $\omega_p(v_1, \ldots, v_n) > 0$.

If we fix a volume form $\omega \in \Omega^n(M)$ on the orientable manifold M, we can define the **integral** of any compactly supported function $f \in C^\infty(M, \mathbb{R})$ as

$$\int_M f = \int_M f\omega$$

(where the orientation of M is determined by ω). If M is compact, we define its **volume** to be

$$\operatorname{vol}(M) = \int_M 1 = \int_M \omega.$$

Exercises 6.4.

(1) Show that $M \times N$ is orientable if and only if both M and N are orientable.
(2) Let M be an oriented manifold with volume element $\omega \in \Omega^n(M)$. Prove that if $f > 0$ then $\int_M f\omega > 0$. (**Remark:** In particular, the volume of a compact manifold is always positive).

(3) Let M^n be a compact orientable manifold, and let ω be an $(n-1)$-form in M.
   (a) Show that there exists a point $p \in M$ for which $(d\omega)_p = 0$.
   (b) Prove that there exists no immersion $f : S^1 \to \mathbb{R}$, of the unit circle into $\mathbb{R}$.

(4) Let $f : S^n \to S^n$ be the antipodal map. Recall that the n-dimensional projective space is the differential manifold $\mathbb{R}P^n = S^n/\mathbb{Z}_2$, where the group $\mathbb{Z}_2 = \{1, -1\}$ acts on S^n through $1 \cdot x = x$ and $(-1) \cdot x = f(x)$. Let $\pi : S^n \to \mathbb{R}P^n$ be the natural projection.
   (a) Prove that $\omega \in \Omega^k(S^n)$ is of the form $\omega = \pi^*\theta$ for some $\theta \in \Omega^k(\mathbb{R}P^n)$ *iff* $f^*\omega = \omega$.
   (b) Show that $\mathbb{R}P^n$ is orientable *iff* n is odd, and that, in this case,
$$\int_{S^n} \pi^*\theta = 2 \int_{\mathbb{R}P^n} \theta.$$
   (c) Show that for n even the sphere S^n is the orientable double covering of $\mathbb{R}P^n$ (cf. Exercise 8.6.9 in Chapter 1).

(5) Let M be a compact oriented manifold with boundary and $\omega \in \Omega^n(M)$ a volume element. The **divergence** of a vector field $X \in \mathfrak{X}(M)$ is the function $\operatorname{div}(X)$ such that
$$L_X\omega = (\operatorname{div}(X))\omega$$
(cf. Exercise 3.8.8). Show that
$$\int_M \operatorname{div}(X) = \int_{\partial M} \iota(X)\omega.$$

(6) (*Brouwer Fixed Point Theorem*)
   (a) Let M^n be a compact orientable manifold with boundary $\partial M \neq \varnothing$. Show that there exists no smooth map $f : M \to \partial M$ satisfying $f|_{\partial M} = \mathrm{id}$.
   (b) Prove the **Brouwer Fixed Point Theorem**: Let $B = \{x \in \mathbb{R}^n : |x| \leq 1\}$. Any smooth map $g : B \to B$ has a fixed point, that is, there exists a point $p \in B$ such that $g(p) = p$. (**Hint:** For each point $x \in B$, consider the ray r_x starting at $g(x)$ and passing through x. There is only one point $y(x)$ different from $g(x)$ on $r_x \cap \partial B$. Consider the map $f : B \to \partial B$, that maps $x \in B$ to $y(x)$).

7. Notes on Chapter 2

7.1. Section 1.

(1) Given a finite dimensional vector space V we define its **dual space** as the space of linear functionals on V.

Proposition 7.1. *If $\{v_1, \ldots, v_n\}$ is a basis for V then there is a unique basis $\{T_1, \ldots, T_n\}$ of V^* dual to $\{v_1, \ldots, v_n\}$, that is, such that $T_i(v_j) = \delta_{ij}$.*

PROOF. By linearity, the equations $T_i(v_j) = \delta_{ij}$ define a unique set of functionals $T_i \in V^*$. Indeed, for any $v \in V$, we have $v = \sum_{j=1}^n a_j v_j$ and so

$$T_i(v) = \sum_{j=1}^n a_j T_i(v_j) = \sum_{j=1}^n a_j \delta_{ij} = a_i.$$

Moreover, these uniquely defined functionals are linearly independent. In fact, if

$$T := \sum_{i=1}^n b_i T_i = 0,$$

then, for each $j = 1, \dots, n$, we have $0 = T(v_j) = \sum_{i=1}^n b_i T_i(v_j) = b_j$. To show that $\{T_1, \dots, T_n\}$ generates V^*, we take any $S \in V^*$ and set $b_i := S(v_i)$. Then, defining $T := \sum_{i=1}^n b_i T_i$, we see that $S(v_j) = T(v_j)$ for all $j = 1, \dots, n$. Since $\{v_1, \dots, v_n\}$ is a basis for V, we have $S = T$. $\square$

Moreover, if $\{v_1, \dots, v_n\}$ is a basis for V and $\{T_1, \dots, T_n\}$ is its dual basis, then, for any $v = \sum a_j v_j \in V$ and $T = \sum b_i T_i \in V^*$, we have

$$T(v) = \sum_{j=i}^n b_i T_i(v) = \sum_{i,j=1}^n a_j b_i T_i(v_j) = \sum_{i,j=1}^n a_j b_i \delta_{ij} = \sum_{i=1}^n a_i b_i.$$

If we now consider a linear functional F on V^*, that is, an element of $(V^*)^*$, we have $F(T) = T(v_0)$ for some fixed vector $v_0 \in V$. Indeed, let $\{v_1, \dots, v_n\}$ be a basis for V and let $\{T_1, \dots, T_n\}$ be its dual basis. Then if $T = \sum_{i=1}^n b_i T_i$, we have $F(T) = \sum_{i=1}^n b_i F(T_i)$. Denoting the values $F(T_i)$ by a_i, we get $F(T) = \sum_{i=1}^n a_i b_i = T(v_0)$ for $v_0 = \sum_{i=1}^n a_i v_i$. This establishes a one-to-one correspondence between $(V^*)^*$ and V, and allows us to view V as the space of linear functionals on V^*. For $v \in V$ and $T \in V^*$, we write $v(T) = T(v)$.

(2) Changing from a basis $\{v_1, \dots, v_n\}$ to a new basis $\{v'_1, \dots, v'_n\}$ in V, we obtain a **change of basis matrix** S, whose jth column is the vector of coordinates of the new basis vector v'_j in the old basis. We can then write the symbolic matrix equation

$$(v'_1, \dots, v'_n) = (v_1, \dots, v_n)S.$$

The coordinate (column) vectors a and b of a vector $v \in V$ (a contravariant 1-tensor on V) with respect to the old basis and to the new basis are related by

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = S^{-1} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = S^{-1} a,$$

since we must have $(v_1',\dots,v_n')b=(v_1,\dots,v_n)a=(v_1',\dots,v_n')S^{-1}a$. On the other hand, if $\{T_1,\dots,T_n\}$ and $\{T_1',\dots,T_n'\}$ are the dual bases of $\{v_1,\dots,v_n\}$ and $\{v_1',\dots,v_n'\}$, we have

$$\begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}(v_1,\dots,v_n)=\begin{pmatrix} T_1' \\ \vdots \\ T_n' \end{pmatrix}(v_1',\dots,v_n')=I$$

(where, in the symbolic matrix multiplication above, each coordinate is obtained by applying the covectors to the vectors). Hence,

$$\begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}(v_1',\dots,v_n')\,S^{-1}=I \quad\Leftrightarrow\quad S^{-1}\begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}(v_1',\dots,v_n')=I,$$

implying that

$$\begin{pmatrix} T_1' \\ \vdots \\ T_n' \end{pmatrix}=S^{-1}\begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}.$$

The coordinate (row) vectors $a=(a_1,\dots,a_n)$ and $b=(b_1,\dots,b_n)$ of a 1-tensor $T\in V^*$ (a covariant 1-tensor on V) with respect to the old basis $\{T_1,\dots,T_n\}$ and to the new basis $\{T_1',\dots,T_n'\}$ are related by

$$a\begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}=b\begin{pmatrix} T_1' \\ \vdots \\ T_n' \end{pmatrix} \quad\Leftrightarrow\quad aS\begin{pmatrix} T_1' \\ \vdots \\ T_n' \end{pmatrix}=b\begin{pmatrix} T_1' \\ \vdots \\ T_n' \end{pmatrix}$$

and so $b=aS$. Note that the coordinate vectors of the covariant 1-tensors on V transform like the basis vectors of V (that is, by means of the matrix S) whereas the coordinate vectors of the contravariant 1-tensors on V transform by means of the inverse of this matrix.

7.2. Section 4.

(1) *(Change of Variables Theorem)* Let $U,V\subset\mathbb{R}^n$ be open sets, $g:U\to V$ a diffeomorphism and $f:V\to\mathbb{R}$ an integrable function. Then

$$\int_V f=\int_U (f\circ g)|\det dg|.$$

(2) To define smooth objects on manifolds it is often useful to define them first on coordinate neighborhoods and then glue the pieces together by means of a **partition of unity**.

THEOREM 7.2. *Let M be a smooth manifold and $\mathcal{V}$ an open cover of M. Then there is a family of differentiable functions on M, $\{\rho_i\}_{i\in I}$, such that:*

(i) for every point $p \in M$, there exists a neighborhood U of p such that $U \cap \operatorname{supp} \rho_i = \varnothing$ except for a finite number of ρ_i's;
(ii) for every point $p \in M$, $\sum_{i \in I} \rho_i(p) = 1$;
(iii) $0 \leq \rho_i \leq 1$ and $\operatorname{supp} \rho_i \subset V$ for some element $V \in \mathcal{V}$.

Remark 7.3. This collection ρ_i of smooth functions is called partition of unity subordinate to the cover $\mathcal{V}$.

Proof. Let us first assume that M is compact. For every point $p \in M$ we consider a coordinate neighborhood $W_p = \varphi_p(U_p)$ around p contained in an element V_p of $\mathcal{V}$, such that $\varphi_p(0) = p$ and $B_3(0) \subset U_p$ (where $B_3(0)$ denotes the ball of radius 3 around 0). Then we consider the C^∞-functions (cf. Figure 1)

$$\begin{aligned} \lambda : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} e^{\frac{1}{(x-1)(x-2)}} & \text{if } 1 < x < 2 \\ 0 & \text{otherwise} \end{cases} , \end{aligned}$$

$$\begin{aligned} h : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \frac{\int_x^2 \lambda(t)\ dt}{\int_1^2 \lambda(t)\ dt} , \end{aligned}$$

$$\begin{aligned} \beta : \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto h(|x|) . \end{aligned}$$

Notice that h is a decreasing function with values $0 \leq h(x) \leq 1$,

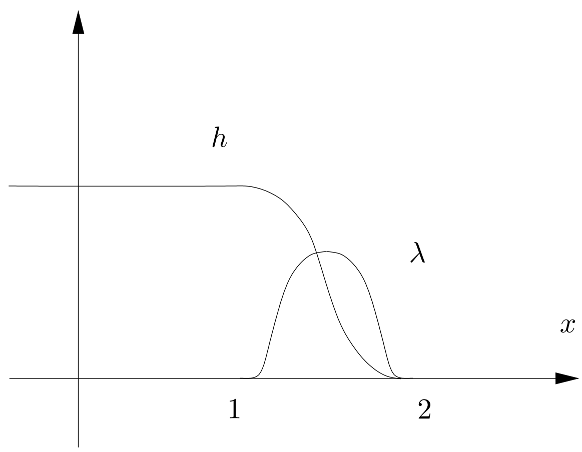

Figure 1

equal to zero for $x \geq 2$ and equal to 1 for $x \leq 1$. Hence, we can consider **bump functions** $\gamma_p : M \to [0,1]$ defined by

$$\gamma_p(q) = \begin{cases} \beta(\varphi_p^{-1}(q)) & \text{if } q \in \varphi_p(U_p) \\ \\ 0 & \text{otherwise.} \end{cases}$$

Then $\operatorname{supp} \gamma_p = \overline{\varphi_p(B_2(0))} \subset \varphi_p(B_3(0)) \subset W_p$ is contained inside an element V_p of the cover. Moreover, $\{\varphi_p(B_1(0))\}_{p \in M}$ is an open cover of M and so we can consider a finite subcover $\{\varphi_{p_i}(B_1(0))\}_{i=1}^k$ such that $M = \cup_{i=1}^k \varphi_{p_i}(B_1(0))$. Finally we take the functions

$$\rho_i = \frac{\gamma_{p_i}}{\sum_{j=1}^k \gamma_{p_j}}.$$

Note that $\sum_{j=1}^k \gamma_{p_j}(q) \neq 0$ since q is necessarily contained inside some $\varphi_{p_i}(B_1(0))$ and so $\gamma_i(q) \neq 0$. Moreover, $0 \leq \rho_i \leq 1$, $\sum \rho_i = 1$ and $\operatorname{supp} \rho_i = \operatorname{supp} \gamma_{p_i} \subset V_{p_i}$.

If M is not compact we can use a **compact exhaustion**, that is, a sequence $\{K_i\}_{i \in \mathbb{N}}$ of compact subsets of M such that $K_i \subset \operatorname{int} K_{i+1}$ and $M = \cup_{i=1}^\infty K_i$. The partition of unity is then obtained as follows. The family $\{\varphi_p(B_1(0))\}_{p \in M}$ is a cover of K_1, so we can consider a finite subcover of K_1,

$$\left\{\varphi_{p_1}(B_1(0)), \ldots, \varphi_{p_{k_1}}(B_1(0))\right\}.$$

By induction, we obtain a finite number of points such that

$$\left\{\varphi_{p_1^i}(B_1(0)), \ldots, \varphi_{p_{k_i}^i}(B_1(0))\right\}$$

covers $K_i \backslash \operatorname{int} K_{i-1}$ (a compact set). Then, for each i, we consider the corresponding bump functions

$$\gamma_{p_1^i}, \ldots, \gamma_{p_{k_i}^i} : M \to [0,1].$$

Note that $\gamma_{p_1{}^i} + \cdots + \gamma_{p_{k_i}^i} > 0$ for every $q \in K_i \backslash \operatorname{int} K_{i-1}$ (as there is always one of these functions which is different from zero). As in the compact case, we can choose these bump functions so that $\operatorname{supp} \gamma_{p_j^i}$ is contained in some element of $\mathcal{V}$. We will also choose them so that $\operatorname{supp} \gamma_{p_j^i} \subset \operatorname{int} K_{i+1} \backslash K_{i-2}$ (an open set). Hence, $\{\gamma_{p_j^i}\}_{i \in \mathbb{N}, 1 \leq j \leq k_i}$ is locally finite, meaning that, given a point $p \in M$, there exists an open neighborhood V of p such that only a finite number of these functions is different from zero in V. Consequently, the sum $\sum_{i=1}^\infty \sum_{j=1}^{k_i} \gamma_{p_j^i}$ is a positive, differentiable function on M. Finally, making

$$\rho_j^i = \frac{\gamma_{p_j^i}}{\sum_{i=1}^\infty \sum_{j=1}^{k_i} \gamma_{p_j^i}},$$

we obtain the desired partition of unity (subordinate to $\mathcal{V}$). $\square$

REMARK 7.4. Compact exhaustions always exist on manifolds. In fact, if U is a bounded open set of $\mathbb{R}^n$, one can easily construct a compact exhaustion $\{K_i\}_{i\in\mathbb{N}}$ for U by setting

$$K_i = \left\{ x \in U : \operatorname{dist}(x, \partial U) \geq \frac{1}{n} \right\}.$$

If M is a differentiable manifold, one can always take a countable atlas $\mathcal{A} = \{(U_j, \varphi_j)\}_{j\in\mathbb{N}}$ such that each U_j is a bounded open set, thus admitting a compact exhaustion $\{K_i^j\}_{i\in\mathbb{N}}$. Therefore

$$\left\{ \bigcup_{i+j=l} \varphi_j \left(K_i^j \right) \right\}_{l\in\mathbb{N}}$$

is a compact exhaustion of M.

7.3. Section 5. *(Fubini Theorem)* Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ be compact intervals and let $f : A \times B \to \mathbb{R}$ be a continuous function. Then

$$\begin{aligned} \int_{A\times B} f &= \int_A \left(\int_B f(x,y) dy^1 \cdots dy^m \right) dx^1 \cdots dx^n \\ &= \int_B \left(\int_A f(x,y) dx^1 \cdots dx^n \right) dy^1 \cdots dy^m. \end{aligned}$$

7.4. Bibliographical notes. The material in this chapter can be found in most books on differential geometry (e.g. [**Boo03, GHL04**]). A text entirely dedicated to differential forms and their applications is [**dC94**]. The study of de Rham cohomology leads to a beautiful and powerful theory, whose details can be found for instance in [**BT82**].

CHAPTER 3

Riemannian Manifolds

In this chapter we begin our study of Riemannian geometry.

Section 1 introduces the concept of a **Riemannian metric** on a smooth manifold, which is simply a tensor field determining an inner product at each tangent space.

In Section 2 we define **affine connections**, which provide a notion of **parallelism** of vectors along curves, and consequently of **geodesics** (curves whose tangent vector is parallel). Riemannian manifolds carry a special connection, called the **Levi-Civita connection** (Section 3), whose geodesics have special distance-minimizing properties (Section 4).

In Section 5 we prove the **Hopf-Rinow Theorem**, relating the properties of a Riemannian manifold as a metric space with the properties of its geodesics.

1. Riemannian Manifolds

The metric properties of $\mathbb{R}^n$ (distances, angles, volumes) are determined by the canonical Cartesian coordinates. In a general differentiable manifold, however, there are no such preferred coordinates; to define distances, angles and volumes we must add more structure, namely a special tensor field called a **Riemannian metric**.

DEFINITION 1.1. *A tensor $g \in \mathcal{T}^2(T_p^*M)$ is said to be*

(i) **symmetric** *if $g(v,w) = g(w,v)$ for all $v, w \in T_pM$;*
(ii) **nondegenerate** *if $g(v,w) = 0$ for all $w \in T_pM$ implies $v = 0$;*
(iii) **positive definite** *if $g(v,v) > 0$ for all $v \in T_pM \setminus \{0\}$.*

A 2-covariant tensor field g is said to be symmetric, nondegenerate or positive definite if g_p is symmetric, nondegenerate or positive definite for all $p \in M$. If $x : V \to \mathbb{R}^n$ is a local chart, we have

$$g = \sum_{i,j=1}^{n} g_{ij} dx^i \otimes dx^j$$

in V, where

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right).$$

It is easy to see that g is symmetric, nondegenerate or positive definite if and only if the matrix (g_{ij}) has these properties (see Exercise 1.11.1).

DEFINITION 1.2. *A* **Riemannian metric** *on a smooth manifold M is a symmetric positive definite smooth 2-covariant tensor field g. A smooth manifold M equipped with a Riemannian metric g is called a* **Riemannian manifold***, and denoted by (M, g).*

A Riemannian metric is therefore a smooth assignment of an inner product to each tangent space. It is usual to write

$$g_p(v, w) = \langle v, w \rangle_p.$$

EXAMPLE 1.3. (**Euclidean n-space**) It should be clear that $M = \mathbb{R}^n$ and

$$g = \sum_{i=1}^{n} dx^i \otimes dx^i$$

define a Riemannian manifold.

PROPOSITION 1.4. *Let (N, g) be a Riemannian manifold and $f : M \to N$ an immersion. Then f^*g is a Riemannian metric in M (called the* **induced metric***).*

PROOF. We just have to prove that f^*g is symmetric and positive definite. Let $p \in M$ and $v, w \in T_pM$. Since g is symmetric,

$$(f^*g)_p(v, w) = g_{f(p)}((df)_p v, (df)_p w) = g_{f(p)}((df)_p w, (df)_p v) = (f^*g)_p(w, v).$$

On the other hand, it is clear that $(f^*g)_p(v, v) \geq 0$, and

$$(f^*g)_p(v, v) = 0 \Rightarrow g_{f(p)}((df)_p v, (df)_p v) = 0 \Rightarrow (df)_p v = 0 \Rightarrow v = 0$$

(as $(df)_p$ is injective). □

In particular, any submanifold M of a Riemannian manifold (N, g) is itself a Riemannian manifold. Notice that, in this case, the induced metric at each point $p \in M$ is just the restriction of g_p to $T_pM \subset T_pN$. Since $\mathbb{R}^n$ is a Riemannian manifold (cf. Example 1.3), we see that any submanifold of $\mathbb{R}^n$ is a Riemannian manifold. Whitney's Theorem then implies that any manifold admits a Riemannian metric.

It was proved in 1954 by John Nash [**Nas56**] that any compact n-dimensional Riemannian manifold can be isometrically embedded in $\mathbb{R}^N$ for $N = \frac{n(3n+11)}{2}$ (that is, embedded in such a way that its metric is induced by the Euclidean metric of $\mathbb{R}^N$). Gromov [**Gro70**] later proved that one can take $N = \frac{(n+2)(n+3)}{2}$. Notice that for $n = 2$ Nash's result gives an isometric embedding of any compact surface into $\mathbb{R}^{17}$, and Gromov's into $\mathbb{R}^{10}$. In fact, Gromov has further showed that any surface isometrically embeds in $\mathbb{R}^5$. This result cannot be improved, as the real projective plane with the standard metric (see Exercise 1.11.3) cannot be isometrically embedded into $\mathbb{R}^4$.

EXAMPLE 1.5. The **standard metric** on

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

is the metric induced on S^n by the Euclidean metric on $\mathbb{R}^{n+1}$. A parametrization of the open set

$$U = \{x \in S^n : x^{n+1} > 0\}$$

is for instance

$$\varphi(x^1, \ldots, x^n) = \left(x^1, \ldots, x^n, \sqrt{1 - (x^1)^2 - \ldots - (x^n)^2}\right),$$

and hence the coefficients of the metric tensor are

$$g_{ij} = \left\langle \frac{\partial \varphi}{\partial x^i}, \frac{\partial \varphi}{\partial x^j} \right\rangle = \delta_{ij} + \frac{x^i x^j}{1 - (x^1)^2 - \ldots - (x^n)^2}.$$

Two Riemannian manifolds will be regarded to be the same if they are **isometric**:

Definition 1.6. *Let (M, g) and (N, h) be Riemannian manifolds. A diffeomorphism $f : M \to N$ is said to be an* **isometry** *if $f^*h = g$. Similarly, a local diffeomorphism $f : M \to N$ is said to be a* **local isometry** *if $f^*h = g$.*

Particularly simple examples of Riemannian manifolds can be constructed from Lie groups. Recall that given a Lie group G and $x \in G$, the **left translation** by x is the diffeomorphism $L_x : G \to G$ given by $L_x(y) = xy$ for all $y \in G$. A Riemannian metric on G is said to be **left-invariant** if L_x is an isometry for all $x \in G$.

Proposition 1.7. *Let G be a Lie group. Then $g(\cdot, \cdot) \equiv \langle \cdot, \cdot \rangle$ is a left-invariant metric if and only if*

$$\langle v, w \rangle_x = \langle (dL_{x^{-1}})_x \, v, (dL_{x^{-1}})_x \, w \rangle_e \tag{5}$$

for all $x \in G$ and $v, w \in T_xG$, where $e \in G$ is the identity element and $\langle \cdot, \cdot \rangle_e$ is an inner product on the Lie algebra $\mathfrak{g} = T_eG$.

Proof. If g is left-invariant, then (5) must obviously hold. Thus we just have to show that (5) defines indeed a left-invariant metric on G. We leave it as an exercise to show that the smoothness of the map

$$G \times G \ni (x, y) \mapsto x^{-1}y = L_{x^{-1}}y \in G$$

implies the smoothness of the map

$$G \times TG \ni (x, v) \mapsto (dL_{x^{-1}})_x \, v \in TG,$$

and that therefore (5) defines a smooth tensor field g on G. It should also be clear from (5) that g is symmetric and positive definite. All that remains to be proved is that g is left-invariant, that is,

$$\langle (dL_y)_x \, v, (dL_y)_x \, w \rangle_{yx} = \langle v, w \rangle_x$$

for all $v, w \in T_xG$ and all $x, y \in G$. We have

$$\begin{aligned}\left\langle (dL_y)_x v, (dL_y)_x w \right\rangle_{yx} &= \left\langle \left(dL_{(yx)^{-1}}\right)_{yx} (dL_y)_x v, \left(dL_{(yx)^{-1}}\right)_{yx} (dL_y)_x w \right\rangle_e \\ &= \left\langle \left(d\left(L_{x^{-1}y^{-1}} \circ L_y\right)\right)_x v, \left(d\left(L_{x^{-1}y^{-1}} \circ L_y\right)\right)_x w \right\rangle_e \\ &= \left\langle (dL_{x^{-1}})_x v, (dL_{x^{-1}})_x w \right\rangle_e \\ &= \langle v, w \rangle_x .\end{aligned}$$

□

Thus any inner product on the Lie algebra $\mathfrak{g} = T_eG$ determines a left-invariant metric on G.

A Riemannian metric allows us to compute the length of any vector (as well as the angle between two vectors with the same base point). Therefore we can measure the length of curves:

Definition 1.8. *If $(M, \langle \cdot, \cdot \rangle)$ is a Riemannian manifold and $c : I \subset \mathbb{R} \to M$ is a differentiable curve, the* **length** *of its restriction to $[a, b] \subset I$ is*

$$l(c) = \int_a^b \langle \dot{c}(t), \dot{c}(t) \rangle^{\frac{1}{2}} dt.$$

The length of a curve segment does not depend on the parametrization (see Exercise 1.11.5).

If M is an orientable n-dimensional manifold then it possesses volume elements, that is, differential forms $\omega \in \Omega^n(M)$ such that $\omega_p \neq 0$ for all $p \in M$. Clearly, there are as many volume elements as differentiable functions $f \in C^\infty(M)$ without zeros.

Definition 1.9. *If (M, g) is an orientable Riemannian manifold, $\omega \in \Omega^n(M)$ is said to be a* **Riemannian volume element** *if*

$$\omega_p(v_1, \ldots, v_n) = \pm 1$$

for any orthonormal basis $\{v_1, \ldots, v_n\}$ of T_pM and all $p \in M$.

Notice that if M is connected there exist exactly two Riemannian volume elements (one for each choice of orientation). Moreover, if ω is a Riemannian volume element and $x : V \to \mathbb{R}$ is a chart compatible with the orientation induced by ω, one has

$$\omega = f dx^1 \wedge \ldots \wedge dx^n$$

for some positive function

$$f = \omega\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right).$$

If S is the matrix whose columns are the components of $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ on some orthonormal basis with the same orientation, we have

$$f = \det S = \left(\det\left(S^2\right)\right)^{\frac{1}{2}} = \left(\det\left(S^tS\right)\right)^{\frac{1}{2}} = (\det(g_{ij}))^{\frac{1}{2}}$$

since clearly the matrix S^tS is the matrix whose (i,j)-th entry is the inner product $g\left(\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\right)=g_{ij}$.

A Riemannian metric $\langle\cdot,\cdot\rangle$ on M determines a linear isomorphism between T_pM and T_p^*M for all $p\in M$, by mapping any vector $v_p\in T_pM$ to the covector ω_p given by $\omega_p(w_p)=\langle v_p,w_p\rangle$. This extends to an isomorphism between $\mathfrak{X}(M)$ and $\Omega^1(M)$. In particular, we have

Definition 1.10. *Let (M,g) be a Riemannian manifold and $f:M\to\mathbb{R}$ a smooth function. The* **gradient** *of f is the vector field* $\operatorname{grad} f$ *associated to the 1-form df through the isomorphism determined by g.*

Exercises 1.11.

(1) Let $g=\sum_{i,j=1}^n g_{ij}\,dx^i\otimes dx^j\in\mathcal{T}^2(T_p^*M)$. Show that:
   (i) g is symmetric *iff* $g_{ij}=g_{ji}$ $(i,j=1,\dots,n)$;
   (ii) g is nondegenerate *iff* $\det(g_{ij})\neq 0$;
   (iii) g is positive definite *iff* (g_{ij}) is a positive definite matrix;
   (iv) if g is nondegenerate, the map $\Phi_g:T_pM\to T_p^*M$ given by
   $$\Phi_g(v)(w)=g(v,w)$$
   for all $v,w\in T_pM$ is a linear isomorphism;
   (v) if g is positive definite then g is nondegenerate.

(2) Prove that any differentiable manifold admits a Riemannian structure without invoking Whitney's Theorem. (**Hint:** Use partitions of unity).

(3) (a) Let (M,g) be a Riemannian manifold and let G be a Lie group acting freely and properly on M by isometries. Show that M/G has a natural Riemannian structure (called the **quotient** structure).
   (b) How would you define the **standard metric** on the standard n-torus $T^n=\mathbb{R}^n/\mathbb{Z}^n$?
   (c) How would you define the **standard metric** on the real projective n-space $\mathbb{R}P^n=S^n/\mathbb{Z}_2$?

(4) Show that the standard metric on $S^3\cong SU(2)$ is left-invariant.

(5) We say that a differentiable curve $\gamma:J\to M$ is obtained from the curve $c:I\to M$ by **reparametrization** if there exists a smooth bijection $f:I\to J$ (the reparametrization) such that $c=\gamma\circ f$. Show that if γ is obtained from c by reparametrization then the length of the restriction of c to $[a,b]\subset I$ is equal to the length of the restriction of γ to $f([a,b])\subset J$.

(6) Let (M,g) be a Riemannian manifold and $f\in C^\infty(M)$. Show that if $a\in\mathbb{R}$ is a regular value of f then $\operatorname{grad}(f)$ is orthogonal to the submanifold $f^{-1}(a)$.

(7) Let (M,g) be an oriented Riemannian manifold with boundary. For each point $p\in M$ we define the linear isomorphism $\tilde g_p:T_pM\to T_p^*M$ given by $(\tilde g_p(v))(w)=g_p(v,w)$ for all $v,w\in T_pM$. Therefore,

we can identify T_pM and T_p^*M, and extend this identification to the spaces $\mathfrak{X}(M)$ and $\Omega^1(M)$ of vector fields and 1-forms on M.

(a) Given two 1-forms $\omega, \eta \in \Omega^1(M)$, we can define their inner product $\langle \omega, \eta \rangle$ as the inner product of the associated vector fields. If $k \geq 1$, we define the inner product of $\alpha := \alpha_1 \wedge \cdots \wedge \alpha_k$ and $\beta := \beta_1 \wedge \cdots \wedge \beta_k$ (with $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_k \in \Omega^1(M)$) to be $\langle \alpha, \beta \rangle = \det\left(\langle \alpha_i, \beta_j \rangle\right)$. By linearity, we can define the inner product of any two k-forms $\alpha, \beta \in \Omega^k(M)$. Show that this inner product is well defined, i.e., does not depend on the representations of α, β. Compute $\langle \omega, \eta \rangle$ for the following 2-forms in $\mathbb{R}^3$:

$$\begin{aligned} \omega &:= a\, dx \wedge dy + b\, dy \wedge dz + c\, dz \wedge dx \\ \eta &:= e\, dx \wedge dy + f\, dy \wedge dz + g\, dz \wedge dx \end{aligned}$$

(**Remark:** For $k = 0$ we define the inner product of functions f, g to be the usual product fg).

(b) (*Hodge $*$-operator*) Consider the linear isomorphism $* : \Lambda^k T_p^*M \to \Lambda^{n-k} T_p^*M$ defined as follows: if $\{\theta_1, \ldots, \theta_k, \theta_{k+1}, \ldots, \theta_n\}$ is any positively oriented orthonormal basis of T_p^*M then $*(\theta_1 \wedge \cdots \wedge \theta_k) = \theta_{k+1} \wedge \cdots \wedge \theta_n$. Show that $*$ is well defined.

(c) We can define $* : \Omega^k(M) \to \Omega^{n-k}(M)$ by setting $(*\omega)_p = *(\omega_p)$ for all $p \in M$ and $\omega \in \Omega^k(M)$. Write out an expression for $*\omega$ in local coordinates, and show that it is a differential form.

(d) Prove that for all $f, g \in C^\infty(M, \mathbb{R})$ and $\omega, \eta \in \Omega^k(M)$
   (i) $*(f\,\omega + g\,\eta) = f * \omega + g * \eta$;
   (ii) $* * \omega = (-1)^{k(n-k)}\,\omega$;
   (iii) $\omega \wedge *\eta = \eta \wedge *\omega = \langle \omega, \eta \rangle \vartheta_M$;
   (iv) $*(\omega \wedge *\eta) = *(\eta \wedge *\omega) = \langle \omega, \eta \rangle$;
   (v) $\langle *\omega, *\eta \rangle = \langle \omega, \eta \rangle$,

   where $\vartheta_M = *1$ is the Riemannian volume element determined by the metric g and the orientation of M.

(e) (*Divergence Theorem*) Let $X \in \mathfrak{X}(M)$ be a vector field on M and $\omega_X \in \Omega^1(M)$ be the 1-form determined by X. Defining the **divergence** of X to be $\operatorname{div} X := *d * \omega_X$, show that if M is compact

$$\int_M \operatorname{div} X\, \vartheta_M = \int_{\partial M} \langle X, n \rangle\, \vartheta_{\partial M}$$

where n is the outward-pointing unit vector field on ∂M.

(f) Assume that $\partial M = \varnothing$. Show that

$$(\omega, \eta) := \int_M \langle \omega, \eta \rangle\, \vartheta_M$$

is an inner product on $\Omega^k(M)$. Moreover, show that $(\omega, \eta) = \int_M \omega \wedge *\,\eta = \int_M \eta \wedge *\,\omega$ and $(*\omega, *\eta) = (\omega, \eta)$.

(g) Define the linear operator $\delta : \Omega^k(M) \to \Omega^{k-1}(M)$ as $\delta := (-1)^k(*^{-1})d*$. Show that:
  (i) $\delta = (-1)^{n(k+1)+1} * d*$;
  (ii) $*\delta = (-1)^k d*$;
  (iii) $\delta* = (-1)^{k+1} * d$;
  (iv) $\delta \circ \delta = 0$;
  (v) $(d\omega, \eta) = (\omega, \delta\eta)$.

(h) (*Laplacian*) Consider the **Laplacian operator** $\Delta := d\delta + \delta d : \Omega^k(M) \to \Omega^k(M)$. Show that if ω, η are differential forms and f is a differentiable function,
  (i) $*\Delta = \Delta*$;
  (ii) $(\Delta\omega, \eta) = (\omega, \Delta\eta)$;
  (iii) $\Delta\omega = 0 \Leftrightarrow (d\omega = 0 \text{ and } \delta\omega = 0)$;
  (iv) $\Delta f = -\operatorname{div}(\operatorname{grad}(f))$;
  (v) $\Delta(fg) = f\Delta g + g\Delta f - 2\langle \operatorname{grad}(f), \operatorname{grad}(g)\rangle$.

(i) A **harmonic form** is a differential form ω such that $\Delta\omega = 0$. Show that if M is connected then any harmonic function on M must be constant, and any harmonic n-form ($n = \dim M$) must be a constant multiple of the volume element ϑ_M.

(j) Assume the following result (**Hodge decomposition**): Any k-form ω on a compact oriented Riemannian manifold M can be uniquely decomposed in a sum $\omega = \omega_H + d\alpha + \delta\beta$, where ω_H is a harmonic form, $\alpha \in \Omega^{k-1}(M)$ and $\beta \in \Omega^{k+1}(M)$. Show that any cohomology class on M (cf. Exercise 3.8.5 in Chapter 2) can be uniquely represented by a harmonic form.

(k) (*Green's formula*) Let M be a compact Riemannian manifold with boundary. The **normal derivative** of a smooth map $f : M \to \mathbb{R}$ is the differentiable map defined on ∂M by $\frac{\partial f}{\partial n} := \langle \operatorname{grad}(f), n\rangle$, where n is the outward-pointing unit normal field on ∂M. Show that

$$\int_M (f_1\Delta f_2 - f_2\Delta f_1)\, \vartheta_M = -\int_{\partial M} \left(f_1\frac{\partial f_2}{\partial n} - f_2\frac{\partial f_1}{\partial n} \right) \vartheta_{\partial M}.$$

(l) Let M be a compact Riemannian manifold with boundary, and suppose that $\Delta f \equiv 0$ in $M \setminus \partial M$ and that one of the following boundary conditions holds:
  (i) $f|_{\partial M} \equiv 0$ (**Dirichlet condition**);
  (ii) $\frac{\partial f}{\partial n} \equiv 0$ (**Neumann condition**).

Show that $f \equiv 0$ in the first case, and that f is constant in the second case.

(8) (*Degree of a map*) Let M, N be compact, connected oriented manifolds of dimension n, and let $f : M \to N$ be a smooth map.

(a) Show that there exists a real number k (called the **degree** of f, and denoted by $\deg(f)$) such that, for any n-form $\omega \in \Omega^n(N)$,
$$\int_M f^*\omega = k \int_N \omega$$
(**Hint:** Use the Hodge decomposition).

(b) If f is not surjective then there exists an open set $W \subset N$ such that $f^{-1}(W) = \varnothing$. Deduce that if f is not surjective then $k = 0$.

(c) Show that if f is an orientation-preserving diffeomorphism then $\deg(f) = 1$, and that if f is an orientation-reversing diffeomorphism then $\deg(f) = -1$.

(d) Let $f : M \to N$ be surjective and let $q \in N$ be a regular value of f. Show that $f^{-1}(q)$ is a finite set and that there exists a neighborhood W of q in N such that $f^{-1}(W)$ is a disjoint union of opens sets V_i of M with $f|_{V_i} : V_i \to W$ a diffeomorphism.

(e) Admitting the existence of a regular value of f, show that $\deg(f)$ is an integer (**Remark: Sard's Theorem** guarantees that the set of critical values of a differentiable map f between manifolds with the same dimension has zero measure, which in turn guarantees the existence of a regular value of f).

(f) What is the degree of the natural projection $\pi : S^n \to \mathbb{R}P^n$ for n odd?

(g) Given $n \in \mathbb{N}$, indicate a smooth map $f : S^1 \to S^1$ of degree n.

(h) Let $S^n \subset \mathbb{R}^{n+1}$ be the unit sphere with the metric induced by the Euclidean metric of $\mathbb{R}^{n+1}$. Let X be a vector field tangent to S^n such that $\|X\| = 1$. Consider the map $F_t : S^n \to \mathbb{R}^{n+1}$ given by $F_t(x) = \cos(\pi t)x + \sin(\pi t)X_x$. Show that F_t is a smooth map of S^n on S^n, and define $k(t) = \deg(F_t)$. Show that the map $t \mapsto k(t)$ is continuous.

(i) What are the values of $k(0)$ and $k(1)$? Show that if n is even then there exists no vector field X on S^n such that $X_p \neq 0$ for all $p \in S^n$.

2. Affine Connections

If X and Y are vector fields in Euclidean space, we can define the **directional derivative** $\nabla_X Y$ of Y along X. This definition, however, uses the existence of Cartesian coordinates, which no longer holds in a general manifold. To overcome this difficulty we must introduce more structure:

Definition 2.1. *Let M be a differentiable manifold. An* **affine connection** *on M is a map $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ such that*

(i) $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$;
(ii) $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$;
(iii) $\nabla_X(fY) = (X \cdot f)Y + f\nabla_X Y$

for all $X, Y, Z \in \mathfrak{X}(M)$ *and* $f, g \in C^\infty(M, \mathbb{R})$ *(we write* $\nabla_X Y := \nabla(X, Y)$*).*

The vector field $\nabla_X Y$ is sometimes known as the **covariant derivative** of Y along X.

PROPOSITION 2.2. *Let* ∇ *be an affine connection on* M*, let* $X, Y \in \mathfrak{X}(M)$ *and* $p \in M$*. Then* $(\nabla_X Y)_p \in T_pM$ *depends only on* X_p *and on the values of* Y *along a curve tangent to* X *at* p*. Moreover, if* $x : W \to \mathbb{R}^n$ *are local coordinates on some open set* $W \subset M$*, we have*

$$\nabla_X Y = \sum_{i=1}^n \left(X \cdot Y^i + \sum_{j,k=1}^n \Gamma^i_{jk} X^j Y^k \right) \frac{\partial}{\partial x^i} \tag{6}$$

where the n^3 *differentiable functions* $\Gamma^i_{jk} : W \to \mathbb{R}$*, called the* **Christoffel symbols***, are defined by*

$$\nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k} = \sum_{i=1}^n \Gamma^i_{jk} \frac{\partial}{\partial x^i}. \tag{7}$$

PROOF. It is easy to show that an affine connection is **local**, that is, if $X, Y \in \mathfrak{X}(M)$ coincide with $\tilde{X}, \tilde{Y} \in \mathfrak{X}(M)$ in some open set $W \subset M$ then $\nabla_X Y = \nabla_{\tilde{X}} \tilde{Y}$ on W (see Exercise 2.6.1). Consequently, we can compute $\nabla_X Y$ for vector fields X, Y defined on W only. Let W be a coordinate neighborhood for the local coordinates $x : W \to \mathbb{R}^n$, and define the Christoffel symbols associated with these local coordinates through (7). Writing out

$$\nabla_X Y = \nabla_{\sum_{i=1}^n X^i \frac{\partial}{\partial x^i}} \sum_{j=1}^n Y^j \frac{\partial}{\partial x^j}$$

and using the properties listed in definition (2.1), we obtain (6). This formula clearly shows that $(\nabla_X Y)_p$ depends only on $X^i(p), Y^i(p)$ and $(X \cdot Y^i)(p)$. However, $X^i(p)$ and $Y^i(p)$ depend only on X_p and Y_p, and $(X \cdot Y^i)(p) = \frac{d}{dt} Y^i(c(t))|_{t=0}$ depends only on the values of Y^i (or Y) along a curve c whose tangent vector at $p = c(0)$ is X_p. $\square$

REMARK 2.3. Locally, an affine connection is uniquely determined by specifying its Christoffel symbols on a coordinate neighborhood. However, the choices of Christoffel symbols on different charts are not independent, as the covariant derivative must agree on the overlap.

A **vector field defined along a differentiable curve** $c : I \to M$ is a differentiable map $V : I \to TM$ such that $V(t) \in T_{c(t)}M$ for all $t \in I$. An obvious example is the tangent vector $\dot{c}(t)$. If V is a vector field defined along the differentiable curve $c : I \to M$ with $\dot{c} \neq 0$, its **covariant derivative** along c is the vector field defined along c given by

$$\frac{DV}{dt}(t) := \nabla_{\dot{c}(t)} V = (\nabla_X Y)_{c(t)}$$

for any vector fields $X, Y \in \mathfrak{X}(M)$ such that $X_{c(t)} = \dot{c}(t)$ and $Y_{c(s)} = V(s)$, with $s \in (t-\varepsilon, t+\varepsilon)$ for some $\varepsilon > 0$ (if $\dot{c}(t) \neq 0$ such extensions always exist). Proposition 2.2 guarantees that $(\nabla_X Y)_{c(t)}$ does not depend on the choice of X, Y; in fact, if in local coordinates $x : W \to \mathbb{R}^n$ we have $x^i(t) := x^i(c(t))$ and

$$V(t) = \sum_{i=1}^{n} V^i(t) \left(\frac{\partial}{\partial x^i} \right)_{c(t)},$$

then

$$\frac{DV}{dt}(t) = \sum_{i=1}^{n} \left(\dot{V}^i(t) + \sum_{j,k=1}^{n} \Gamma^i_{jk}(c(t)) \dot{x}^j(t) V^k(t) \right) \left(\frac{\partial}{\partial x^i} \right)_{c(t)}.$$

Definition 2.4. *A vector field V defined along a curve $c : I \to M$ is said to be* **parallel along** *c if*

$$\frac{DV}{dt}(t) = 0$$

for all $t \in I$. The curve c is said to be a **geodesic** *of the connection ∇ if $\dot{c}$ is parallel along c, i.e., if*

$$\frac{D\dot{c}}{dt}(t) = 0$$

for all $t \in I$.

In local coordinates $x : W \to \mathbb{R}^n$, the condition for V to be parallel along c is written as

$$\dot{V}^i + \sum_{j,k=1}^{n} \Gamma^i_{jk} \dot{x}^j V^k = 0 \quad (i = 1, \dots, n). \tag{8}$$

This is a system of first order linear ODE's for the components of V. By the Picard-Lindelöf Theorem, given a curve $c : I \to M$, a point $p \in c(I)$ and a vector $v_p \in T_pM$, there exists a unique vector field $V : I \to TM$ parallel along c such that $V(0) = v_p$, which is called the **parallel transport** of v_p along c.

Moreover, the geodesic equations are

$$\ddot{x}^i + \sum_{j,k=1}^{n} \Gamma^i_{jk} \dot{x}^j \dot{x}^k = 0 \quad (i = 1, \dots, n). \tag{9}$$

This is a system of second order (nonlinear) ODE's for the coordinates of $c(t)$. Therefore the Picard-Lindelöf Theorem implies that, given a point $p \in M$ and a vector $v_p \in T_pM$, there exists a unique geodesic $c : I \to M$, defined on a maximal open interval I such that $0 \in I$, satisfying $c(0) = p$ and $\dot{c}(0) = v_p$.

We will now define the torsion of an affine connection ∇. For that, we note that, in local coordinates $x : W \to \mathbb{R}^n$, we have

$$\begin{aligned}\nabla_X Y - \nabla_Y X &= \sum_{i=1}^n \left(X \cdot Y^i - Y \cdot X^i + \sum_{j,k=1}^n \Gamma^i_{jk} \left(X^j Y^k - Y^j X^k \right) \right) \frac{\partial}{\partial x^i} \\ &= [X,Y] + \sum_{i,j,k=1}^n \left(\Gamma^i_{jk} - \Gamma^i_{kj} \right) X^j Y^k \frac{\partial}{\partial x^i}.\end{aligned}$$

Definition 2.5. *The* **torsion operator** *of a connection ∇ on M is the operator $T : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ given by*

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y],$$

for all $X, Y \in \mathfrak{X}(M)$. The connection is said to be **symmetric** *if $T = 0$.*

The local expression of $T(X,Y)$ makes it clear that $T(X,Y)_p$ depends linearly on X_p and Y_p. In other words, T is the $(2,1)$-tensor field on M given in local coordinates by

$$T = \sum_{i,j,k=1}^n \left(\Gamma^i_{jk} - \Gamma^i_{kj} \right) dx^j \otimes dx^k \otimes \frac{\partial}{\partial x^i}$$

(recall that any $(2,1)$-tensor $T \in \mathcal{T}^{2,1}(V^*, V)$ is naturally identified with a bilinear map $\Phi_T : V^* \times V^* \to V \cong V^{**}$ through $\Phi_T(v,w)(\alpha) = T(v,w,\alpha)$ for all $v, w \in V,\ \alpha \in V^*$).

Notice that the connection is symmetric *iff* $\nabla_X Y - \nabla_Y X = [X,Y]$ for all $X, Y \in \mathfrak{X}(M)$. In local coordinates, the condition for the connection to be symmetric is

$$\Gamma^i_{jk} = \Gamma^i_{kj} \quad (i,j,k = 1, \ldots, n)$$

(hence the name).

Exercises 2.6.

(1) (a) Show that if $X, Y \in \mathfrak{X}(M)$ coincide with $\tilde{X}, \tilde{Y} \in \mathfrak{X}(M)$ in some open set $W \subset M$ then $\nabla_X Y = \nabla_{\tilde{X}} \tilde{Y}$ on W. (**Hint:** Use bump functions with support contained on W and the properties listed in definition (2.1)).
    (b) Obtain the local coordinate expression (6) for $\nabla_X Y$.
    (c) Obtain the local coordinate equations (8) for the parallel transport law.
    (d) Obtain the local coordinate equations (9) for the geodesics of the connection ∇.
(2) Determine all affine connections on $\mathbb{R}^n$. Of these, determine the connections whose geodesics are straight lines.
(3) Let ∇ be an affine connection on M. If $\omega \in \Omega^1(M)$ and $X \in \mathfrak{X}(M)$, we define $\nabla_X \omega \in \Omega^1(M)$ by

$$\nabla_X \omega(Y) = X \cdot (\omega(Y)) - \omega(\nabla_X Y)$$

for all $Y \in \mathfrak{X}(M)$.

(a) Show that this formula defines indeed a 1-form, i.e., show that $(\nabla_X\omega(Y))\,(p)$ is a linear function of Y_p.

(b) Show that
   (i) $\nabla_{fX+gY}\omega = f\nabla_X\omega + g\nabla_Y\omega$;
   (ii) $\nabla_X(\omega+\eta) = \nabla_X\omega + \nabla_X\eta$;
   (iii) $\nabla_X(f\omega) = (X\cdot f)\omega + f\nabla_X\omega$

   for all $X, Y \in \mathfrak{X}(M)$, $f, g \in C^\infty(M,\mathbb{R})$ and $\omega,\eta \in \Omega^1(M)$.

(c) Let $x : W \to \mathbb{R}^n$ be local coordinates on $W \subset M$, and take

$$\omega = \sum_{i=1}^n \omega_i dx^i$$

on W. Show that

$$\nabla_X\omega = \sum_{i=1}^n \left(X\cdot\omega_i - \sum_{j,k=1}^n \Gamma^k_{ji}X^j\omega_k \right) dx^i.$$

(d) Define $\nabla_X T$ for an arbitrary tensor field T in M, and write its expression in local coordinates.

3. Levi-Civita Connection

In the case of a Riemannian manifold, there is a special choice of connection called the **Levi-Civita connection**, with very important geometric properties.

Definition 3.1. *A connection ∇ in a Riemannian manifold $(M,\langle\cdot,\cdot\rangle)$ is said to be* **compatible** *with the metric if*

$$X\cdot\langle Y,Z\rangle = \langle\nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$$

for all $X, Y, Z \in \mathfrak{X}(M)$.

If ∇ is compatible with the metric, then the inner product of two vector fields V_1 and V_2, parallel along a curve, is constant along the curve:

$$\frac{d}{dt}\langle V_1(t), V_2(t)\rangle = \left\langle \nabla_{\dot c(t)}V_1(t), V_2(t)\right\rangle + \left\langle V_1(t), \nabla_{\dot c(t)}V_2(t)\right\rangle = 0.$$

In particular, parallel transport preserves lengths of vectors and angles between vectors. Therefore, if $c : I \to M$ is a geodesic, then $\langle\dot c(t),\dot c(t)\rangle^{\frac12} = k$ is constant. If $a \in I$, the length s of the geodesic between a and t is

$$s = \int_a^t \langle\dot c(u),\dot c(u)\rangle^{\frac12}\,du = \int_a^t k\,du = k(t-a).$$

In other words, t is an affine function of the arclength s (and is therefore called an **affine parameter**); this shows in particular that the parameters of two geodesics with the same image are affine functions of each other).

Theorem 3.2. (Levi-Civita) *If $(M,\langle\cdot,\cdot\rangle)$ is a Riemannian manifold then there exists a unique connection ∇ on M such that*

(i) ∇ *is symmetric;*
(ii) ∇ *is compatible with* $\langle\cdot,\cdot\rangle$.

In local coordinates $(x^1,\ldots,x^n)$*, the Christoffel symbols for this connection are*

$$\Gamma^i_{jk}=\frac{1}{2}\sum_{l=1}^n g^{il}\left(\frac{\partial g_{kl}}{\partial x^j}+\frac{\partial g_{jl}}{\partial x^k}-\frac{\partial g_{jk}}{\partial x^l}\right) \tag{10}$$

where $\left(g^{ij}\right)=(g_{ij})^{-1}$.

Proof. Let $X,Y,Z\in\mathfrak{X}(M)$. If the Levi-Civita connection exists then we must have

$$\begin{aligned} X\cdot\langle Y,Z\rangle &= \langle\nabla_XY,Z\rangle+\langle Y,\nabla_XZ\rangle;\\ Y\cdot\langle X,Z\rangle &= \langle\nabla_YX,Z\rangle+\langle X,\nabla_YZ\rangle;\\ -Z\cdot\langle X,Y\rangle &= -\langle\nabla_ZX,Y\rangle-\langle X,\nabla_ZY\rangle, \end{aligned}$$

as ∇ is compatible with the metric. Moreover, since ∇ is symmetric, we must also have

$$\begin{aligned} -\langle[X,Z],Y\rangle &= -\langle\nabla_XZ,Y\rangle+\langle\nabla_ZX,Y\rangle,\\ -\langle[Y,Z],X\rangle &= -\langle\nabla_YZ,X\rangle+\langle\nabla_ZY,X\rangle,\\ \langle[X,Y],Z\rangle &= \langle\nabla_XY,Z\rangle-\langle\nabla_YX,Z\rangle. \end{aligned}$$

Adding these six equalities, we obtain the **Koszul formula**

$$\begin{aligned} 2\langle\nabla_XY,Z\rangle &= X\cdot\langle Y,Z\rangle+Y\cdot\langle X,Z\rangle-Z\cdot\langle X,Y\rangle\\ &\quad-\langle[X,Z],Y\rangle-\langle[Y,Z],X\rangle+\langle[X,Y],Z\rangle. \end{aligned}$$

Since $\langle\cdot,\cdot\rangle$ is nondegenerate and Z is arbitrary, this formula determines ∇_XY. Thus, if the Levi-Civita connection exists, it must be unique.

To prove existence, we **define** ∇_XY through the Koszul formula. It is not difficult to show that this defines indeed a connection (cf. Exercise 3.3.1). Also, using this formula, we obtain

$$2\langle\nabla_XY-\nabla_YX,Z\rangle=2\langle\nabla_XY,Z\rangle-2\langle\nabla_YX,Z\rangle=2\langle[X,Y],Z\rangle$$

for all $X,Y,Z\in\mathfrak{X}(M)$, and hence ∇ is symmetric. Finally, again using the Koszul formula, we have

$$2\langle\nabla_XY,Z\rangle+2\langle Y,\nabla_XZ\rangle=2X\cdot\langle Y,Z\rangle$$

and therefore the connection defined by this formula is compatible with the metric.

Choosing local coordinates $(x^1,\ldots,x^n)$, we have

$$\left[\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\right]=0\quad\text{ and }\quad\left\langle\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\right\rangle=g_{ij}.$$

Therefore the Koszul formula yields

$$
\begin{aligned}
&2\left\langle \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l} \right\rangle = \frac{\partial}{\partial x^j} \cdot g_{kl} + \frac{\partial}{\partial x^k} \cdot g_{jl} - \frac{\partial}{\partial x^l} \cdot g_{jk} \\
&\Leftrightarrow \left\langle \sum_{i=1}^n \Gamma^i_{jk} \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^l} \right\rangle = \frac{1}{2} \left(\frac{\partial g_{kl}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l} \right) \\
&\Leftrightarrow \sum_{i=1}^n g_{il} \Gamma^i_{jk} = \frac{1}{2} \left(\frac{\partial g_{kl}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l} \right).
\end{aligned}
$$

□

EXERCISES 3.3.

(1) Show that the Koszul formula defines a connection.
(2) We introduce in $\mathbb{R}^3$, with the usual Euclidean metric $\langle \cdot, \cdot \rangle$, the connection ∇ defined in Cartesian coordinates (x^1, x^2, x^3) by

$$\Gamma^i_{jk} = \omega \varepsilon_{ijk},$$

where $\omega : \mathbb{R}^3 \to \mathbb{R}$ is a smooth function and

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) \text{ is an even permutation of } (1,2,3) \\ -1 & \text{if } (i,j,k) \text{ is an odd permutation of } (1,2,3) \\ 0 & \text{otherwise.} \end{cases}$$

Show that:
  (a) ∇ is compatible with $\langle \cdot, \cdot \rangle$;
  (b) the geodesics of ∇ are straight lines;
  (c) the torsion of ∇ is not zero in all points where $\omega \neq 0$ (therefore ∇ is not the Levi-Civita connection unless $\omega \equiv 0$);
  (d) the parallel transport equation is

$$\dot{V}^i + \sum_{j,k=1}^3 \omega \varepsilon_{ijk} \dot{x}^j V^k = 0 \Leftrightarrow \dot{V} + \omega(\dot{x} \times V) = 0$$

  (where $\times$ is the cross product in $\mathbb{R}^3$); therefore, a vector parallel along a straight line rotates about it with angular velocity $-\omega \dot{x}$.
(3) Let (M, g) and $(N, \tilde{g})$ be isometric Riemannian manifolds with Levi-Civita connections ∇ and $\widetilde{\nabla}$, and let $f : M \to N$ be an isometry. Show that:
  (a) $f_* \nabla_X Y = \widetilde{\nabla}_{f_* X} f_* Y$ for all $X, Y \in \mathfrak{X}(M)$;
  (b) if $c : I \to M$ is a geodesic then $f \circ c : I \to N$ is also a geodesic.
(4) Consider the usual local coordinates (θ, φ) in $S^2 \subset \mathbb{R}^3$ defined by the parametrization $\phi : (0, \pi) \times (0, 2\pi) \to \mathbb{R}^3$ given by

$$\phi(\theta, \varphi) = (\sin\theta \cos\varphi, \sin\theta \sin\varphi, \cos\theta).$$

(a) Using these coordinates, determine the expression of the Riemannian metric induced in S^2 by the usual Euclidean metric of $\mathbb{R}^3$.
(b) Compute the Christoffel symbols for the Levi-Civita connection in these coordinates.
(c) Show that the equator is the image of a geodesic.
(d) Show that any rotation about an axis through the origin in $\mathbb{R}^3$ induces an isometry of S^2.
(e) Show that the geodesics of S^2 traverse great circles.
(f) Find a geodesic triangle whose internal angles add up to $\frac{3\pi}{2}$.
(g) Let $c : \mathbb{R} \to S^2$ be given by $c(t) = (\sin\theta_0 \cos t, \sin\theta_0 \sin t, \cos\theta_0)$, where $\theta_0 \in \left(0, \frac{\pi}{2}\right)$ (therefore c is not a geodesic). Let V be a vector field parallel along c such that $V(0) = \frac{\partial}{\partial\theta}$ ($\frac{\partial}{\partial\theta}$ is well defined at $(\sin\theta_0, 0, \cos\theta_0)$ by continuity). Compute the angle by which V is rotated when it returns to the initial point. (**Remark:** The angle you have computed is exactly the angle by which the oscillation plane of the **Foulcaut pendulum** - which is just any sufficiently long and heavy pendulum - rotates during a day in a place at latitude $\frac{\pi}{2} - \theta_0$, as it tries to remain fixed with respect to the stars in a rotating Earth).
(h) Use this result to prove that no open set $U \subset S^2$ is isometric to an open set $V \subset \mathbb{R}^2$ with the Euclidean metric.
(i) Given a geodesic $c : \mathbb{R} \to \mathbb{R}^2$ of $\mathbb{R}^2$ with the Euclidean metric and a point $p \notin c(\mathbb{R})$, there exists a unique geodesic $\tilde{c} : \mathbb{R} \to \mathbb{R}^2$ (up to reparametrization) such that $p \in \tilde{c}(\mathbb{R})$ and $c(\mathbb{R}) \cap \tilde{c}(\mathbb{R}) = \varnothing$ (**parallel postulate**). Is this true in S^2?

(5) Let H be the group of proper affine transformations of $\mathbb{R}$, that is, the group of functions $g : \mathbb{R} \to \mathbb{R}$ of the form

$$g(t) = yt + x$$

with $y > 0$ and $x \in \mathbb{R}$ (the group operation being composition). Taking $(x, y) \in \mathbb{R} \times \mathbb{R}^+$ as global coordinates, we induce a differentiable structure in H, and H, with this differentiable structure, is a Lie group (cf. Exercise 7.17.3 in Chapter 1).

(a) Determine the left-invariant metric induced by the Euclidean inner product

$$g = dx \otimes dx + dy \otimes dy$$

in $\mathfrak{h} = T_{(0,1)}H$ (H endowed with this metric is called the **hyperbolic plane**).
(b) Compute the Christoffel symbols of the Levi-Civita connection in the coordinates (x, y).

(c) Show that the curves $\alpha, \beta : \mathbb{R} \to H$ given in these coordinates by

$$\alpha(t) = \big(0, e^t\big)$$
$$\beta(t) = \left(\tanh t, \frac{1}{\cosh t}\right)$$

are geodesics. What are the sets $\alpha(\mathbb{R})$ and $\beta(\mathbb{R})$?

(d) Determine all images of geodesics.

(e) Show that, given two points $p, q \in H$, there exists a unique geodesic through them (up to reparametrization).

(f) Give examples of connected Riemannian manifolds containing two points through which there are (i) infinitely many geodesics (up to reparametrization); (ii) no geodesics.

(g) Show that no open set $U \subset H$ is isometric to an open set $V \subset \mathbb{R}^2$ with the Euclidean metric. (**Hint:** Show that in any neighborhood of any point $p \in H$ there is always a geodesic quadrilateral whose internal angles add up to less than 2π).

(h) Does the parallel postulate hold in the hyperbolic plane?

(6) Let $(M, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold with Levi-Civita connection $\widetilde{\nabla}$, and let $(N, \langle\langle\cdot,\cdot\rangle\rangle)$ be a submanifold with the induced metric and Levi-Civita connection ∇.

(a) Show that

$$\nabla_X Y = \left(\widetilde{\nabla}_{\widetilde{X}} \widetilde{Y}\right)^{\top}$$

for all $X, Y \in \mathfrak{X}(N)$, where $\widetilde{X}, \widetilde{Y}$ are any extensions of X, Y to $\mathfrak{X}(M)$ and $^{\top} : TM|_N \to TN$ is the orthogonal projection.

(b) Use this result to indicate curves that are, and curves that are not, geodesics of the following surfaces in $\mathbb{R}^3$:

(i) the sphere S^2;
(ii) the torus of revolution;
(iii) the surface of a cone;
(iv) a general surface of revolution.

(c) Show that if two surfaces in $\mathbb{R}^3$ are tangent along a curve, then the parallel transport of vectors along this curve in both surfaces coincides.

(d) Use this result to compute the angle $\Delta\theta$ by which a vector V is rotated when it is parallel transported along a circle on the sphere (**Hint:** Consider the cone which is tangent to the sphere along the circle (cf. Figure 1); notice that the cone minus a ray through the vertex is isometric to an open set of the Euclidean plane).

(7) Let (M, g) be a Riemannian manifold with Levi-Civita connection ∇. Show that g is **parallel** along any curve, i.e., show that

$$\nabla_X g = 0$$

for all $X \in \mathfrak{X}(M)$ (cf. Exercise 2.6.3).

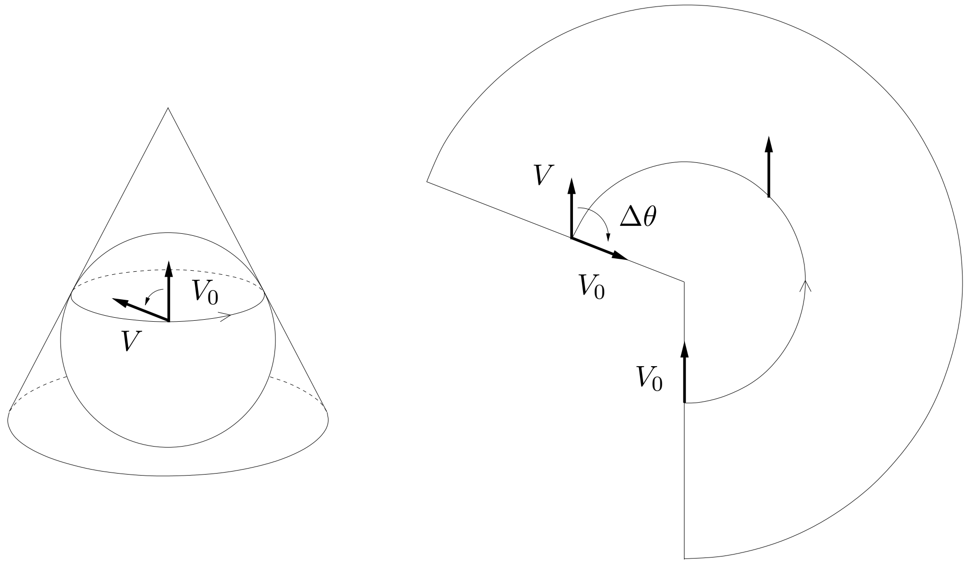

FIGURE 1. Parallel transport along a circle on the sphere.

(8) Let (M, g) be a Riemannian manifold with Levi-Civita connection ∇, and let $\psi_t : M \to M$ be a one-parameter group of isometries. The vector field $X \in \mathfrak{X}(M)$ defined by

$$X_p = \left.\frac{d}{dt}\right|_{t=0} \psi_t(p)$$

is called the **Killing vector field** associated to ψ_t. Show that:
  (a) $L_X g = 0$ (cf. Exercise 2.8.3);
  (b) X satisfies $\langle \nabla_Y X, Z\rangle + \langle \nabla_Z X, Y\rangle = 0$ for all vector fields $Y, Z \in \mathfrak{X}(M)$;
  (c) if $c : I \to M$ is a geodesic then $\left\langle \dot{c}(t), X_{c(t)} \right\rangle$ is constant.

(9) Recall that if M is an oriented differential manifold with volume element $\omega \in \Omega^n(M)$, the **divergence** of X is the function $\operatorname{div}(X)$ such that

$$L_X \omega = (\operatorname{div}(X))\omega$$

(cf. Exercise 6.4.5 in Chapter 2). Suppose that M has a Riemannian structure and ω is a Riemannian volume element.
  (a) Show that this definition of divergence coincides with the definition in Exercise 1.11.7.
  (b) Show that at each point $p \in M$,

$$\operatorname{div}(X) = \sum_{i=1}^{n} \langle \nabla_{Y_i} X, Y_i \rangle,$$

  where $\{Y_1, \ldots, Y_n\}$ is an orthonormal basis of T_pM and ∇ is the Levi-Civita connection.

(10) Let M be an oriented Riemannian manifold of dimension 3. The **curl** of a vector field $X \in \mathfrak{X}(M)$ is the vector field $\operatorname{curl}(X)$ associated to the 1-form $*d\omega_X$, where $\omega_X \in \Omega^1(M)$ is the 1-form associated to X (cf. Exercise 1.11.7). Show that:
 (a) $\operatorname{curl}(\operatorname{grad}(f)) = 0$ for $f \in C^\infty(M, \mathbb{R})$;
 (b) $\operatorname{div}(\operatorname{curl}(X)) = 0$ for $X \in \mathfrak{X}(M)$;
 (c) $\operatorname{curl}(X) = \sum_{i,j,k=1}^{3} \varepsilon_{ijk} \left\langle \nabla_{Y_j} X, Y_k \right\rangle Y_i$, where $\{Y_1, Y_2, Y_3\}$ is a positive basis of orthonormal vector fields, $X = \sum_{i=1}^{n} X^i Y_i$ and ε_{ijk} was defined on Exercise 3.3.2.

4. Minimizing Properties of Geodesics

Let M be a differentiable manifold with an affine connection ∇. As we saw in Section 2, given a point $p \in M$ and a tangent vector $v \in T_pM$, there exists a unique geodesic $c_v : I \to M$ defined on a maximal open interval $I \subset \mathbb{R}$ such that $0 \in I$, $c_v(0) = p$ and $\dot{c_v}(0) = v$. Consider now the curve $\gamma : J \to M$ defined by $\gamma(t) = c_v(at)$, where $a \in \mathbb{R}$ and J is the inverse image of I by the map $t \mapsto at$. We have

$$\dot{\gamma}(t) = a\dot{c}_v(at)$$

and, consequently,

$$\nabla_{\dot{\gamma}}\dot{\gamma} = \nabla_{a\dot{c}_v}(a\dot{c}_v) = a^2 \nabla_{\dot{c}_v}\dot{c}_v = 0.$$

Therefore γ is also a geodesic. Since $\gamma(0) = c_v(0) = p$ and $\dot{\gamma}(0) = a\dot{c}_v(0) = av$, we see that γ is the unique geodesic with initial velocity $av \in T_pM$, that is, $\gamma = c_{av}$. Therefore, we have $c_{av}(t) = c_v(at)$ for all $t \in I$. This property is sometimes referred to as the **homogeneity** of geodesics. Notice that we can make the interval J arbitrarily large by making a sufficiently small.

If $1 \in I$, we define $\exp_p(v) = c_v(1)$. By homogeneity of geodesics, we can define $\exp_p(v)$ for v in some open neighborhood U of the origin in T_pM. The map $\exp_p : U \subset T_pM \to M$ thus obtained is called the **exponential map** at p.

Proposition 4.1. *There exists an open set $U \subset T_pM$ containing the origin such that $\exp_p : U \to M$ is a diffeomorphism onto some open set $V \subset M$ containing p (called a* **normal neighborhood***).*

Proof. The exponential map is clearly differentiable as a consequence of the smooth dependence of the solution of an ODE on its initial data (cf. [**Arn92**]). If $v \in T_pM$ is such that $\exp_p(v)$ is defined, we have, by homogeneity, that $\exp_p(tv) = c_{tv}(1) = c_v(t)$. Consequently,

$$\left(d\exp_p\right)_0 v = \frac{d}{dt}\exp_p(tv)|_{t=0} = \frac{d}{dt}c_v(t)|_{t=0} = v.$$

We conclude that $\left(d\exp_p\right)_0 : T_0(T_pM) \cong T_pM \to T_pM$ is the identity map. By the Inverse Function Theorem, $\exp_p$ is then a diffeomorphism of some

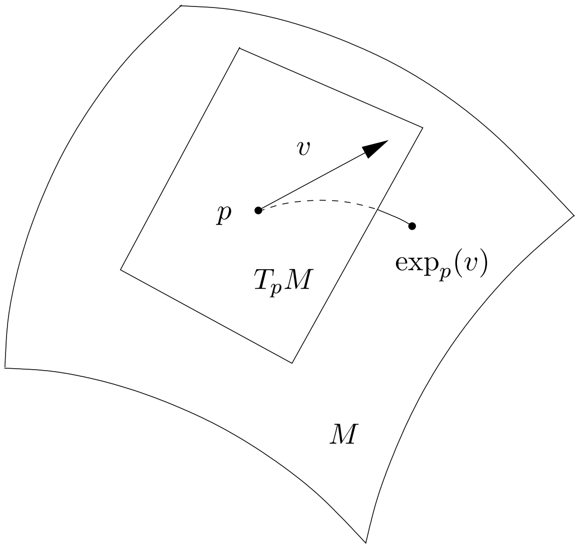

FIGURE 2. The exponential map.

open neighbourhood U of $0 \in T_pM$ onto some open set $V \subset M$ containing $p = \exp_p(0)$. $\square$

EXAMPLE 4.2. Consider the Levi-Civita connection in S^2 with the standard metric, and let $p \in S^2$. Then $\exp_p(v)$ is well defined for all $v \in T_pS^2$, but is not a diffeomorphism, as it clearly is not injective. However, its restriction to the open ball $B_\pi(0) \subset T_pS^2$ is a diffeomorphism onto $S^2 \setminus \{-p\}$.

Now let $(M, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold and ∇ its Levi-Civita connection. Since $\langle\cdot,\cdot\rangle$ defines an inner product in T_pM, we can think of T_pM as the Euclidean n-space $\mathbb{R}^n$.

Let E be the vector field defined on $T_pM \setminus \{0\}$ by

$$E_v = \frac{v}{\|v\|},$$

and define $X = (\exp_p)_* E$ on $V \setminus \{p\}$, where $V \subset M$ is a normal neighborhood. We have

$$\begin{aligned} X_{\exp_p(v)} &= \left(d\exp_p\right)_v E_v = \frac{d}{dt}\exp_p\left(v + t\frac{v}{\|v\|}\right)\Big|_{t=0} \\ &= \frac{d}{dt}c_v\left(1 + \frac{t}{\|v\|}\right)\Big|_{t=0} = \frac{1}{\|v\|}\dot{c}_v(1). \end{aligned}$$

Since $\|\dot{c}_v(1)\| = \|\dot{c}_v(0)\| = \|v\|$, we see that $X_{\exp_p(v)}$ is the unit tangent vector to the geodesics c_v. In particular, X must satisfy

$$\nabla_X X = 0.$$

If $\varepsilon > 0$ is such that $\overline{B_\varepsilon(0)} \subset U := \exp_p^{-1}(V)$, the **normal ball** with center p and radius ε is the open set $B_\varepsilon(p) := \exp_p(B_\varepsilon(0))$, and the **normal sphere** of radius ε centered at p is the compact submanifold $S_\varepsilon(p) := \exp_p(\partial B_\varepsilon(0))$. We will now prove that X is (and hence the geodesics through p are) orthogonal to normal spheres.

For that, we choose a local parametrization $\varphi : W \subset \mathbb{R}^{n-1} \to S^{n-1} \subset T_pM$, and use it to define a parametrization $\tilde{\varphi} : (0, +\infty) \times W \to T_pM$ through

$$\tilde{\varphi}(r, \theta^1, \ldots, \theta^{n-1}) = r\varphi(\theta^1, \ldots, \theta^{n-1})$$

(hence $(r, \theta^1, \ldots, \theta^{n-1})$ are spherical coordinates on T_pM). Notice that

$$\frac{\partial}{\partial r} = E,$$

and consequently

$$X = (\exp_p)_* \frac{\partial}{\partial r}. \tag{11}$$

Since $\frac{\partial}{\partial \theta^i}$ is tangent to $\{r = \varepsilon\}$, the vector fields

$$Y_i := (\exp_p)_* \frac{\partial}{\partial \theta^i} \tag{12}$$

are tangent to $S_\varepsilon(p)$. Notice also that $\left\|\frac{\partial}{\partial \theta^i}\right\| = \left\|\frac{\partial \tilde{\varphi}}{\partial \theta^i}\right\| = r \left\|\frac{\partial \varphi}{\partial \theta^i}\right\|$ is proportional to r, and consequently $\frac{\partial}{\partial \theta^i} \to 0$ as $r \to 0$, implying that $(Y_i)_q \to 0_p$ as $q \to p$.

Since $\exp_p$ is a local diffeomorphism, the vector fields X and Y_i are linearly independent at each point. Also,

$$[X, Y_i] = \left[(\exp_p)_* \frac{\partial}{\partial r}, (\exp_p)_* \frac{\partial}{\partial \theta^i}\right] = (\exp_p)_* \left[\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta^i}\right] = 0,$$

or, since the Levi-Civita connection is symmetric,

$$\nabla_X Y_i = \nabla_{Y_i} X.$$

To prove that X is orthogonal to the normal spheres $S_\varepsilon(p)$, we show that X is orthogonal to each of the vector fields Y_i. In fact, since $\nabla_X X = 0$ and $\|X\| = 1$, we have

$$X \cdot \langle X, Y_i \rangle = \langle \nabla_X X, Y_i \rangle + \langle X, \nabla_X Y_i \rangle = \langle X, \nabla_{Y_i} X \rangle = Y_i \cdot \left(\frac{1}{2}\langle X, X\rangle\right) = 0,$$

and hence $\langle X, Y_i \rangle$ is constant along each geodesic through p. Consequently,

$$\langle X, Y_i \rangle(\exp_p v) = \left\langle X_{\exp_p(v)}, (Y_i)_{\exp_p(v)} \right\rangle = \lim_{t \to 0} \left\langle X_{\exp_p(tv)}, (Y_i)_{\exp_p(tv)} \right\rangle = 0$$

(as $\|X\| = 1$ and $(Y_i)_q \to 0_p$ as $q \to p$).

PROPOSITION 4.3. *Let $\gamma : I \to M$ be a differentiable curve such that $\gamma(0) = p$, $\gamma(1) \in S_\varepsilon(p)$, where $S_\varepsilon(p)$ is a normal sphere. Then the length $l(\gamma)$ of the restriction of γ to $[0,1]$ satisfies $l(\gamma) \geq \varepsilon$, and $l(\gamma) = \varepsilon$ if and only if γ is a reparametrized geodesic.*

PROOF. We can assume that $\gamma(t) \neq p$ for all $t \in (0,1)$: if that were not so, we could easily construct a curve $\tilde{\gamma} : \tilde{I} \to M$ with $\tilde{\gamma}(0) = p$, $\tilde{\gamma}(1) = \gamma(1) \in S_\varepsilon(p)$ and $l(\tilde{\gamma}) < l(\gamma)$. For the same reason, we can assume that $\gamma([0,1)) \subset B_\varepsilon(p)$. Let

$$\gamma(t) = \exp_p(r(t)n(t)),$$

where $r(t) \in (0,\varepsilon]$ and $n(t) \in S^{n-1}$ are well defined for $t \in (0,1]$. Note that r can be extended to $[0,1]$ as a smooth function. We have

$$\dot{\gamma}(t) = (\exp_p)_* \left(\dot{r}(t)n(t) + r(t)\dot{n}(t)\right).$$

Since $\langle n(t), n(t)\rangle = 1$, we have $\langle \dot{n}(t), n(t)\rangle = 0$, and consequently $\dot{n}(t)$ is tangent to $\partial B_{r(t)}(0)$. Noticing that $n(t) = \left(\frac{\partial}{\partial r}\right)_{r(t)n(t)}$, we conclude that

$$\dot{\gamma}(t) = \dot{r}(t)X_{\gamma(t)} + Y(t),$$

where $Y(t) = r(t)(\exp_p)_*\dot{n}(t)$ is tangent to $S_{r(t)}(p)$, and hence orthogonal to $X_{\gamma(t)}$. Consequently,

$$\begin{aligned} l(\gamma) &= \int_0^1 \left\langle \dot{r}(t)X_{\gamma(t)} + Y(t), \dot{r}(t)X_{\gamma(t)} + Y(t)\right\rangle^{\frac{1}{2}} dt \\ &= \int_0^1 \left(\dot{r}(t)^2 + \|Y(t)\|^2\right)^{\frac{1}{2}} dt \\ &\geq \int_0^1 \dot{r}(t)dt = r(1) - r(0) = \varepsilon. \end{aligned}$$

It should be clear that $l(\gamma) = \varepsilon$ if and only if $\|Y(t)\| = 0$ and $\dot{r}(t) \geq 0$ for all $t \in [0,1]$; but then $\dot{n}(t) = 0$ (implying that n is constant), and $\gamma(t) = \exp_p(r(t)n) = c_{r(t)n}(1) = c_n(r(t))$ is, up to reparametrization, the geodesic through p with initial condition $n \in T_pM$. $\square$

DEFINITION 4.4. *A* **piecewise differentiable curve** *is a continuous map $c : [a,b] \to M$ such that the restriction of c to $[t_{i-1}, t_i]$ coincides with the restriction of a differentiable curve to the same interval for $i = 1, \dots, n$, where $a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$. We say that c* **connects** *$p \in M$ to $q \in M$ if $c(a) = p$ and $c(b) = q$.*

The definition of **length** of a piecewise differentiable curve offers no difficulties. It should also be clear that Proposition 4.3 easily extends to piecewise differentiable curves, if we now allow for piecewise differentiable reparametrizations. Using this extended version of Proposition 4.3, the properties of the exponential map and the invariance of length under reparametrization, one easily shows the following result:

THEOREM 4.5. *Let $(M, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold, $p \in M$ and $B_\varepsilon(p)$ a normal ball centered at p. Then, for each point $q \in B_\varepsilon(p)$, there exists a geodesic $c : I \to M$ connecting p to q; moreover, if $\gamma : J \to M$ is any other piecewise differentiable curve connecting p to q, then $l(\gamma) \geq l(c)$, and $l(\gamma) = l(c)$ if and only if γ is a reparametrization of c.*

Conversely, we have

THEOREM 4.6. *Let $(M, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold and $p, q \in M$. If $c : I \to M$ is a piecewise differentiable curve connecting p to q and $l(c) \leq l(\gamma)$ for any piecewise differentiable curve $\gamma : J \to M$ connecting p to q then c is a reparametrized geodesic.*

To prove this theorem, we need the following definition:

DEFINITION 4.7. *A normal neighborhood $V \subset M$ is called a* **totally normal neighborhood** *if there exists $\varepsilon > 0$ such that $V \subset B_\varepsilon(p)$ for all $p \in V$.*

We will now prove that totally normal neighborhoods always exist. To do so, we recall that local coordinates $(x^1, \ldots, x^n)$ on M yield local coordinates $(x^1, \ldots, x^n, v^1, \ldots, v^n)$ on TM labeling the vector

$$v^1 \frac{\partial}{\partial x^1} + \ldots + v^n \frac{\partial}{\partial x^n}.$$

The geodesic equations,

$$\ddot{x}^i + \sum_{j,k=1}^n \Gamma^i_{jk} \dot{x}^j \dot{x}^k = 0 \qquad (i = 1, \ldots, n),$$

correspond to the system of first order ODE's

$$\begin{cases} \dot{x}^i = v^i \\ \dot{v}^i = -\sum_{j,k=1}^n \Gamma^i_{jk} v^j v^k \end{cases} \qquad (i = 1, \ldots, n).$$

These equations define the local flow of the vector field $X \in \mathfrak{X}(TM)$ given in local coordinates by

$$X = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i} - \sum_{i,j,k=1}^n \Gamma^i_{jk} v^j v^k \frac{\partial}{\partial v^i},$$

called the **geodesic flow**. As was seen in Chapter 1, for each point $v \in TM$ there exists an open neighborhood $W \subset TM$ and an open interval $I \subset \mathbb{R}$ containing 0 such that the local flow $F : W \times I \to TM$ of X is well defined. In particular, for each point $p \in M$ we can choose an open neighborhood U containing p and $\varepsilon > 0$ such that the geodesic flow is well defined in $W \times I$ with

$$W = \{v_q \in TM \mid q \in U, \|v_q\| < \varepsilon\}.$$

Using homogeneity of geodesics, we can make the interval I as large as we want by making ε sufficiently small. Therefore, for ε small enough we can define a map $G : W \to M \times M$ by $G(v_q) := (q, \exp_q(v_q))$. Since $\exp_q(0_q) = q$, the matrix representation of $(dG)_{0_p}$ in the above local coordinates is $\left(\begin{smallmatrix} I & 0 \\ I & I \end{smallmatrix}\right)$, and hence G is a local diffeomorphism. Reducing U and ε if necessary, we can therefore assume that G is a diffeomorphism onto its image $G(W)$, which contains the point $(p, p) = G(0_p)$. Choosing an open neighborhood V of p such that $V \times V \subset G(W)$, it is clear that V is a totally normal neighborhood:

for each point $q \in V$ we have $\{q\} \times \exp_q(B_\varepsilon(0_q)) = G(W) \cap (\{q\} \times M) \supset \{q\} \times V$, that is, $\exp_q(B_\varepsilon(0_q)) \supset V$.

Notice that, given any two points p, q in a totally normal neighborhood V, there exists a geodesic $c : I \to M$ connecting p to q; if $\gamma : J \to M$ is any other piecewise differentiable curve connecting p to q, then $l(\gamma) \geq l(c)$, and $l(\gamma) = l(c)$ if and only if γ is a reparametrization of c. The proof of Theorem 4.6 is now an immediate consequence of the following observation: if $c : I \to M$ is a piecewise differentiable curve connecting p to q such that $l(c) \leq l(\gamma)$ for any curve $\gamma : J \to M$ connecting p to q, we see that c must be a reparametrized geodesic in each totally normal neighborhood it intersects.

Exercises 4.8.

(1) Let (M, g) be a Riemannian manifold and $f : M \to \mathbb{R}$ a smooth function. Show that if $\|\operatorname{grad}(f)\| \equiv 1$ then the integral curves of $\operatorname{grad}(f)$ are geodesics.

(2) Let M be a Riemannian manifold and ∇ the Levi-Civita connection on M. Given $p \in M$ and a basis $\{v_1, \ldots, v_n\}$ for T_pM, we consider the parametrization $\varphi : U \subset \mathbb{R}^n \to M$ given by

$$\varphi(x^1, \ldots, x^n) = \exp_p(x^1 v_1 + \ldots + x^n v_n)$$

(the local coordinates $(x^1, \ldots, x^n)$ are called **normal coordinates**). Show that:

 (a) in these coordinates, $\Gamma^i_{jk}(p) = 0$ (**Hint:** Consider the geodesic equation);

 (b) if $\{v_1, \ldots, v_n\}$ is an orthonormal basis then $g_{ij}(p) = \delta_{ij}$.

(3) Let G be a Lie group endowed with a bi-invariant Riemannian metric (i.e., such that L_x and R_x are isometries for all $x \in G$), and let $i : G \to G$ be the diffeomorphism defined by $i(x) = x^{-1}$.

 (a) Compute $(di)_e$ and show that

$$(di)_x = (dR_{x^{-1}})_e \,(di)_e \,(dL_{x^{-1}})_x$$

 for all $x \in G$. Conclude that i is an isometry.

 (b) Let $v \in \mathfrak{g} = T_eG$ and c_v be the geodesic satisfying $c_v(0) = e$ and $\dot{c}_v(0) = v$. Show that if t is sufficiently small then $c_v(-t) = (c_v(t))^{-1}$. Conclude that c_v is defined in $\mathbb{R}$ and satisfies $c_v(t+s) = c_v(t)c_v(s)$ for all $t, s \in \mathbb{R}$ (**Hint:** Recall that any two points in a totally normal neighborhood are connected by a unique geodesic in that neighbourhood).

 (c) Show that the geodesics of G are the integral curves of left-invariant vector fields, and that the maps exp (in the Lie group) and $\exp_e$ (in the Riemannian manifold) coincide.

 (d) Let ∇ be the Levi-Civita connection of the bi-invariant metric and X, Y two left-invariant vector fields. Show that

$$\nabla_X Y = \frac{1}{2}[X, Y].$$

(4) Use Theorem 4.5 to prove that if $f : M \to N$ is an isometry and $c : I \to M$ is a geodesic then $f \circ c : I \to N$ is also a geodesic.
(5) Let $f : M \to M$ be an isometry whose set of fixed points is a connected 1-dimensional submanifold $N \subset M$. Show that N is the image of a geodesic.
(6) Let $(M, \langle \cdot, \cdot \rangle)$ be a geodesically complete Riemannian manifold and let $p \in M$.
   (a) Consider a geodesic $c : \mathbb{R} \to M$ parametrized by the arclength such that $c(0) = p$. Let X and Y_i be the vector fields defined as in (11) and (12) (so that $X_{c(t)} = \dot{c}(t)$). Show that Y_i satisfies the **Jacobi equation**

$$\frac{D^2 Y_i}{dt^2} = R(X, Y_i)X,$$

   where $R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, defined by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z,$$

   is called the **curvature operator** (cf. Chapter 4). A solution of the Jacobi equation is called a **Jacobi field**.
   (b) Show that Y is a Jacobi field with $Y(0) = 0$ if and only if

$$Y(t) = \left.\frac{\partial}{\partial \alpha}\right|_{\alpha=0} \gamma(t, \alpha),$$

   where $\gamma : \mathbb{R} \times (-\varepsilon, \varepsilon) \to M$ is such that $\gamma(t, 0) = c(t)$ and for each α the curve $\gamma(t, \alpha)$ is a geodesic with $\gamma(0, \alpha) = p$.
   (c) A point $q \in M$ is said to be **conjugate** to p if it is a critical value of $\exp_p$. Show that q is conjugate to p if and only if there exists a nonvanishing Jacobi field Y along a geodesic c connecting $p = c(0)$ to $q = c(r)$ such that $Y(0) = Y(r) = 0$. Conclude that if q is conjugate to p then p is conjugate to q.
   (d) The manifold M is said to have **nonpositive curvature** if $\langle R(X, Y)X, Y \rangle \geq 0$ for all $X, Y \in \mathfrak{X}(M)$. Show that for such a manifold no two points are conjugate.
   (e) Given a geodesic $c : I \to M$ parametrized by the arclength such that $c(0) = p$, let t_c be the supremum of the set of values of t such that c is the minimizing curve connecting p to $c(t)$ (hence $t_c > 0$). The **cut locus** of p is defined to be the set of all points of the form $c(t_c)$ for $t_c < +\infty$. Determine the cut locus of a given point $p \in M$ when M is:
      (i) the torus T^n with the standard metric.
      (ii) the sphere S^n with the standard metric;
      (iii) the projective space $\mathbb{R}P^n$ with the standard metric.

   Check that any point in the cut locus is either conjugate to p or joined to p by two geodesic arcs with the same length but different images.

5. Hopf-Rinow Theorem

Let (M, g) be a Riemannian manifold. The existence of totally normal neighborhoods implies that it is always possible to connect two sufficiently close points $p, q \in M$ by a minimizing geodesic. We now address the same question globally.

EXAMPLE 5.1.

(1) Given two distinct points $p, q \in \mathbb{R}^n$ there exists a unique (up to reparametrization) geodesic arc for the Euclidean metric connecting them.
(2) Given two distinct points $p, q \in S^n$ there exist at least two geodesic arcs for the standard metric connecting them which are not reparametrizations of each other.
(3) If $p \neq 0$ then there exists no geodesic arc for the Euclidean metric in $\mathbb{R}^n \setminus \{0\}$ connecting p to $-p$.

In many cases (for example in $\mathbb{R}^n \setminus \{0\}$) there exist geodesics which cannot be extended for all values of its parameter. In other words, $\exp_p(v)$ is not defined for all $v \in T_pM$.

DEFINITION 5.2. *A Riemannian manifold* $(M, \langle \cdot, \cdot \rangle)$ *is said to be* **geodesically complete** *if, for every point* $p \in M$*, the map* $\exp_p$ *is defined in* T_pM*.*

There exists another notion of completeness of a connected Riemannian manifold, coming from the fact that any such manifold is naturally a metric space.

DEFINITION 5.3. *Let* $(M, \langle \cdot, \cdot \rangle)$ *be a connected Riemannian manifold and* $p, q \in M$*. The* **distance** *between* p *and* q *is defined as*

$$d(p, q) = \inf\{l(\gamma) \mid \gamma \textit{ is a piecewise differentiable curve connecting } p \textit{ to } q\}.$$

Notice that if there exists a minimizing geodesic c connecting p to q then $d(p, q) = l(c)$. The function $d : M \times M \to [0, +\infty)$ is indeed a distance, as stated in the following proposition (whose proof is left as an exercise):

PROPOSITION 5.4. (M, d) *is a metric space, that is,*

(i) $d(p, q) \geq 0$ *and* $d(p, q) = 0$ *if and only if* $p = q$*;*
(ii) $d(p, q) = d(q, p)$*;*
(iii) $d(p, r) \leq d(p, q) + d(q, r)$*,*

for all $p, q, r \in M$*. The metric space topology induced on* M *coincides with the topology of* M *as a differentiable manifold.*

Therefore, we can discuss the completeness of M as a metric space (that is, whether Cauchy sequences converge). The fact that completeness and geodesic completeness are equivalent is the content of

THEOREM 5.5. (Hopf-Rinow) *Let* $(M, \langle \cdot, \cdot \rangle)$ *be a connected Riemannian manifold and* $p \in M$*. The following assertions are equivalent:*

(i) M *is geodesically complete.*
(ii) (M, d) *is a complete metric space;*
(iii) $\exp_p$ *is defined in* T_pM.

Moreover, if $(M, \langle \cdot, \cdot \rangle)$ *is geodesically complete then for all* $q \in M$ *there exists a geodesic* c *connecting* p *to* q *with* $l(c) = d(p, q)$.

PROOF. It is clear that $(i) \Rightarrow (iii)$.

We begin by showing that if (iii) holds then for all $q \in M$ there exists a geodesic c connecting p to q with $l(c) = d(p, q)$. Let $d(p, q) = \rho$. If $\rho = 0$ then $q = p$ and there is nothing to prove. If $\rho > 0$, let $\varepsilon \in (0, \rho)$ be such that $S_\varepsilon(p)$ is a normal sphere (which is a compact submanifold of M). The continuous function $x \mapsto d(x, q)$ will then have a point of minimum $x_0 \in S_\varepsilon(p)$. Moreover, $x_0 = \exp_p(\varepsilon v)$, where $\|v\| = 1$. Let us consider the geodesic $c_v(t) = \exp_p(tv)$. We will show that $q = c_v(\rho)$. For that, we consider the set

$$A = \{t \in [0, \rho] \mid d(c_v(t), q) = \rho - t\}.$$

Since the map $t \mapsto d(c_v(t), q)$ is continuous, A is a closed set. Moreover,

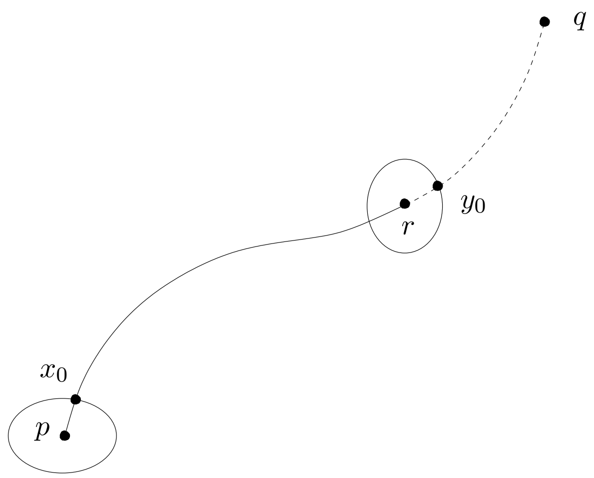

FIGURE 3. Proof of the Hopf-Rinow Theorem.

$A \neq \varnothing$, as clearly $0 \in A$. We will now show that no point $t_0 \in [0, \rho)$ can be the maximum of A, which implies that the maximum of A must be ρ, and consequently that $d(c_v(\rho), q) = 0$, i.e., $c_v(\rho) = q$ (hence c_v connects p to q and $l(c_v) = \rho$). Let $t_0 \in A \cap [0, \rho)$, $r = c_v(t_0)$ and $\delta \in (0, \rho - t_0)$ such that $S_\delta(r)$ is a normal sphere. Let y_0 be a point of minimum of the continuous function $y \mapsto d(y, q)$ on the compact set $S_\delta(r)$. Then $y_0 = c_v(t_0 + \delta)$. In fact, we have

$$\rho - t_0 = d(r, q) = \delta + \min_{y \in S_\delta(r)} d(y, q) = \delta + d(y_0, q),$$

and so

$$d(y_0, q) = \rho - t_0 - \delta. \tag{13}$$

The triangular inequality then implies that

$$d(p, y_0) \geq d(p, q) - d(y_0, q) = \rho - (\rho - t_0 - \delta) = t_0 + \delta,$$

and since the piecewise differentiable curve which connects p to r through c_v and r to y_0 through a geodesic arc has length $t_0 + \delta$, we conclude that this is a minimizing curve, hence a (reparametrized) geodesic. Therefore, $y_0 = c_v(t_0 + \delta)$. Consequently, equation (13) can be written as

$$d(c_v(t_0 + \delta), q) = \rho - (t_0 + \delta),$$

indicating that $t_0 + \delta \in A$. Therefore t_0 cannot be the maximum of A.

We can now prove that $(iii) \Rightarrow (ii)$. To do so, we begin by showing that any bounded closed subset $K \subset M$ is compact. Indeed, if K is bounded then $K \subset B_R(p)$ for some $R > 0$, where

$$B_R(p) = \{q \in M \mid d(p, q) < R\}.$$

As we have seen, p can be connected to any point in $B_R(p)$ by a geodesic of length smaller than R, and so $B_R(p) \subset \exp_p\left(\overline{B_R(0)}\right)$. Since $\exp_p : T_pM \to M$ is continuous and $\overline{B_R(0)}$ is compact, the set $\exp_p\left(\overline{B_R(0)}\right)$ is also compact. Therefore K is a closed subset of a compact set, hence compact. Now, if $\{p_n\}$ is a Cauchy sequence in M, then its closure is compact. Thus $\{p_n\}$ must have a convergent subsequence, and therefore must itself converge.

Finally, we show that $(ii) \Rightarrow (i)$. Let c be a geodesic defined for $t < t_0$, which we can assume without loss of generality to be **normalized**, that is, $\|\dot{c}(t)\| = 1$. Let $\{t_n\}$ be an increasing sequence of real numbers converging to t_0. Since $d(c(t_m), c(t_n)) \leq |t_m - t_n|$, we see that $\{c(t_n)\}$ is a Cauchy sequence. As we are assuming M to be complete, we conclude that $c(t_n) \to p \in M$, and it is easily seen that $c(t) \to p$ as $t \to t_0$. Let $B_\varepsilon(p)$ be a normal ball centered at p. Then c can be extended past t_0 in this normal ball. □

Corollary 5.6. *If M is compact then M is geodesically complete.*

Proof. Any compact metric space is complete. □

Corollary 5.7. *If M is a closed connected submanifold of a complete connected Riemannian manifold with the induced metric then M is complete.*

Proof. Let M be a closed connected submanifold of a complete connected Riemannian manifold N. Let d be the distance determined by the metric on N, and let d^* be the distance determined by the induced metric on M. Then $d \leq d^*$. Let $\{p_n\}$ be a Cauchy sequence on (M, d^*). Then $\{p_n\}$ is a Cauchy sequence on (N, d), and consequently converges in N to a point $p \in M$ (as N is complete and M is closed). Since the topology of M is induced by the topology of N, we conclude that $p_n \to p$ on M. □

EXERCISES 5.8.

(1) Prove Proposition 5.4.

(2) Consider $\mathbb{R}^2 \setminus \{(x, 0) \mid -3 \leq x \leq 3\}$ with the Euclidean metric. Determine $B_7(0, 4)$.

(3) (a) Prove that a connected Riemannian manifold is complete if and only if the compact sets are the closed bounded sets.

(b) Give an example of a connected Riemannian manifold containing a noncompact closed bounded set.

(c) A Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ is said to be **homogeneous** if given any two points $p, q \in M$ there exists an isometry $f : M \to M$ such that $f(p) = q$. Show that any homogenous Riemannian manifold is complete.

6. Notes on Chapter 3

6.1. Section 5. In this Section we use several definitions and results about metric spaces, which we now discuss. A **metric space** is a pair (M, d), where M is a set and $d : M \times M \to [0, +\infty)$ is a map satisfying the properties enumerated in Proposition 5.4. The set

$$B_\varepsilon(p) = \{q \in M \mid d(p, q) < \varepsilon\}$$

is called the **open ball** with center p and radius ε. The family of all such balls is a basis for a Hausdorff topology on M, called the **metric topology**. Notice that in this topology $p_n \to p$ if and only if $d(p_n, p) \to 0$. Although a metric space (M, d) is not necessarily second countable, it is still true that $F \subset M$ is closed if and only if every convergent sequence in F has limit in F, and $K \subset M$ is compact if and only if every sequence in K has a sublimit in K.

A sequence $\{p_n\}$ in M is said to be a **Cauchy sequence** if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(p_n, p_m) < \varepsilon$ for all $m, n > N$. It is easily seen that all convergent sequences are Cauchy sequences; the converse, however, is not necessarily true (but if a Cauchy sequence has a convergent subsequence then it must converge). A metric space is said to be **complete** if all its Cauchy sequences converge. A closed subset of a complete metric space is itself complete.

A set is said to be bounded if it is a subset of some ball. For instance, the set of all terms of a Cauchy sequence is bounded. It is easily shown that if $K \subset M$ is compact then K must be bounded and closed (but the converse is not necessarily true). A compact metric space is necessarily complete.

6.2. Bibliographical notes. The material in this chapter can be found in most books on Riemannian geometry (e.g. [**Boo03, dC93, GHL04**]). For more details on general affine connections, see [**KN96**]. Bi-invariant metrics on a Lie group are examples of symmetric spaces, whose beautiful theory is studied in [**Hel01**].

CHAPTER 4

Curvature

This chapter addresses the fundamental notion of **curvature** of a Riemannian manifold.

In Section 1 we define the **curvature operator** of a general affine connection, and, for Riemannian manifolds, the equivalent (more geometric) notion of **sectional curvature**.

Section 2 establishes **Cartan's structure equations**, a powerful computational method which employs differential forms to calculate the curvature. We use these equations in Section 3 to prove the **Gauss-Bonnet Theorem**, relating the curvature of a compact surface to its topology; we show in the Exercises how to use this theorem to interpret the curvature of a surface as a measure of the excess of the sum of the inner angles of a geodesic triangle over π.

We enumerate all complete Riemannian manifolds with **constant curvature** in Section 4. These provide important examples of curved geometries.

Finally, in Section 5 we study the relation between the curvature of a Riemannian manifold and the curvature of a submanifold (with the induced metric). This can again be used to give different geometric interpretations of the curvature. In particular, as shown in the Exercises, any sectional curvature is the curvature of a submanifold of dimension 2.

1. Curvature

As we saw in Exercise 3.3.4 of Chapter 3, no open set of the 2-sphere S^2 with the standard metric is isometric to an open set of the Euclidean plane. The geometric object that locally distinguishes these two Riemannian manifolds is the so-called **curvature operator**, which appears in many other situations (cf. Exercise 4.8.6 of Chapter 3):

Definition 1.1. *The* **curvature** *R of a connection ∇ is a correspondence that, to each pair of vector fields $X, Y \in \mathfrak{X}(M)$, associates a map $R(X,Y) : \mathfrak{X}(M) \to \mathfrak{X}(M)$ defined by*

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

Hence, it is a way of measuring the non-commutativity of the connection. We leave it as an exercise to show that this defines a $(3,1)$-tensor (called the **Riemann tensor**), meaning that

(i) $R(fX_1 + gX_2, Y)Z = fR(X_1, Y)Z + gR(X_2, Y)Z$,
(ii) $R(X, fY_1 + gY_2)Z = fR(X, Y_1)Z + gR(X, Y_2)Z$,
(iii) $R(X, Y)(fZ_1 + gZ_2) = fR(X, Y)Z_1 + gR(X, Y)Z_2$,

for all vector fields $X, X_1, X_2, Y, Y_1, Y_2, Z, Z_1, Z_2 \in \mathfrak{X}(M)$ and all smooth functions $f, g \in C^\infty(M, \mathbb{R})$. Locally, choosing a coordinate system $x : V \to \mathbb{R}^n$ on M, this tensor can be written as

$$R = \sum_{i,j,k,l=1}^{n} R_{ijk}{}^{l}\, dx^i \otimes dx^j \otimes dx^k \otimes \frac{\partial}{\partial x^l},$$

where each coefficient $R_{ijk}{}^{l}$ is the l-coordinate of the vector field $R(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})\frac{\partial}{\partial x^k}$, that is,

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^k} = \sum_{l=1}^{n} R_{ijk}{}^{l} \frac{\partial}{\partial x^l}.$$

Using $[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] = 0$, we have

$$\begin{aligned}
R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^k} &= \nabla_{\frac{\partial}{\partial x^i}} \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k} - \nabla_{\frac{\partial}{\partial x^j}} \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^k} \\
&= \nabla_{\frac{\partial}{\partial x^i}} \left(\sum_{m=1}^{n} \Gamma_{jk}^{m} \frac{\partial}{\partial x^m} \right) - \nabla_{\frac{\partial}{\partial x^j}} \left(\sum_{m=1}^{n} \Gamma_{ik}^{m} \frac{\partial}{\partial x^m} \right) \\
&= \sum_{m=1}^{n} \left(\frac{\partial}{\partial x^i} \cdot \Gamma_{jk}^{m} - \frac{\partial}{\partial x^j} \cdot \Gamma_{ik}^{m} \right) \frac{\partial}{\partial x^m} + \sum_{l,m=1}^{n} (\Gamma_{jk}^{m}\Gamma_{im}^{l} - \Gamma_{ik}^{m}\Gamma_{jm}^{l}) \frac{\partial}{\partial x^l} \\
&= \sum_{l=1}^{n} \left(\frac{\partial \Gamma_{jk}^{l}}{\partial x^i} - \frac{\partial \Gamma_{ik}^{l}}{\partial x^j} + \sum_{m=1}^{n} \Gamma_{jk}^{m}\Gamma_{im}^{l} - \sum_{m=1}^{n} \Gamma_{ik}^{m}\Gamma_{jm}^{l} \right) \frac{\partial}{\partial x^l},
\end{aligned}$$

and so

$$R_{ijk}{}^{l} = \frac{\partial \Gamma_{jk}^{l}}{\partial x^i} - \frac{\partial \Gamma_{ik}^{l}}{\partial x^j} + \sum_{m=1}^{n} \Gamma_{jk}^{m}\Gamma_{im}^{l} - \sum_{m=1}^{n} \Gamma_{ik}^{m}\Gamma_{jm}^{l}.$$

Example 1.2. Consider $M = \mathbb{R}^n$ with the Euclidean metric and the corresponding Levi-Civita connection (that is, with Christoffel symbols $\Gamma_{ij}^{k} \equiv 0$). Then $R_{ijk}{}^{l} = 0$, and the curvature R is zero. Thus, we interpret the curvature as a measure of how much a connection on a given manifold differs from the Levi-Civita connection of Euclidean space.

When the connection is symmetric (as in the case of the Levi-Civita connection), the tensor R satisfies the following property, known as the **Bianchi Identity**:

Proposition 1.3. *(*Bianchi Identity*) If M is a manifold with a symmetric connection then the associated curvature satisfies*

$$R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0.$$

Proof. This property is a direct consequence of the Jacobi identity of vector fields. Indeed,

$$\begin{aligned} R(X,Y)Z + R(Y,Z)X + R(Z,X)Y &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \\ &+\nabla_Y \nabla_Z X - \nabla_Z \nabla_Y X - \nabla_{[Y,Z]} X + \nabla_Z \nabla_X Y - \nabla_X \nabla_Z Y - \nabla_{[Z,X]} Y \\ &= \nabla_X (\nabla_Y Z - \nabla_Z Y) + \nabla_Y (\nabla_Z X - \nabla_X Z) + \nabla_Z (\nabla_X Y - \nabla_Y X) \\ &-\nabla_{[X,Y]} Z - \nabla_{[Y,Z]} X - \nabla_{[Z,X]} Y, \end{aligned}$$

and so, since the connection is symmetric, we have

$$\begin{aligned} &R(X,Y)Z + R(Y,Z)X + R(Z,X)Y \\ &= \nabla_X [Y,Z] + \nabla_Y [Z,X] + \nabla_Z [X,Y] - \nabla_{[Y,Z]} X - \nabla_{[Z,X]} Y - \nabla_{[X,Y]} Z \\ &= [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0. \end{aligned}$$

□

We will assume from this point on that (M,g) is a Riemannian manifold and ∇ its Levi-Civita connection. We can define a new covariant 4-tensor, known as the **curvature tensor**:

$$R(X,Y,Z,W) := g(R(X,Y)Z,W).$$

Again, choosing a coordinate system $x : V \to \mathbb{R}^n$ on M, we can write this tensor as

$$R(X,Y,Z,W) = \left(\sum_{i,j,k,l=1}^{n} R_{ijkl}\, dx^i \otimes dx^j \otimes dx^k \otimes dx^l \right) (X,Y,Z,W)$$

where

$$R_{ijkl} = g\left(R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right) \frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l} \right) = g\left(\sum_{m=1}^{n} R_{ijk}{}^m \frac{\partial}{\partial x^m}, \frac{\partial}{\partial x^l} \right) = \sum_{m=1}^{n} R_{ijk}{}^m g_{ml}.$$

This tensor satisfies the following symmetry properties:

Proposition 1.4. *If X,Y,Z,W are vector fields in M and ∇ is the Levi-Civita connection, then*

(i) $R(X,Y,Z,W) + R(Y,Z,X,W) + R(Z,X,Y,W) = 0$;
(ii) $R(X,Y,Z,W) = -R(Y,X,Z,W)$;
(iii) $R(X,Y,Z,W) = -R(X,Y,W,Z)$;
(iv) $R(X,Y,Z,W) = R(Z,W,X,Y)$.

Proof. Property (i) is an immediate consequence of the Bianchi identity, and property (ii) holds trivially.

Property (iii) is equivalent to showing that $R(X,Y,Z,Z) = 0$. Indeed, if (iii) holds then clearly $R(X,Y,Z,Z) = 0$. Conversely, if this is true, we have

$$R(X,Y,Z+W,Z+W) = 0 \Leftrightarrow R(X,Y,Z,W) + R(X,Y,W,Z) = 0.$$

Now, using the fact that the Levi-Civita connection is compatible with the metric, we have

$$X \cdot \langle \nabla_Y Z, Z \rangle = \langle \nabla_X \nabla_Y Z, Z \rangle + \langle \nabla_Y Z, \nabla_X Z \rangle$$

and

$$[X,Y] \cdot \langle Z, Z \rangle = 2 \langle \nabla_{[X,Y]} Z, Z \rangle.$$

Hence,

$$\begin{aligned} R(X,Y,Z,Z) &= \langle \nabla_X \nabla_Y Z, Z \rangle - \langle \nabla_Y \nabla_X Z, Z \rangle - \langle \nabla_{[X,Y]} Z, Z \rangle \\ &= X \cdot \langle \nabla_Y Z, Z \rangle - \langle \nabla_Y Z, \nabla_X Z \rangle - Y \cdot \langle \nabla_X Z, Z \rangle \\ &\quad + \langle \nabla_X Z, \nabla_Y Z \rangle - \frac{1}{2}[X,Y] \cdot \langle Z, Z \rangle \\ &= \frac{1}{2} X \cdot (Y \cdot \langle Z, Z \rangle) - \frac{1}{2} Y \cdot (X \cdot \langle Z, Z \rangle) - \frac{1}{2}[X,Y] \cdot \langle Z, Z \rangle \\ &= \frac{1}{2}[X,Y] \cdot \langle Z, Z \rangle - \frac{1}{2}[X,Y] \cdot \langle Z, Z \rangle = 0. \end{aligned}$$

To show (iv), we use (i) to get

$$\begin{array}{ccccccc} R(X,Y,Z,W) & + & R(Y,Z,X,W) & + & R(Z,X,Y,W) & = & 0 \\ R(Y,Z,W,X) & + & R(Z,W,Y,X) & + & R(W,Y,Z,X) & = & 0 \\ R(Z,W,X,Y) & + & R(W,X,Z,Y) & + & R(X,Z,W,Y) & = & 0 \\ R(W,X,Y,Z) & + & R(X,Y,W,Z) & + & R(Y,W,X,Z) & = & 0 \end{array}$$

and so, adding these and using (c), we have

$$R(Z,X,Y,W) + R(W,Y,Z,X) + R(X,Z,W,Y) + R(Y,W,X,Z) = 0.$$

Using (b) and (c), we obtain

$$2R(Z,X,Y,W) - 2R(Y,W,Z,X) = 0.$$

□

An equivalent way of encoding the information about the curvature of a Riemannian manifold is by considering the following definition:

Definition 1.5. *Let* Π *be a* 2*-dimensional subspace of* T_pM *and let* X_p, Y_p *be two linearly independent elements of* Π*. Then, the* **sectional curvature** *of* Π *is defined as*

$$K(\Pi) := -\frac{R(X_p, Y_p, X_p, Y_p)}{\|X_p\|^2 \|Y_p\|^2 - \langle X_p, Y_p \rangle^2}.$$

Note that $\|X_p\|^2 \|Y_p\|^2 - \langle X_p, Y_p \rangle^2$ is the square of the area of the parallelogram in T_pM spanned by X_p, Y_p, and so the above definition of sectional curvature does not depend on the choice of the linearly independent vectors X_p, Y_p. Indeed, when we change of basis on Π, both $R(X_p, Y_p, X_p, Y_p)$ and $\|X_p\|^2 \|Y_p\|^2 - \langle X_p, Y_p \rangle^2$ change by the square of the determinant of the change of basis matrix (cf. Exercise 1.11.2.). We will now see that knowing the sectional curvature of every section of T_pM completely determines the curvature tensor on this space.

PROPOSITION 1.6. *The Riemannian curvature tensor at p is uniquely determined by the values of the sectional curvatures of sections (that is, 2-dimensional subspaces) of T_pM.*

PROOF. Let us consider two covariant 4-tensors R_1, R_2 on T_pM satisfying the symmetry properties of Proposition 1.4. Then the tensor $T := R_1 - R_2$ also satisfies these symmetry properties. We will see that, if the values $R_1(X_p, Y_p, X_p, Y_p)$ and $R_2(X_p, Y_p, X_p, Y_p)$ agree for every $X_p, Y_p \in T_pM$ (that is, if $T(X_p, Y_p, X_p, Y_p) = 0$ for every $X_p, Y_p \in T_pM$), then $R_1 = R_2$ (that is, $T \equiv 0$). Indeed, for vectors $X_p, Y_p, Z_p \in T_pM$,

$$\begin{aligned} 0 &= T(X_p + Z_p, Y_p, X_p + Z_p, Y_p) = T(X_p, Y_p, Z_p, Y_p) + T(Z_p, Y_p, X_p, Y_p) \\ &= 2T(X_p, Y_p, Z_p, Y_p). \end{aligned}$$

Then $T(X_p, Y_p, Z_p, Y_p) = 0$ for all $X_p, Y_p, Z_p \in T_pM$, and so

$$\begin{aligned} 0 &= T(X_p, Y_p + W_p, Z_p, Y_p + W_p) = T(X_p, Y_p, Z_p, W_p) + T(X_p, W_p, Z_p, Y_p) \\ &= T(Z_p, W_p, X_p, Y_p) - T(W_p, X_p, Z_p, Y_p), \end{aligned}$$

that is, $T(Z_p, W_p, X_p, Y_p) = T(W_p, X_p, Z_p, Y_p)$. Hence T is invariant by cyclic permutations of the first three elements and so, by the Bianchi Identity, we have $3T(X_p, Y_p, Z_p, W_p) = 0$. □

A manifold is called **isotropic at a point** $p \in M$ if its sectional curvature is constant K_p for every section $\Pi \subset T_pM$. Moreover, it is called **isotropic** if it is isotropic at all points. Note that every 2-dimensional manifold is trivially isotropic. Its sectional curvature $K(p) := K_p$ is called the **Gauss curvature**. We will see later on other equivalent definitions of this curvature (cf. Exercise 2.8.9, Exercise 3.6.7 and Section 5). We will also see that the sectional curvature is actually the Gaussian curvature of special 2-dimensional submanifolds, formed by geodesics tangent to the sections (cf. Exercise 5.7.5).

PROPOSITION 1.7. *If M is isotropic at p and $x : V \to \mathbb{R}^n$ is a coordinate system around p, then the coefficients of the Riemannian curvature tensor at p are given by*

$$R_{ijkl}(p) = -K_p(g_{ik}\, g_{jl} - g_{il}\, g_{jk}).$$

PROOF. We first define a covariant 4-tensor A on T_pM as

$$A := \sum_{i,j,k,l=1}^{n} -K_p(\, g_{ik}\, g_{jl} - g_{il}\, g_{jk})\, dx^i \otimes dx^j \otimes dx^k \otimes dx^l.$$

We leave it as an exercise to check that A satisfies the symmetry properties of Proposition 1.4. Moreover,

$$\begin{aligned} A(X_p, Y_p, X_p, Y_p) &= \sum_{i,j,k,l=1}^{n} -K_p(\, g_{ik}\, g_{jl} - g_{il}\, g_{jk})\, X_p^i\, Y_p^j\, X_p^k\, Y_p^l \\ &= -K_p\left(\langle X_p, X_p\rangle\langle Y_p, Y_p\rangle - \langle X_p, Y_p\rangle^2\right) \\ &= R(X_p, Y_p, X_p, Y_p), \end{aligned}$$

and so, from Proposition 1.6, we conclude that $A = R$. $\square$

Definition 1.8. *A Riemannian manifold is called a manifold of* **constant curvature** *if it is isotropic and K_p is the same at all points of M.*

Example 1.9. The Euclidean space is a manifold of constant curvature $K_p \equiv 0$.

Another geometric object, very important in General Relativity, is defined as follows:

Definition 1.10. *The* **Ricci curvature** *tensor is the covariant 2-tensor locally defined as*

$$Ric(X,Y) := \sum_{k=1}^{n} dx^k \left(R\left(\frac{\partial}{\partial x^k}, X\right) Y \right).$$

Note that the above definition is independent of the choice of coordinates. Indeed, we can see $Ric_p(X_p, Y_p)$ as the trace of the linear map from T_pM to T_pM given by $Z_p \mapsto R(Z_p, X_p)Y_p$, hence independent of the choice of basis. Moreover, this tensor is symmetric. In fact, choosing an orthonormal basis $\{E_1 \ldots, E_n\}$ of T_pM we have

$$\begin{aligned} Ric_p(X_p, Y_p) &= \sum_{k=1}^{n} R(E_k, X_p, Y_p, E_k) = \sum_{k=1}^{n} R(Y_p, E_k, E_k, X_p) \\ &= \sum_{k=1}^{n} R(E_k, Y_p, X_p, E_k) = Ric_p(Y_p, X_p). \end{aligned}$$

Locally, we can write

$$Ric = \sum_{i,j=1}^{n} R_{ij} dx^i \otimes dx^j$$

where the coefficients R_{ij} are given by

$$R_{ij} := Ric\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \sum_{k=1}^{n} dx^k \left(R\left(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^i}\right) \frac{\partial}{\partial x^j} \right) = \sum_{k=1}^{n} R_{kij}{}^k,$$

that is, $R_{ij} = \sum_{k=1}^{n} R_{kij}{}^k$.

Note that from a $(3,1)$-tensor we obtained a $(2,0)$-tensor. This is an example of a general procedure called **contraction**, where we obtain a $(k-1, m-1)$-tensor from a (k,m)-tensor. To do so, we first choose two indices, one covariant and other contravariant, and then set them equal and take summations, obtaining a $(k-1, m-1)$-tensor. On the example of the Ricci tensor, we took the $(3,1)$-tensor $\widetilde{R}$ defined by the curvature,

$$\widetilde{R}(X,Y,Z,\omega) = \omega(R(X,Y)Z),$$

chose the first covariant index and the first contravariant index, set them equal and summed over them:

$$Ric(X,Y) = \sum_{k=1}^{n} \widetilde{R}\left(\frac{\partial}{\partial x^k}, X, Y, dx^k\right).$$

Similarly, we can use contraction to obtain a function (0-tensor) from the Ricci tensor (a covariant 2-tensor). For that, we first need to define a new $(1,1)$-tensor field T using the metric,

$$T(X,\omega) := Ric(X,Y),$$

where Y is such that $\omega(Z) = \langle Y, Z\rangle$ for every vector field Z. Then, we set the covariant index equal to the contravariant one and add, obtaining a function $S : M \to \mathbb{R}$ called the **scalar curvature**. Locally, choosing a coordinate system $x : V \to \mathbb{R}^n$, we have

$$S(p) := \sum_{k=1}^{n} T\left(\frac{\partial}{\partial x^k}, dx^k\right) = \sum_{k=1}^{n} Ric\left(\frac{\partial}{\partial x^k}, Y_k\right),$$

where, for every vector field Z on V,

$$Z^k = dx^k(Z) = \langle Z, Y_k\rangle = \sum_{i,j=1}^{n} g_{ij} Z^i Y_k^j.$$

Therefore, we must have $Y_k^j = g^{jk}$ (where $(g^{ij}) = (g_{ij})^{-1}$), and hence $Y_k = \sum_{i=1}^{n} g^{ik}\frac{\partial}{\partial x^i}$. We conclude that the scalar curvature is given by

$$S(p) = \sum_{k=1}^{n} Ric\left(\frac{\partial}{\partial x^k}, \sum_{i=1}^{n} g^{ik}\frac{\partial}{\partial x^i}\right) = \sum_{i,k=1}^{n} R_{ki} g^{ik} = \sum_{i,k=1}^{n} g^{ik} R_{ik}.$$

(since Ric is symmetric).

EXERCISES 1.11.

(1) (a) Show that the curvature operator satisfies
  (i) $R(fX_1 + gX_2, Y)Z = fR(X_1,Y)Z + gR(X_2,Y)Z$;
  (ii) $R(X, fY_1 + gY_2)Z = fR(X,Y_1)Z + gR(X,Y_2)Z$;
  (iii) $R(X,Y)(fZ_1 + gZ_2) = fR(X,Y)Z_1 + gR(X,Y)Z_2$,

  for all vector fields $X, X_1, X_2, Y, Y_1, Y_2, Z, Z_1, Z_2 \in \mathfrak{X}(M)$ and smooth functions $f, g \in C^\infty(M,\mathbb{R})$.

(b) Show that $(R(X,Y)Z)_p \in T_pM$ depends only on X_p, Y_p, Z_p. Conclude that R defines a $(3,1)$-tensor. (**Hint:** Choose local coordinates around $p \in M$).

(c) Recall that if G is a Lie group endowed with a bi-invariant Riemannian metric, ∇ is the Levi-Civita connection and X, Y are two left-invariant vector fields then

$$\nabla_X Y = \frac{1}{2}[X,Y]$$

(cf. Exercise 4.8.3 in Chapter 3). Show that if Z is also left-invariant, then

$$R(X,Y)Z = \frac{1}{4}[Z,[X,Y]].$$

(2) Show that $\|X_p\|^2\|Y_p\|^2 - \langle X_p, Y_p\rangle^2$ gives us the square of the area of the parallelogram in T_pM spanned by X_p, Y_p. Conclude that the sectional curvature does not depend on the choice of the linearly independent vectors X_p, Y_p, that is, when we change of basis on Π, both $R(X_p, Y_p, X_p, Y_p)$ and $\|X_p\|^2\|Y_p\|^2 - \langle X_p, Y_p\rangle^2$ change by the square of the determinant of the change of basis matrix.
(3) Show that Ric is the only independent contraction of the curvature tensor: choosing any other two indices and contracting, one either gets 0 or $\pm Ric$.
(4) Let M be a 3-dimensional manifold. Show that the curvature tensor is entirely determined by the Ricci tensor.
(5) Let (M, g) be an n-dimensional isotropic Riemannian manifold with sectional curvature K. Show that $Ric = (n-1)Kg$ and $S = n(n-1)K$.
(6) Let g_1, g_2 be two Riemannian metrics on a manifold M such that $g_1 = \rho g_2$, for some constant $\rho > 0$. Show that:
   (a) the corresponding sectional curvatures K_1 and K_2 satisfy $K_1(\Pi) = \rho^{-1}K_2(\Pi)$ for any 2-dimensional section of a tangent space of M;
   (b) the corresponding Ricci curvature tensors satisfy $Ric_1 = Ric_2$;
   (c) the corresponding scalar curvatures satisfy $S_1 = \rho^{-1}S_2$.
(7) If ∇ is not the Levi-Civita connection can we still define the Ricci curvature tensor Ric? Is it necessarily symmetric?

2. Cartan's Structure Equations

In this section we will reformulate the properties of the Levi-Civita connection and of the Riemannian curvature tensor in terms of differential forms. For that we will take an open subset V of M where we have defined a **field of frames** $X_1, \dots X_n$, that is, a set of n vector fields that, at each point p of V, form a basis for T_pM (for example, we can take a coordinate neighborhood V and the vector fields $X_i = \frac{\partial}{\partial x^i}$; however, in general, the X_i's are not associated to a coordinate system). Then we consider a **field of dual co-frames**, that is, 1-forms $\omega^1, \dots, \omega^n$ on V such that $\omega^i(X_j) = \delta_{ij}$. Note that, at each point $p \in V$, $\omega^1_p, \dots, \omega^n_p$ is a basis for T^*_pM. From the properties of a connection, in order to define $\nabla_X Y$ we just have to establish the values of

$$\nabla_{X_i} X_j = \sum_{k=1}^{n} \Gamma^k_{ij} X_k,$$

where Γ_{ij}^k is defined as the k^{th} component of the vector field $\nabla_{X_i} X_j$ on the basis $\{X_i\}_{i=1}^n$. Note that if the X_i's are not associated to a coordinate system then the Γ_{ij}^k's cannot be computed using formula (10), and, in general, they are not even symmetric in the indices i, j. Given the values of the Γ_{ij}^k's on V, we can define 1-forms ω_j^k $(j, k = 1, \ldots, n)$ in the following way:

$$\omega_j^k := \sum_{i=1}^n \Gamma_{ij}^k \omega^i. \tag{14}$$

Conversely, given these forms, we can obtain the values of Γ_{ij}^k through

$$\Gamma_{ij}^k = \omega_j^k(X_i).$$

The connection is then completely determined from these forms: given two vector fields $X = \sum_{i=1}^n a^i X_i$ and $Y = \sum_{i=1}^n b^i X_i$, we have

$$\begin{aligned} \nabla_X X_j &= \nabla_{\sum_{i=1}^n a^i X_i} X_j = \sum_{i=1}^n a^i \nabla_{X_i} X_j = \sum_{i,k=1}^n a^i\, \Gamma_{ij}^k\, X_k \\ &= \sum_{i,k=1}^n a^i\, \omega_j^k(X_i)\, X_k = \sum_{k=1}^n \omega_j^k(X)\, X_k \end{aligned} \tag{15}$$

and hence

$$\begin{aligned} \nabla_X Y &= \nabla_X \left(\sum_{i=1}^n b^i X_i \right) = \sum_{i=1}^n \big((X \cdot b^i) X_i + b^i \nabla_X X_i \big) \\ &= \sum_{j=1}^n \left(X \cdot b^j + \sum_{i=1}^n b^i \omega_i^j(X) \right) X_j. \end{aligned} \tag{16}$$

Note that the values of the forms ω_j^k at X are the components of $\nabla_X X_j$ relative to the field of frames, that is,

$$\omega_j^i(X) = \omega^i \left(\nabla_X X_j \right). \tag{17}$$

The ω_j^k's are called the **connection forms**. For the Levi-Civita connection, these forms cannot be arbitrary. Indeed, they have to satisfy some equations corresponding to the properties of symmetry and compatibility with the metric.

Theorem 2.1. (Cartan) *Let V be an open subset of a Riemannian manifold M on which we have defined a field of frames $X_1, \ldots, X_n$. Let $\omega^1, \ldots, \omega^n$ be the corresponding field of co-frames. Then the connection forms of the Levi-Civita connection are the unique solution of the equations*

(i) $d\omega^i = \sum_{j=1}^n \omega^j \wedge \omega_j^i$,

(ii) $dg_{ij} = \sum_{k=1}^n (g_{kj}\, \omega_i^k + g_{ki}\, \omega_j^k)$,

where $g_{ij} = \langle X_i, X_j \rangle$.

PROOF. We begin by showing that the Levi-Civita connection forms, defined by (14), satisfy (i) and (ii). For this, we will use the following property of 1-forms (cf. Exercise 3.8.2 of Chapter 2):

$$d\omega(X,Y) = X \cdot (\omega(Y)) - Y \cdot (\omega(X)) - \omega([X,Y]).$$

We have

$$\nabla_Y X = \nabla_Y \left(\sum_{j=1}^n \omega^j(X) X_j \right) = \sum_{j=1}^n \left(Y \cdot \omega^j(X)\, X_j + \omega^j(X)\, \nabla_Y X_j \right),$$

which implies

$$\omega^i(\nabla_Y X) = Y \cdot \omega^i(X) + \sum_{j=1}^n \omega^j(X) \omega^i(\nabla_Y X_j). \tag{18}$$

Using (17) and (18), we have

$$\begin{aligned}
\left(\sum_{j=1}^n \omega^j \wedge \omega^i_j \right)(X,Y) &= \sum_{j=1}^n \left(\omega^j(X)\, \omega^i_j(Y) - \omega^j(Y)\, \omega^i_j(X) \right) \\
&= \sum_{j=1}^n \left(\omega^j(X)\, \omega^i(\nabla_Y X_j) - \omega^j(Y)\, \omega^i(\nabla_X X_j) \right) \\
&= \omega^i(\nabla_Y X) - Y \cdot \omega^i(X) - \omega^i(\nabla_X Y) + X \cdot \omega^i(Y),
\end{aligned}$$

and so

$$\begin{aligned}
&\left(d\omega^i - \sum_{j=1}^n \omega^j \wedge \omega^i_j \right)(X,Y) = \\
&= X \cdot \omega^i(Y) - Y \cdot \omega^i(X) - \omega^i\left([X,Y]\right) - \sum_{j=1}^n \omega^j \wedge \omega^i_j(X,Y) \\
&= \omega^i\left(\nabla_X Y - \nabla_Y X - [X,Y] \right) = 0.
\end{aligned}$$

Note that equation (i) is equivalent to symmetry of the connection. To show that (ii) holds, we notice that

$$dg_{ij}(Y) = Y \cdot \langle X_i, X_j \rangle,$$

and, on the other hand,

$$\begin{aligned}
\left(\sum_{k=1}^n g_{kj}\, \omega^k_i + g_{ki}\, \omega^k_j \right)(Y) &= \sum_{k=1}^n g_{kj}\, \omega^k_i(Y) + g_{ki}\, \omega^k_j(Y) \\
&= \left\langle \sum_{k=1}^n \omega^k_i(Y)\, X_k, X_j \right\rangle + \left\langle \sum_{k=1}^n \omega^k_j(Y)\, X_k, X_i \right\rangle \\
&= \langle \nabla_Y X_i, X_j \rangle + \langle \nabla_Y X_j, X_i \rangle.
\end{aligned}$$

Hence, equation (ii) is equivalent to

$$Y \cdot \langle X_i, X_j \rangle = \langle \nabla_Y X_i, X_j \rangle + \langle X_i, \nabla_Y X_j \rangle,$$

for every i, j, that is, it is equivalent to compatibility with the metric (cf. Exercise 2.8.1). We conclude that the Levi-Civita connection forms satisfy (i) and (ii).

To prove unicity, we take 1-forms ω_i^j $(i, j = 1, \ldots, n)$ satisfying (i) and (ii). Using (15) and (16), we can define a connection, which is necessarily symmetric and compatible with the metric. By uniqueness of the Levi-Civita connection, we have uniqueness of the set of forms ω_i^j satisfying (i) and (ii) (note that each connection determines a unique set of n^2 connection forms and vice-versa). $\square$

REMARK 2.2. If on an open set we have a field of frames, we can perform Gram-Schmidt orthogonalization and obtain a smooth field of orthonormal frames $\{E_1, \ldots, E_n\}$ (the norm function is smooth on $T_pM \setminus \{0\}$). Then, as $g_{ij} = \langle E_i, E_j \rangle = \delta_{ij}$, equations (i) and (ii) above become

(i) $d\omega^i = \sum_{j=1}^n \omega^j \wedge \omega_j^i$,

(ii) $\omega_i^j + \omega_j^i = 0$.

In addition to connection forms, we can also define curvature forms. Again we consider an open subset V of M where we have a field of frames $\{X_1, \ldots, X_n\}$ (hence a corresponding field of dual coframes $\omega^1, \ldots, \omega^n$). We then define 2-forms Ω_k^l $(k, l = 1, \ldots, n)$ by

$$\Omega_k^l(X, Y) := \omega^l(R(X, Y)X_k),$$

for all vector fields X, Y in V (i.e., $R(X, Y)X_k = \sum_{l=1}^n \Omega_k^l(X, Y)X_l$). Using the basis $\{\omega^i \wedge \omega^j\}_{i<j}$ for 2-forms, we have

$$\begin{aligned} \Omega_k^l &= \sum_{i<j} \Omega_k^l(X_i, X_j)\, \omega^i \wedge \omega^j = \sum_{i<j} \omega^l(R(X_i, X_j)X_k)\, \omega^i \wedge \omega^j \\ &= \sum_{i<j} R_{ijk}{}^l\, \omega^i \wedge \omega^j = \frac{1}{2} \sum_{i,j=1}^n R_{ijk}{}^l\, \omega^i \wedge \omega^j, \end{aligned}$$

where $R_{ijk}{}^l$ are the coefficients of the curvature relative to these frames:

$$R(X_i, X_j)X_k = \sum_{l=1}^n R_{ijk}{}^l X_l.$$

These forms satisfy the following equation:

PROPOSITION 2.3. *In the above notation,*

(iii) $\Omega_i^j = d\omega_i^j - \sum_{k=1}^n \omega_i^k \wedge \omega_k^j$,

for every $i, j = 1, \ldots, n$.

PROOF. We will show that

$$R(X,Y)X_i = \sum_{j=1}^n \Omega_i^j(X,Y)X_j = \sum_{j=1}^n \left(\left(d\omega_i^j - \sum_{k=1}^n \omega_i^k \wedge \omega_k^j \right)(X,Y) \right) X_j.$$

Indeed,

$$\begin{aligned}
R(X,Y)X_i &= \nabla_X \nabla_Y X_i - \nabla_Y \nabla_X X_i - \nabla_{[X,Y]} X_i = \\
&= \nabla_X \left(\sum_{k=1}^n \omega_i^k(Y)X_k \right) - \nabla_Y \left(\sum_{k=1}^n \omega_i^k(X)X_k \right) - \sum_{k=1}^n \omega_i^k([X,Y])X_k \\
&= \sum_{k=1}^n \left(X \cdot \omega_i^k(Y) - Y \cdot \omega_i^k(X) - \omega_i^k([X,Y]) \right) X_k + \\
&\quad + \sum_{k=1}^n \omega_i^k(Y)\nabla_X X_k - \sum_{k=1}^n \omega_i^k(X)\nabla_Y X_k \\
&= \sum_{k=1}^n d\omega_i^k(X,Y)X_k + \sum_{k,j=1}^n \left(\omega_i^k(Y)\,\omega_k^j(X)\,X_j - \omega_i^k(X)\,\omega_k^j(Y)\,X_j \right) \\
&= \sum_{j=1}^n \left(d\omega_i^j(X,Y) - \sum_{k=1}^n (\omega_i^k \wedge \omega_k^j)(X,Y) \right) X_j.
\end{aligned}$$

$\square$

Equations (i), (ii) and (iii) are known as **Cartan's structure equations**. We list these equations below, as well as the main definitions:

(i) $d\omega^i = \sum_{j=1}^n \omega^j \wedge \omega_j^i$,
(ii) $dg_{ij} = \sum_{k=1}^n (g_{kj}\,\omega_i^k + g_{ki}\,\omega_j^k)$,
(iii) $d\omega_i^j = \Omega_i^j + \sum_{k=1}^n \omega_i^k \wedge \omega_k^j$,

where $\omega^i(X_j) = \delta_{ij}$, $\omega_j^k = \sum_{i=1}^n \Gamma_{ij}^k \omega^i$ and $\Omega_i^j = \sum_{k<l} {R_{kli}}^j\, \omega^k \wedge \omega^l$.

REMARK 2.4. If we consider an orthonormal field of frames $\{E_1, \ldots, E_n\}$, the above equations become:

(i) $d\omega^i = \sum_{j=1}^n \omega^j \wedge \omega_j^i$,
(ii) $\omega_i^j + \omega_j^i = 0$,
(iii) $d\omega_i^j = \Omega_i^j + \sum_{k=1}^n \omega_i^k \wedge \omega_k^j$ (and so $\Omega_i^j + \Omega_j^i = 0$).

EXAMPLE 2.5. For an orthonormal field of frames in $\mathbb{R}^n$ with the Euclidean metric, the curvature forms must vanish (as $R = 0$), and we obtain the following structure equations:

(i) $d\omega^i = \sum_{j=1}^n \omega^j \wedge \omega_j^i$,
(ii) $\omega_i^j + \omega_j^i = 0$,
(iii) $d\omega_i^j = \sum_{k=1}^n \omega_i^k \wedge \omega_k^j$.

To finish this section, we will consider in detail the special case of a 2-dimensional Riemannian manifold. In this case, the structure equations for an orthonormal field of frames are particularly simple: equation (ii) implies that there is only one independent connection form ($\omega_1^1 = \omega_2^2 = 0$ and $\omega_2^1 = -\omega_1^2$), which can be computed from equation (i):

$$\begin{aligned} d\omega^1 &= -\omega^2 \wedge \omega_1^2; \\ d\omega^2 &= \omega^1 \wedge \omega_1^2. \end{aligned}$$

Equation (iii) then yields that there is only one independent curvature form $\Omega_1^2 = d\omega_1^2$. This form is closely related to the Gauss curvature of the manifold:

PROPOSITION 2.6. *If M is a 2-dimensional manifold, then for an orthonormal frame we have $\Omega_1^2 = -K\omega^1 \wedge \omega^2$, where $K = K(p)$ is the Gauss curvature of M (that is, its sectional curvature).*

PROOF. Let p be a point in M and let us choose an open set containing p where we have defined an orthonormal field of frames $\{E_1, E_2\}$. Then

$$K = -R(E_1, E_2, E_1, E_2) = -R_{1212},$$

and consequently

$$\begin{aligned} \Omega_1^2 &= \Omega_1^2(E_1, E_2)\, \omega^1 \wedge \omega^2 = \omega^2(R(E_1, E_2)E_1)\, \omega^1 \wedge \omega^2 \\ &= \langle R(E_1, E_2)E_1, E_2\rangle\, \omega^1 \wedge \omega^2 = R_{1212}\, \omega^1 \wedge \omega^2 = -K\, \omega^1 \wedge \omega^2. \end{aligned}$$

□

Note that K does not depend on the choice of the field of frames, since it is a sectional curvature (cf. Definition 1.5). However, the connection forms do: Let $\{E_1, E_2\}, \{F_1, F_2\}$ be two orthonormal fields of frames on an open subset V of M. Then

$$\begin{pmatrix} F_1 & F_2 \end{pmatrix} = \begin{pmatrix} E_1 & E_2 \end{pmatrix} S$$

where $S : V \to O(2)$ has values in the orthogonal group of 2×2 matrices. Note that S has one of the following two forms

$$S = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \quad \text{or} \quad S = \begin{pmatrix} a & b \\ b & -a \end{pmatrix},$$

where $a, b : V \to \mathbb{R}$ are such that $a^2 + b^2 = 1$. The determinant of S is then ± 1 depending on whether the two frames have the same orientation or not. Then we have the following proposition:

PROPOSITION 2.7. *If $\{E_1, E_2\}$ and $\{F_1, F_2\}$ have the same orientation then, denoting by ω_1^2 and $\overline{\omega}_1^2$ the corresponding connection forms, we have $\overline{\omega}_1^2 - \omega_1^2 = \sigma$, where $\sigma = a\, db - b\, da$.*

Proof. Denoting by $\{\omega^1, \omega^2\}$ and $\{\overline{\omega}^1, \overline{\omega}^2\}$ the fields of dual co-frames corresponding to $\{E_1, E_2\}$ and $\{F_1, F_2\}$, we define the column vectors of 1-forms

$$\omega = \begin{pmatrix} \omega^1 \\ \omega^2 \end{pmatrix} \quad \text{and} \quad \overline{\omega} = \begin{pmatrix} \overline{\omega}^1 \\ \overline{\omega}^2 \end{pmatrix}$$

and the matrices of 1-forms

$$A = \begin{pmatrix} 0 & -\omega_1^2 \\ \omega_1^2 & 0 \end{pmatrix} \quad \text{and} \quad \overline{A} = \begin{pmatrix} 0 & -\overline{\omega}_1^2 \\ \overline{\omega}_1^2 & 0 \end{pmatrix}.$$

The relation between the frames can be written as

$$\overline{\omega} = S^{-1}\omega \Leftrightarrow \omega = S\,\overline{\omega}$$

and the Cartan structure equations as

$$d\omega = -A \wedge \omega \quad \text{and} \quad d\overline{\omega} = -\overline{A} \wedge \overline{\omega}.$$

Therefore

$$\begin{aligned} d\omega &= S\,d\overline{\omega} + dS \wedge \overline{\omega} = -S\overline{A} \wedge \overline{\omega} + dS \wedge S^{-1}\omega \\ &= -S\overline{A} \wedge S^{-1}\omega + dS \wedge S^{-1}\omega = -\left(S\overline{A}S^{-1} - dS\,S^{-1}\right) \wedge \omega, \end{aligned}$$

and unicity of solutions of the Cartan structure equations implies

$$A = S\overline{A}S^{-1} - dSS^{-1}.$$

Writing this out in full one obtains

$$\begin{pmatrix} 0 & -\omega_1^2 \\ \omega_1^2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\overline{\omega}_1^2 \\ \overline{\omega}_1^2 & 0 \end{pmatrix} - \begin{pmatrix} a\,da + b\,db & b\,da - a\,db \\ a\,db - b\,da & a\,da + b\,db \end{pmatrix},$$

and the result follows (we also obtain $a\,da + b\,db = 0$, which is clear from $\det A = a^2 + b^2 = 1$). $\square$

Let us now give a geometric interpretation of σ. Locally, we can define at each point $p \in M$ the angle $\theta(p)$ between $(E_1)_p$ and $(F_1)_p$. Then the change of basis matrix S has the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Hence,

$$\begin{aligned} \sigma &= a\,db - b\,da = \cos\theta d\,(\sin\theta) - \sin\theta d\,(\cos\theta) \\ &= \cos^2\theta d\theta + \sin^2\theta d\theta = d\theta. \end{aligned}$$

Therefore, integrating σ along a curve yields the angle by which F_1 rotates with respect to E_1 along the curve.

Notice that in particular σ is closed. This is also clear from

$$d\sigma = d\,\overline{\omega}_1^2 - d\omega_1^2 = -K\,\overline{\omega}^1 \wedge \overline{\omega}^2 + K\omega^1 \wedge \omega^2 = 0.$$

We can use the connection form ω_1^2 to define the **geodesic curvature** of a curve on an oriented Riemannian 2-manifold M. Let $c : I \to M$ be a smooth curve in M parametrized by its arclength s (hence $\|\dot{c}(s)\| = 1$). Let V be a neighborhood of a point $c(s)$ in this curve where we have a field of orthonormal frames $\{E_1, E_2\}$ satisfying $(E_1)_{c(s)} = \dot{c}(s)$. Note that it is always possible to consider such a field of frames: we start by extending the vector field $\dot{c}(s)$ to a unit vector field E_1 defined on a neighborhood of $c(s)$, and then consider a unit vector field E_2 orthogonal to the first, such that $\{E_1, E_2\}$ is positively oriented. Since

$$\nabla_{E_1} E_1 = \omega_1^1(E_1)E_1 + \omega_1^2(E_1)E_2 = \omega_1^2(E_1)E_2,$$

the **covariant acceleration** of c is

$$\nabla_{\dot{c}(s)}\dot{c}(s) = \nabla_{E_1(s)}E_1(s) = \omega_1^2(E_1(s))E_2(s).$$

We define the **geodesic curvature** of the curve c to be $k_g(s) := \omega_1^2(E_1(s))$ (in particular $|k_g(s)| = \|\nabla_{\dot{c}(s)}\dot{c}(s)\|$). It is a measure of how much the curve fails to be a geodesic at $c(s)$. In particular, c is a geodesic if and only if its geodesic curvature vanishes.

Exercises 2.8.

(1) Let $X_1, \ldots, X_n$ be a field of frames on an open set V of a Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$.Show that a connection ∇ on M is compatible with the metric on V if and only if

$$X_k \cdot \langle X_i, X_j \rangle = \langle \nabla_{X_k} X_i, X_j \rangle + \langle X_i, \nabla_{X_k} X_j \rangle$$

for all i, j, k.

(2) Show that Cartan's structure equations (i) and (iii) hold for any symmetric connection.

(3) Compute the Gauss curvature of:
 (a) the sphere S^2 with the standard metric;
 (b) the hyperbolic plane, i.e., the upper half-plane

$$H = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$$

 with the metric

$$g = \frac{1}{y^2}(dx \otimes dx + dy \otimes dy)$$

 (cf. Exercise 3.3.5 of Chapter 3).

(4) Determine all surfaces of revolution with constant Gauss curvature.

(5) Compute the Gauss curvature of the graph of a function $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ with the metric induced by the Euclidean metric of $\mathbb{R}^3$.

(6) Let M be the image of the parametrization $\varphi : (0, +\infty) \times \mathbb{R} \to \mathbb{R}^3$ given by

$$\varphi(u, v) = (u \cos v, u \sin v, v),$$

and let N be the image of the parametrization $\psi : (0, +\infty) \times \mathbb{R} \to \mathbb{R}^3$ given by

$$\psi(u, v) = (u \cos v, u \sin v, \log u).$$

Consider in both M and N the Riemannian metric induced by the Euclidean metric of $\mathbb{R}^3$. Show that the map $f : M \to N$ defined by

$$f(\varphi(u, v)) = \psi(u, v)$$

preserves the Gaussian curvature but is not a local isometry.

(7) Consider the metric

$$g = dr \otimes dr + f^2(r)d\theta \otimes d\theta$$

on $M = I \times S^1$, where r is a local coordinate on $I \subset \mathbb{R}$ and θ is the usual angular coordinate on S^1.

(a) Compute the Gaussian curvature of this metric.

(b) For which functions $f(r)$ is the Gaussian curvature constant?

(8) Consider the metric

$$g = A^2(r)dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2 \theta \, d\varphi \otimes d\varphi$$

on $M = I \times S^2$, where r is a local coordinate on $I \subset \mathbb{R}$ and (θ, φ) are spherical local coordinates on S^2.

(a) Compute the Ricci tensor and the scalar curvature of this metric.

(b) What happens when $A^2(r) = (1 - r^2)^{-1}$ (that is, when M is locally isometric to S^3)?

(c) And when $A^2(r) = (1 + r^2)^{-1}$ (that is, when M is locally isometric to the **hyperbolic 3-space**)?

(d) For which functions $A(r)$ is the scalar curvature constant?

(9) Let M be a Riemannian 2-manifold and let p be a point in M. Let D be a neighborhood of p in M homeomorphic to a disc, with a smooth boundary ∂D. Consider a point $q \in \partial D$ and a unit vector $X_q \in T_qM$. Let X be the parallel transport of X_q along ∂D. When X returns to q it makes an angle $\Delta\theta$ with the initial vector X_q. Parameterizing ∂D with arc length $(c : I \to \partial D)$ and using fields of orthonormal frames $\{E_1, E_2\}$ and $\{F_1, F_2\}$ positively oriented and such that $F_1 = X$, show that

$$\Delta\theta = \int_D K.$$

Conclude that the Gauss curvature of M at p satisfies

$$K(p) = \lim_{D \to p} \frac{\Delta\theta}{\operatorname{vol}(D)}.$$

(10) Compute the geodesic curvature of a circle on:

(a) $\mathbb{R}^2$ with the Euclidean metric and the usual orientation;

(b) S^2 with the usual metric and orientation;

(c) the hyperbolic plane with the usual orientation.

(11) Let c be a smooth curve on an oriented 2-manifold M as in the definition of geodesic curvature. Let X be a vector field parallel along c and let θ be the angle between X and $\dot{c}(s)$ along c in the given orientation. Show that the geodesic curvature of c, k_g, is equal to $\frac{d\theta}{ds}$. (**Hint:** Consider two fields of orthonormal frames $\{E_1, E_2\}$ and $\{F_1, F_2\}$ positively oriented and such that $F_1 = \frac{X}{\|X\|}$).

3. Gauss-Bonnet Theorem

We will now use Cartan's structure equations to prove the **Gauss-Bonnet Theorem**, relating the curvature of a compact surface to its topology. Let M be a compact, oriented, 2-dimensional manifold and X a vector field on M.

DEFINITION 3.1. *A point $p \in M$ is said to be a* **singular point** *of X if $X_p = 0$. A singular point is said to be an* **isolated singularity** *if there exists a neighborhood $V \subset M$ of p such that p is the only singular point of X in V.*

Since M is compact, if all the singularities of X are isolated then they are in finite number (as otherwise they would accumulate on a non-isolated singularity).

To each isolated singularity $p \in V$ of $X \in \mathfrak{X}(M)$ one can associate an integer number, called the **index** of X at p, as follows:

(i) fix a Riemannian metric in M;

(ii) choose a positively oriented orthonormal frame $\{F_1, F_2\}$, defined on $V \setminus \{p\}$, such that
$$F_1 = \frac{X}{\|X\|};$$
let $\{\overline{\omega}^1, \overline{\omega}^2\}$ be the dual co-frame and let $\overline{\omega}_1^2$ be the corresponding connection form;

(iii) possibly shrinking V, choose a positively oriented orthonormal frame $\{E_1, E_2\}$, defined on V, with dual co-frame $\{\omega^1, \omega^2\}$ and connection form ω_1^2;

(iv) take a neighborhood D of p in V, homeomorphic to a disc, with a smooth boundary ∂D, endowed with the induced orientation; we then define the index I_p of X at p through
$$2\pi I_p = \int_{\partial D} \sigma,$$
where $\sigma := \overline{\omega}_1^2 - \omega_1^2$ is the form defined in Section 2.

Recall that σ satisfies $\sigma = d\theta$, where θ is the angle between E_1 and F_1. Therefore I_p must be an integer. Intuitively, the index of a vector field X measures the number of times that X rotates as one goes around the singularity anticlockwise, counted positively if X itself rotates anticlockwise, and negatively otherwise.

EXAMPLE 3.2. In $M = \mathbb{R}^2$ the following vector fields have isolated singularities at the origin with the indicated indices (cf. Figure 1):

(1) $X_{(x,y)} = (x, y)$ has index 1;
(2) $Y_{(x,y)} = (-y, x)$ has index 1;
(3) $Z_{(x,y)} = (y, x)$ has index -1.
(4) $W_{(x,y)} = (x, -y)$ has index -1.

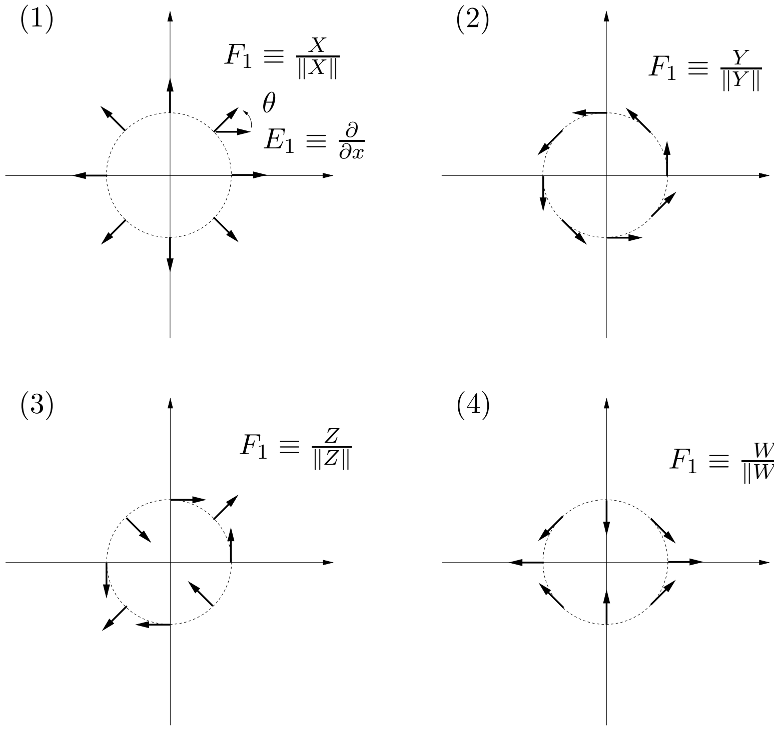

FIGURE 1. Computing the indices of the vector fields X, Y, Z and W.

We will now check that the index is well defined. We begin by observing that, since σ is closed, I_p does not depend on the choice of D. Indeed, the boundaries of any two such discs are necessarily homotopic (cf. Exercise 5.2.2 of Chapter 2). Next we prove that I_p does not depend on the choice of the frame $\{E_1, E_2\}$. More precisely, we have

$$I_p = \lim_{r\to 0} \frac{1}{2\pi} \int_{S_r(p)} \overline{\omega}_1^2,$$

where $S_r(p)$ is the normal sphere of radius r centered at p. Indeed, if $r_1 > r_2 > 0$ are radii of normal spheres, one has

$$\int_{S_{r_1}(p)} \overline{\omega}_1^2 - \int_{S_{r_2}(p)} \overline{\omega}_1^2 = \int_{\Delta_{12}} d\overline{\omega}_1^2 = -\int_{\Delta_{12}} K\overline{\omega}^1 \wedge \overline{\omega}^2 = -\int_{\Delta_{12}} K, \tag{19}$$

where $\Delta_{12} = B_{r_1}(p) \setminus B_{r_2}(p)$. Since K is continuous, we see that

$$\left(\int_{S_{r_1}(p)} \overline{\omega}_1^2 - \int_{S_{r_2}(p)} \overline{\omega}_1^2 \right) \longrightarrow 0$$

as $r_1 \to 0$. Therefore, if $\{r_n\}$ is a decreasing sequence of positive numbers converging to zero, the sequence

$$\left\{ \int_{S_{r_n}(p)} \overline{\omega}_1^2 \right\}$$

is a Cauchy sequence, and therefore converges. Thus the limit

$$\overline{I}_p = \lim_{r \to 0} \frac{1}{2\pi} \int_{S_r(p)} \overline{\omega}_1^2$$

exists. Making $r_2 \to 0$ on (19) one obtains

$$\int_{S_{r_1}(p)} \overline{\omega}_1^2 - 2\pi \overline{I}_p = -\int_{B_{r_1}(p)} K = -\int_{B_{r_1}(p)} K\omega^1 \wedge \omega^2 = \int_{B_{r_1}(p)} d\omega_1^2 = \int_{S_{r_1}(p)} \omega_1^2,$$

and hence

$$2\pi I_p = \int_{S_{r_1}(p)} \sigma = \int_{S_{r_1}(p)} \overline{\omega}_1^2 - \omega_1^2 = 2\pi \overline{I}_p.$$

Finally, we show that I_p does not depend on the choice of Riemannian metric. Indeed, if $\langle \cdot, \cdot \rangle_0, \langle \cdot, \cdot \rangle_1$ are two Riemannian metrics on M, it is easy to check that

$$\langle \cdot, \cdot \rangle_t := (1-t)\langle \cdot, \cdot \rangle_0 + t\langle \cdot, \cdot \rangle_1$$

is also a Riemannian metric on M, and that the index $I_p(t)$ computed using the metric $\langle \cdot, \cdot \rangle_t$ is a continuous function of t (cf. Exercise 3.6.1). Since $I_p(t)$ is an integer for all $t \in [0,1]$, we conclude that $I_p(0) = I_p(1)$.

Therefore I_p depends only on the vector field $X \in \mathfrak{X}(M)$. We are now ready to state the Gauss-Bonnet Theorem:

THEOREM 3.3. (Gauss-Bonnet) *Let M be a compact, oriented, 2-dimensional manifold and let X be a vector field in M with isolated singularities $p_1, \dots, p_k$. Then*

$$\int_M K = 2\pi \sum_{i=1}^k I_{p_i} \tag{20}$$

for any Riemannian metric on M, where K is the Gauss curvature.

PROOF. We consider the positively oriented orthonormal frame $\{F_1, F_2\}$, with

$$F_1 = \frac{X}{\|X\|},$$

defined on $M \setminus \cup_{i=1}^k \{p_i\}$, with dual co-frame $\{\overline{\omega}^1, \overline{\omega}^2\}$ and connection form $\overline{\omega}_1^2$. For $r > 0$ sufficiently small, we take $B_i = B_r(p_i)$ such that $B_i \cap B_j = \varnothing$

for $i \neq j$ and note that

$$\begin{aligned}\int_{M\setminus\cup_{i=1}^k B_i} K &= \int_{M\setminus\cup_{i=1}^k B_i} K\,\overline{\omega}^1 \wedge \overline{\omega}^2 = -\int_{M\setminus\cup_{i=1}^k B_i} d\overline{\omega}_1^2 \\ &= \int_{\cup_{i=1}^k \partial B_i} \overline{\omega}_1^2 = \sum_{i=1}^k \int_{\partial B_i} \overline{\omega}_1^2,\end{aligned}$$

where ∂B_i have the orientation induced by the orientation of B_i. Making $r \to 0$, one obtains

$$\int_M K = 2\pi \sum_{i=1}^k I_{p_i}.$$

□

Remark 3.4.

(1) Since the right-hand side of (20) does not depend on the metric, we conclude that $\int_M K$ is the same for **all** Riemannian metrics on M.

(2) Since the left-hand side of (20) does not depend on the vector field X, we conclude that $\chi(M) := \sum_{i=1}^k I_{p_i}$ is the same for all vector fields on M with isolated singularities. This is the so-called **Euler characteristic** of M.

(3) Recall that a **triangulation** of M is a decomposition of M in a finite number of triangles (i.e., images of Euclidean triangles by parametrizations) such that the intersection of any two triangles is either a common edge, a common vertex or empty (it is possible to prove that such a triangulation always exists). Given a triangulation, one can construct a vector field with the following properties (cf. Figure 2):
   (a) each vertex is a singularity, which is a sink;
   (b) each face contains exactly one singularity, which is a source;
   (c) each edge is formed by integral curves of the vector field and contains exactly one singularity.

   It is easy to see that all singularities are isolated, that the singularities at the vertices and faces have index 1 and that the singularities at the edges have index -1. Therefore,

$$\chi(M) = V - E + F,$$

   where V is the number of vertices, E is the number of edges and F is the number of faces on any triangulation. This is the definition we used in Exercise 1.8.5 of Chapter 1.

Example 3.5.

(1) Choosing the standard metric in S^2, we have

$$\chi(S^2) = \frac{1}{2\pi}\int_{S^2} 1 = \frac{1}{2\pi}\mathrm{vol}(S^2) = 2.$$

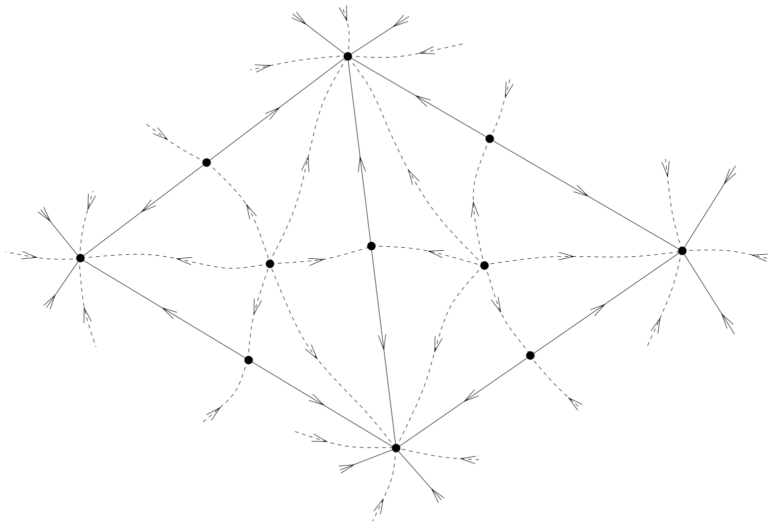

FIGURE 2. Vector field associated to a triangulation.

From this one can derive a number of conclusions:

(a) there is no zero curvature metric on S^2, for this would imply $\chi(S^2) = 0$.

(b) there is no vector field on S^2 without singularities, as this would also imply $\chi(S^2) = 0$.

(c) for any triangulation of S^2, one has $V - E + F = 2$. In particular, this proves Euler's formula for convex polyhedra with triangular faces, as these clearly yield triangulations of S^2.

(2) As we saw in Section 4, the torus T^2 has a zero curvature metric, and hence $\chi(T^2) = 0$. This can also be seen from the fact that there exist vector fields on T^2 without singularities.

EXERCISES 3.6.

(1) Show that if $\langle\cdot,\cdot\rangle_0, \langle\cdot,\cdot\rangle_1$ are two Riemannian metrics on M then

$$\langle\cdot,\cdot\rangle_t := (1-t)\langle\cdot,\cdot\rangle_0 + t\langle\cdot,\cdot\rangle_1$$

is also a Riemannian metric on M, and that the index $I_p(t)$ computed using the metric $\langle\cdot,\cdot\rangle_t$ is a continuous function of t.

(2) (*Gauss-Bonnet Theorem for non-orientable manifolds*) Let (M,g) be a compact, non-orientable, 2-dimensional Riemannian manifold and let $\pi : \overline{M} \to M$ be its orientable double covering (cf. Exercise 8.6.9 in Chapter 1). Show that:

(a) $\chi(\overline{M}) = 2\chi(M)$;

(b) $\overline{K} = \pi^*K$, where $\overline{K}$ is the Gauss curvature of the Riemannian metric $\overline{g} = \pi^*g$ on $\overline{M}$;

(c) $\chi(M) = \frac{1}{2}\int_{\overline{M}} \overline{K}$.

(**Remark:** Even though M is not orientable, we can still define the integral of a function f on M through $\int_M f = \frac{1}{2}\int_{\overline{M}} \pi^* f$; with this definition, the Gauss-Bonnet Theorem holds for non-orientable Riemannian 2-manifolds).

(3) Let M be a compact, oriented, 2-dimensional manifold with boundary and let X be a vector field in M **transverse** to ∂M (i.e., such that $X_p \notin T_p\partial M$ for all $p \in \partial M$), with isolated singularities $p_1, \ldots, p_k \in M \setminus \partial M$. Prove that

$$\int_M K + \int_{\partial M} k_g(s)ds = 2\pi \sum_{i=1}^{k} I_{p_i}$$

for any Riemannian metric on M, where K is the Gauss curvature of M, k_g is the geodesic curvature of ∂M and s is the arclength.

(4) Let (M, g) be a compact orientable 2-dimensional Riemannian manifold, with positive Gauss curvature. Show that any two non-self-intersecting closed geodesics must intersect each other.

(5) (*Hessian*) Let M be a differentiable manifold, $f : M \to \mathbb{R}$ a smooth function and $p \in M$ a critical point of f (i.e. $(df)_p = 0$). For $v, w \in T_pM$ we define the **Hessian** of f at p to be the map $(Hf)_p : T_pM \times T_pM \to \mathbb{R}$ given by

$$(Hf)_p(v, w) = \left.\frac{\partial^2}{\partial t \partial s}\right|_{s=t=0} f \circ \gamma\,(s, t),$$

where $\gamma : U \subset \mathbb{R}^2 \to M$ is such that $\gamma(0, 0) = p$, $\frac{\partial\gamma}{\partial s}(0, 0) = v$ and $\frac{\partial\gamma}{\partial t}(0, 0) = w$. Show that $(Hf)_p$

(a) is well-defined;

(b) is a symmetric 2-tensor (if $(Hf)_p$ is nondegenerate then p is called a **nondegenerate critical point**).

(6) (*Morse Theorem*) A smooth function $f : M \to \mathbb{R}$ is said to be a **Morse function** if all its critical points are nondegenerate. If M is compact then the number of critical points of any Morse function on M is finite. Prove that if M is a 2-dimensional compact manifold and $f : M \to \mathbb{R}$ is a Morse function with m maxima, s saddle points and n minima, then

$$\chi(M) = m - s + n.$$

(**Hint:** Choose a Riemannian metric on M and consider the vector field $X = \operatorname{grad} f$).

(7) Let (M, g) be a 2-dimensional Riemannian manifold and $\Delta \subset M$ a **geodesic triangle**, i.e., an open set homeomorphic to a disc whose boundary is contained in the union of the images of three geodesics. Let α, β, γ be the inner angles of Δ, i.e., the angles between the geodesics at the intersection points contained in $\partial\Delta$. Prove that for small enough Δ one has

$$\alpha + \beta + \gamma = \pi + \int_\Delta K,$$

where K is the Gauss curvature of M, using:

(a) the fact that $\int_\Delta K$ is the angle by which a vector parallel-transported once around $\partial\Delta$ rotates;

(b) the Gauss-Bonnet Theorem for manifolds with boundary.

(**Remark:** We can use this result to give another geometric interpretation of the Gauss curvature: $K(p) = \lim_{\Delta \to p} \frac{\alpha+\beta+\gamma-\pi}{\mathrm{vol}(\Delta)}$).

(8) Let (M, g) be a simply connected 2-dimensional Riemannian manifold with nonpositive Gauss curvature. Show that any two geodesics intersect at most in one point. (**Hint:** Note that if two geodesics intersect in more than one point then one would have a **geodesic biangle**, i.e., an open set homeomorphic to a disc whose boundary is contained in the union of the images of two geodesics.).

4. Manifolds of Constant Curvature

Recall that a manifold is said to have constant curvature if all sectional curvatures at all points have the same constant value K. There is an easy way to identify these manifolds using their curvature forms:

LEMMA 4.1. *If M is a manifold of constant curvature K, then, around each point $p \in M$, all curvature forms Ω_i^j satisfy*

$$\Omega_i^j = -K\omega^i \wedge \omega^j, \tag{21}$$

where $\{\omega^1, \ldots, \omega^n\}$ is any field of orthonormal co-frames defined on a neighborhood of p. Conversely, if on a neighborhood of each point of M there is a field of orthonormal frames $E_1, \ldots, E_n$ such that the corresponding field of co-frames $\{\omega^1, \ldots, \omega^n\}$ satisfies (21) for some constant K, then M has constant curvature K.

PROOF. If M has constant curvature K then

$$\begin{aligned}
\Omega_i^j &= \sum_{k<l} \Omega_i^j(E_k, E_l)\,\omega^k \wedge \omega^l = \sum_{k<l} \omega^j(R(E_k, E_l)E_i)\,\omega^k \wedge \omega^l \\
&= \sum_{k<l} \langle R(E_k, E_l)E_i, E_j\rangle \omega^k \wedge \omega^l = \sum_{k<l} R_{klij}\,\omega^k \wedge \omega^l \\
&= -\sum_{k<l} K(\delta_{ki}\delta_{lj} - \delta_{kj}\delta_{li})\,\omega^k \wedge \omega^l = -K\omega^i \wedge \omega^j.
\end{aligned}$$

Conversely, let us assume that there is a constant K such that on a neighborhood of each point $p \in M$ we have $\Omega_i^j = -K\omega^i \wedge \omega^j$. Then, for every section Π of the tangent space T_pM, the corresponding sectional curvature is given by

$$K(\Pi) = -R(X, Y, X, Y)$$

where X, Y are two linearly independent vectors spanning Π (which we assume to span a parallelogram of unit area). Using the field of orthonormal frames around p, we have $X=\sum_{i=1}^n X^iE_i$ and $Y=\sum_{i=1}^n Y^iE_i$ and so,

$$\begin{aligned}
K(\Pi) &= -\sum_{i,j,k,l=1}^n X^iY^jX^kY^lR(E_i,E_j,E_k,E_l)\\
&= -\sum_{i,j,k,l=1}^n X^iY^jX^kY^l\,\Omega_k^l(E_i,E_j)\\
&= K\sum_{i,j,k,l=1}^n X^iY^jX^kY^l\,\omega^k\wedge\omega^l(E_i,E_j)\\
&= K\sum_{i,j,k,l=1}^n X^iY^jX^kY^l\left(\omega^k(E_i)\omega^l(E_j)-\omega^k(E_j)\omega^l(E_i)\right)\\
&= K\sum_{i,j,k,l=1}^n X^iY^jX^kY^l(\delta_{ik}\delta_{jl}-\delta_{jk}\delta_{il})\\
&= K\left(\|X\|^2\,\|Y\|^2-\langle X,Y\rangle^2\right)=K.
\end{aligned}$$

□

Let us now see an example of how we can use this lemma:

EXAMPLE 4.2. Let a be a positive real number and let

$$H^n(a)=\{(x^1,\dots x^n)\in\mathbb{R}^n : x^n>0\}.$$

We will see that the Riemannian metric in $H^n(a)$ given by

$$g_{ij}(x)=\frac{a^2}{(x^n)^2}\,\delta_{ij},$$

has constant sectional curvature $K=-\frac{1}{a^2}$. Indeed, using the above lemma, we will show that on $H^n(a)$ there is a field of orthonormal frames $E_1\dots,E_n$ whose dual field of co-frames $\omega^1\dots\omega^n$ satisfies

$$\Omega_i^j=-K\omega^i\wedge\omega^j \tag{22}$$

for $K=-\frac{1}{a^2}$. For that, let us consider the natural coordinate system $x:H^n(a)\to\mathbb{R}^n$ and the corresponding field of coordinate frames $X_1,\dots,X_n$ with $X_i=\frac{\partial}{\partial x_i}$. Since

$$\langle X_i,X_j\rangle=\frac{a^2}{(x^n)^2}\,\delta_{ij},$$

we obtain a field of orthonormal frames $E_1,\dots,E_n$ with $E_i=\frac{x^n}{a}X_i$, and the corresponding dual field of co-frames $\omega^1,\dots\omega^n$ where $\omega^i=\frac{a}{x^n}dx^i$. Then

$$d\omega^i=\frac{a}{(x^n)^2}\,dx^i\wedge dx^n=\frac{1}{a}\omega^i\wedge\omega^n=\sum_{k=1}^n\omega^k\wedge\left(-\frac{1}{a}\delta_{kn}\omega^i\right),$$

and so, using the structure equations

$$d\omega^i = \sum_{k=1}^{n} \omega^k \wedge \omega_k^i$$
$$\omega_i^j + \omega_j^i = 0,$$

we can guess that the connection forms are given by $\omega_j^i = \frac{1}{a}(\delta_{in}\omega^j - \delta_{jn}\omega^i)$. We can easily verify that these forms satisfy the above structure equations since

$$\sum_{k=1}^{n} \omega^k \wedge \omega_k^i = \frac{1}{a}\sum_{k=1}^{n} \omega^k \wedge (\delta_{in}\omega^k - \delta_{kn}\omega^i) = \frac{1}{a}\omega^i \wedge \omega^n = d\omega^i$$

and

$$\omega_i^j = \frac{1}{a}(\delta_{jn}\omega^i - \delta_{in}\omega^j) = -\frac{1}{a}(\delta_{in}\omega^j - \delta_{jn}\omega^i) = -\omega_j^i.$$

Hence, by unicity of solution of these equations, we conclude that these forms are indeed given by $\omega_i^j = \frac{1}{a}(\delta_{jn}\omega^i - \delta_{in}\omega^j)$. With the connection forms it is now easy to compute the curvature forms Ω_i^j using the third structure equation

$$d\omega_i^j = \sum_{k=1}^{n} \omega_i^k \wedge \omega_k^j + \Omega_i^j.$$

In fact,

$$d\omega_i^j = d\left(\frac{1}{a}(\delta_{jn}\omega^i - \delta_{in}\omega^j)\right) = \frac{1}{a^2}\left(\delta_{jn}\omega^i \wedge \omega^n - \delta_{in}\omega^j \wedge \omega^n\right)$$

and

$$\begin{aligned}
\sum_{k=1}^{n} \omega_i^k \wedge \omega_k^j &= \frac{1}{a^2}\sum_{k=1}^{n}(\delta_{kn}\omega^i - \delta_{in}\omega^k) \wedge (\delta_{jn}\omega^k - \delta_{kn}\omega^j) \\
&= \frac{1}{a^2}\sum_{k=1}^{n}(\delta_{kn}\delta_{jn}\omega^i \wedge \omega^k - \delta_{kn}\omega^i \wedge \omega^j + \delta_{in}\delta_{kn}\omega^k \wedge \omega^j) \\
&= \frac{1}{a^2}\left(\delta_{jn}\omega^i \wedge \omega^n - \omega^i \wedge \omega^j + \delta_{in}\omega^n \wedge \omega^j\right),
\end{aligned}$$

and so,

$$\Omega_i^j = \frac{1}{a^2}\left(\delta_{jn}\omega^i \wedge \omega^n - \delta_{in}\omega^j \wedge \omega^n - \delta_{jn}\omega^i \wedge \omega^n + \omega^i \wedge \omega^j - \delta_{in}\omega^n \wedge \omega^j\right) = \frac{1}{a^2}\omega^i \wedge \omega^j.$$

We conclude that $K = -\frac{1}{a^2}$. Note that these spaces give us examples in any dimension of Riemannian manifolds with arbitrary constant negative curvature.

The Euclidean spaces $\mathbb{R}^n$ give us examples of Riemannian manifolds with constant curvature equal to zero. Moreover, we can easily see that the spheres $S^n(r) \subset \mathbb{R}^{n+1}$ of radius r have constant curvature equal to $\frac{1}{r^2}$ (cf. Exercise 5.7.2), and so we have examples in any dimension of spaces with arbitrary constant positive curvature. Note that all of the examples given so far in this section are simply connected and are geodesically complete. Indeed, the geodesics of the Euclidean space $\mathbb{R}^n$ traverse straight lines, $S^n(r)$ is compact and the geodesics of $H^n(a)$ traverse either half circles perpendicular to the plane $x^n = 0$ and centered on this plane, or vertical half lines starting at the plane $x^n = 0$.

Every simply connected geodesically complete manifold of constant curvature is isometric to one of these examples as it is stated in the following theorem (which we will not prove). In general, if the manifold is not simply connected (but still geodesically complete), it is isometric to the quotient of one of the above examples by a free and proper action of a discrete subgroup of the group of isometries (it can be proved that the group of isometries of a Riemannian manifold is always a Lie group).

THEOREM 4.3. (Killing-Hopf)

(1) *Let M be a simply connected Riemannian manifold geodesically complete. If M has constant curvature K then it is isometric to one of the following: $S^n\left(\frac{1}{\sqrt{K}}\right)$ if $K > 0$, $\mathbb{R}^n$ if $K = 0$, or $H^n\left(\frac{1}{\sqrt{-K}}\right)$ if $K < 0$.*
(2) *Let M be a geodesically complete manifold (not necessarily simply connected) with constant curvature K. Then M is isometric to a quotient $\widetilde{M}/\Gamma$, where $\widetilde{M}$ is one of the above simply connected spaces and Γ is a discrete subgroup of the group of isometries of $\widetilde{M}$ acting properly and freely on $\widetilde{M}$.*

EXAMPLE 4.4. Let $\widetilde{M} = \mathbb{R}^2$. Then the subgroup of isometries Γ cannot contain any rotation (since it acts freely). Hence it can only contain translations and gliding reflections (that is, reflections followed by a translation in the direction of the reflection axis). Moreover, it is easy to check that Γ has to be generated by at most two elements. Hence we obtain that:

(1) if Γ is generated by one translation, then the resulting surface will be a cylinder;
(2) if Γ is generated by two translations we obtain a torus;
(3) if Γ is generated by a gliding reflection we obtain a Möbius band;
(4) if Γ is generated by a translation and a gliding reflection we obtain a Klein bottle.

Note that if Γ is generated by two gliding reflections then it can also be generated by a translation and a gliding reflection (cf. Exercise 4.7.4). Hence, these are all the possible examples of geodesically complete Euclidean surfaces (2-dimensional manifolds of constant zero curvature).

EXAMPLE 4.5. The group of orientation-preserving isometries of the hyperbolic plane H^2 is $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm Id\}$, acting on H^2 through

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az+b}{cz+d},$$

where we make the identification $\mathbb{R}^2 \cong \mathbb{C}$ (cf. Exercise 4.7.5). To find orientable hyperbolic surfaces, that is, surfaces with constant curvature $K = -1$, we have to find discrete subgroups Γ of $PSL(2,\mathbb{R})$ acting properly and freely on H^2. Here there are many more possibilities. As an example, we can consider the group $\Gamma = \langle t_{2\pi} \rangle$ generated by the translation $t_{2\pi}(z) = z + 2\pi$. The resulting surface is known as **pseudosphere** and is homeomorphic to a cylinder (cf. Figure 3). However, the width of the end where $y \to +\infty$ converges to zero, while the width of the end where $y \to 0$ converges to $+\infty$. Its height towards both ends is infinite. Note that this surface has geodesics which transversely autointersect a finite number of times (cf. Figure 4).

Other examples can be obtained by considering hyperbolic polygons (bounded by geodesics) and identifying their sides through isometries. For instance, the surface in Figure 5-(b) is obtained by identifying the sides of the polygon in Figure 5-(a) through the isometries $g(z) = z+2$ and $h(z) = \frac{z}{2z+1}$. Choosing other polygons it is possible to obtain **compact** hyperbolic surfaces. In fact, there exist compact hyperbolic surfaces homeomorphic to any topological 2-manifold with negative Euler characteristic (the Gauss-Bonnet Theorem does not allow non-negative Euler characteristics).

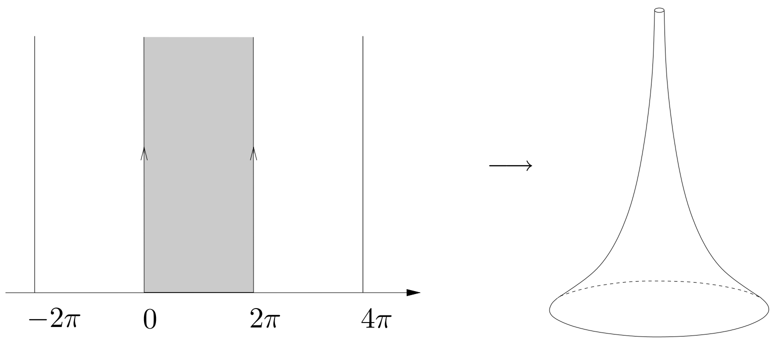

FIGURE 3. Pseudosphere.

EXAMPLE 4.6. To find Riemannian manifolds of constant positive curvature we have to find discrete subgroups of isometries of the sphere that act properly and freely. Let us consider the case where $K = 1$. Then $\Gamma \subset O(n+1)$. Since it must act freely on S^n, no element of $\Gamma \backslash \{Id\}$ can have 1 as an eigenvalue. We will see that, when n is even, S^n and $\mathbb{R}P^n$ are the only geodesically complete manifolds of constant curvature 1. Indeed, if $A \in \Gamma$, then A is an orthogonal $(n+1) \times (n+1)$ matrix and so all its eigenvalues have absolute value equal to 1. Moreover, its characteristic polynomial has odd degree $(n+1)$, implying that, if $A \neq I$, this polynomial has a real root equal to -1 (since it cannot have 1 as an eigenvalue). Consequently, A^2 has 1 as an eigenvalue and so it has to be the identity. Hence,

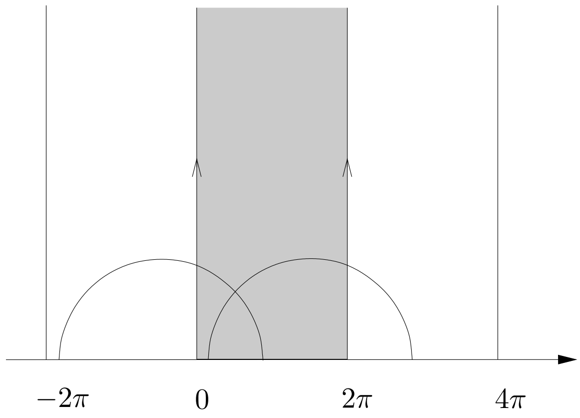

FIGURE 4. Trajectories of geodesics on the pseudosphere.

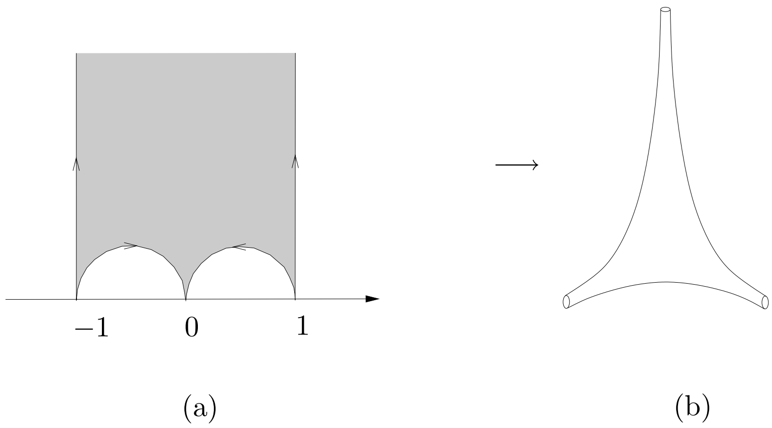

FIGURE 5. (a) Hyperbolic polygon, (b) Pair of pants.

the eigenvalues of A are either all equal to 1 (if $A = Id$) or all equal to -1, in which case $A = -Id$. We conclude that $\Gamma = \{\pm Id\}$ implying that our manifold is either S^n or $\mathbb{R}P^n$. If n is odd there are other possibilities which are classified in [**Wol78**].

EXERCISES 4.7.

(1) Prove that if the forms ω^i in an orthonormal co-frame satisfy $d\omega^i = \alpha \wedge \omega^i$ (with α a 1-form), then the connection forms ω_i^j are given by $\omega_i^j = \alpha(E_i)\omega^j - \alpha(E_j)\omega^i = -\omega_j^i$. Use this to confirm the results in Example 4.2.
(2) Let K be a real number and let $\rho = 1 + (\frac{K}{4})\sum_{i=1}^n (x^i)^2$. Let $V = \varphi(U)$ be a coordinate neighborhood of a manifold M of dimension n, with $U = B_\varepsilon(0) \subset \mathbb{R}^n$ (for some $\varepsilon > 0$). Show that, for the

Riemannian metric defined in V by

$$g_{ij}(p) = \frac{1}{\rho^2}\,\delta_{ij},$$

the sectional curvature is constant equal to K. Note that in this way we obtain manifolds with an arbitrary constant curvature.

(3) (*Schur Theorem*) Let M be a connected isotropic Riemannian manifold of dimension $n \geq 3$. Show that M has constant curvature. (**Hint:** Use the structure equations to show that $dK = 0$).

(4) To complete the details in Example 4.4, show that:
   (a) any discrete group of isometries of the Euclidean plane $\mathbb{R}^2$ acting properly and freely on $\mathbb{R}^2$ can only contain translations and gliding reflections and is generated by at most two elements;
   (b) show that any group generated by two gliding reflections can also be generated by a translation and a gliding reflection.

(5) Let H^2 be the hyperbolic plane. Show that:
   (a)
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z \ := \ \frac{az+b}{cz+d}$$
   defines an action of $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm Id\}$ on H^2 by orientation-preserving isometries;
   (b) for any two geodesics $c_1, c_2 : \mathbb{R} \to H^2$, parametrized by the arclength, there exists $g \in PSL(2,\mathbb{R})$ such that $c_1(s) = g \cdot c_2(s)$ for all $s \in \mathbb{R}$;
   (c) if $f : H^2 \to H^2$ is an orientation-preserving isometry then it must be a holomorphic function. Conclude that all orientation-preserving isometries are of the form $f(z) = g \cdot z$ for some $g \in PSL(2,\mathbb{R})$.

(6) Check that the isometries g, h of the hyperbolic plane in Example 4.5 identify the sides of the hyperbolic polygon in Figure 5.

(7) A **tractrix** is the curve described parametrically by

$$\begin{cases} x = u - \tanh u \\ y = \operatorname{sech} u \end{cases} \qquad (u > 0)$$

(its name derives from the property that the distance between any point in the curve and the x-axis along the tangent is constant equal to 1). Show that the surface of revolution generated by rotating a tractrix about the x-axis (**tractroid**) has constant Gauss curvature $K = -1$. Determine an open subset of the pseudosphere isometric to the tractroid. (**Remark:** The tractroid is not geodesically complete; in fact, it was proved by Hilbert in 1901 that any surface of constant negative curvature embedded in Euclidean 3-space must be incomplete).

(8) Show that the group of isometries of S^n is $O(n+1)$.

(9) Let G be a compact Lie group of dimension 2. Show that:
   (a) G is orientable;

(b) $\chi(G) = 0$;
(c) any left-invariant metric on G has constant curvature;
(d) G is the 2-torus T^2.

5. Isometric Immersions

Many Riemannian manifolds arise as submanifolds of another Riemannian manifold, by taking the induced metric (e.g. $S^n \subset \mathbb{R}^{n+1}$). In this section, we will analyze how the curvatures of the two manifolds are related.

Let $f : N \to M$ be an immersion of an n-manifold N on an m-manifold M. We know from Section 5 of Chapter 1 that, for each point $p \in N$, there is a neighborhood $V \subset N$ of p where f is an embedding onto its image. Hence $f(V)$ is a submanifold of M. To simplify notation, we will proceed as if f were the inclusion map, and will identify V with $f(V)$, as well as every element $v \in T_pN$ with $(df)_pv \in T_{f(p)}M$. Let $\langle \cdot, \cdot \rangle$ be a Riemannian metric on M and $\langle\langle \cdot, \cdot \rangle\rangle$ the induced metric on N (we then call f an **isometric immersion**). Then, for every $p \in V$, the tangent space T_pM can be decomposed as follows:

$$T_pM = T_pN \oplus (T_pN)^\perp.$$

Therefore, every element v of T_pM can be written uniquely as $v = v^\top + v^\perp$, where $v^\top \in T_pN$ is the tangential part of v and $v^\perp \in (T_pN)^\perp$ is the normal part of v. Let $\widetilde{\nabla}$ and ∇ be the Levi-Civita connections of $(M, \langle \cdot, \cdot \rangle)$ and $(N, \langle\langle \cdot, \cdot \rangle\rangle)$, respectively. Let X, Y be two vector fields in $V \subset N$ and let $\widetilde{X}$, $\widetilde{Y}$ be two extensions of X, Y to a neighborhood $W \subset M$ of V. Using the Koszul formula, we can easily check that

$$\nabla_X Y = \left(\widetilde{\nabla}_{\widetilde{X}} \widetilde{Y}\right)^\top$$

(cf. Exercise 3.3.6 in Chapter 3). We define the **second fundamental form** of N as

$$B(X, Y) := \widetilde{\nabla}_{\widetilde{X}} \widetilde{Y} - \nabla_X Y.$$

Note that this map is well defined, that is, it does not depend on the extensions $\widetilde{X}, \widetilde{Y}$ of X and Y (cf. Exercise 5.7.1). Moreover, it is bilinear, symmetric, and, for each $p \in V$, $B(X, Y)_p \in (T_pN)^\perp$ depends only on the values of X_p and Y_p.

Using the second fundamental form, we can define for each vector $n_p \in (T_pN)^\perp$ a symmetric bilinear map $H_{n_p} : T_pN \times T_pN \to \mathbb{R}$ through

$$H_{n_p}(X_p, Y_p) = \langle B(X_p, Y_p), n_p \rangle.$$

Hence, we have a quadratic form $\mathrm{II}_{n_p} : T_pN \to \mathbb{R}$, given by

$$\mathrm{II}_{n_p}(X_p) = H_{n_p}(X_p, X_p),$$

which is often called the **second fundamental form of f at p along the vector n_p**.

Finally, since H_{n_p} is bilinear, there exists a linear map $S_{n_p} : T_pN \to T_pN$ satisfying

$$\langle\langle S_{n_p}(X_p), Y_p\rangle\rangle = H_{n_p}(X_p, Y_p) = \langle B(X_p, Y_p), n_p\rangle$$

for all $X_p, Y_p \in T_pM$. It is easy to check that this linear map is given by

$$S_{n_p}(X_p) = -(\widetilde{\nabla}_{\widetilde{X}} n)_p^\top,$$

where n is a local extension of n_p normal to N. Indeed, since $\langle \tilde{Y}, n\rangle = 0$ on N and $\widetilde{X}$ is tangent to N, we have

$$\begin{aligned}\langle\langle S_n(X), Y\rangle\rangle &= \langle B(X,Y), n\rangle = \langle \widetilde{\nabla}_{\widetilde{X}} \widetilde{Y} - \nabla_X Y, n\rangle \\ &= \langle \widetilde{\nabla}_{\widetilde{X}} \widetilde{Y}, n\rangle = \widetilde{X} \cdot \langle \widetilde{Y}, n\rangle - \langle \widetilde{Y}, \widetilde{\nabla}_{\widetilde{X}} n\rangle \\ &= \langle -\widetilde{\nabla}_{\widetilde{X}} n, \widetilde{Y}\rangle = \langle\langle -(\widetilde{\nabla}_{\widetilde{X}} n)^\top, Y\rangle\rangle.\end{aligned}$$

Therefore

$$\langle\langle S_{n_p}(X_p), Y_p\rangle\rangle = \langle\langle -(\widetilde{\nabla}_{\widetilde{X}} n)_p^\top, Y_p\rangle\rangle$$

for all $Y_p \in T_pN$.

EXAMPLE 5.1. Let N be a **hypersurface** in M, i.e., let $\dim N = n$ and $\dim M = n+1$. Consider a point $p \in V$ (a neighborhood of N where f is an embedding), and a unit vector n_p normal to N at p. As the linear map $S_{n_p} : T_pN \to T_pN$ is symmetric, there exists an orthonormal basis of T_pN formed by eigenvectors $\{(E_1)_p, \dots, (E_n)_p\}$ (called **principal directions** at p) corresponding to a set of real eigenvalues $\lambda_1, \dots, \lambda_n$ (called **principal curvatures** at p). The determinant of the map S_{n_p} (equal to the product $\lambda_1 \cdots \lambda_n$) is called the **Gauss curvature of** f and $H := \frac{1}{n}\operatorname{tr} S_{n_p} = \frac{1}{n}(\lambda_1 + \cdots + \lambda_n)$ is called the **mean curvature** of f. When $n = 2$ and $M = \mathbb{R}^3$ with the Euclidean metric, the Gauss curvature of f is in fact the Gauss curvature of N as defined in Section 1 (cf. Example 5.5).

EXAMPLE 5.2. If, in the above example, $M = \mathbb{R}^{n+1}$ with the Euclidean metric, we can define the Gauss map $g : V \subset N \to S^n$, with values on the unit sphere, which, to each point $p \in V$, assigns the normal unit vector n_p. Since n_p is normal to T_pN, we can identify the tangent spaces T_pN and $T_{g(p)}S^n$ and obtain a well-defined map $(dg)_p : T_pN \to T_pN$. Note that, for each $X_p \in T_pN$, choosing a curve $c : I \to N$ such that $c(0) = p$ and $\dot{c}(0) = X_p$, we have

$$(dg)_p(X_p) = \frac{d}{dt}(g \circ c)|_{t=0} = \frac{d}{dt} n_{c(t)}|_{t=0} = (\widetilde{\nabla}_{\dot{c}} n)_p\,,$$

where we used the fact $\widetilde{\nabla}$ is the Levi-Civita connection for the Euclidean metric. However, since $\|n\| = 1$, we have

$$0 = \dot{c}(t) \cdot \langle n, n\rangle = 2\langle \widetilde{\nabla}_{\dot{c}}\, n, n\rangle,$$

implying that

$$(dg)_p(X_p) = (\widetilde{\nabla}_{\dot{c}}\, n)_p = (\widetilde{\nabla}_{\dot{c}}\, n)_p^\top = -S_{n_p}(X_p).$$

We conclude that the derivative of the Gauss map at p is $-S_{n_p}$.

Let us now relate the curvatures of N and M.

PROPOSITION 5.3. *Let p be a point in N, let X_p and Y_p be two linearly independent vectors in $T_pN \subset T_pM$ and let $\Pi \subset T_pN \subset T_pM$ be the two dimensional subspace generated by these vectors. Let $K^N(\Pi)$ and $K^M(\Pi)$ denote the corresponding sectional curvatures in N and M, respectively. Then*

$$K^N(\Pi) - K^M(\Pi) = \frac{\langle B(X_p,X_p), B(Y_p,Y_p)\rangle - \|B(X_p,Y_p)\|^2}{\|X_p\|^2\|Y_p\|^2 - \langle X_p,Y_p\rangle^2}.$$

PROOF. Observing that the right-hand side depends only on Π, we can assume without loss of generality that $\{X_p, Y_p\}$ is orthonormal. Let X, Y be local extensions of X_p, Y_p, defined on a neighborhood of p in N and tangent to N, also orthonormal. Let $\widetilde{X}, \widetilde{Y}$ be extensions of X, Y to a neighborhood of p in M. Moreover, consider a field of frames $\{E_1, \dots, E_{n+k}\}$, also defined on a neighborhood of p in M, such that $E_1, \dots, E_n$ are tangent to N, $E_1 = X$, $E_2 = Y$ on N, and $E_{n+1}, \dots, E_{n+k}$ are normal to N $(m = n + k)$. Then, since $B(X, Y)$ is normal to N,

$$B(X,Y) = \sum_{i=1}^{k} \langle B(X,Y), E_{n+i}\rangle\, E_{n+i} = \sum_{i=1}^{k} H_{E_{n+i}}(X,Y)\, E_{n+i}.$$

On the other hand,

$$\begin{aligned}
&K^N(\Pi) - K^M(\Pi) = -R^N(X_p,Y_p,X_p,Y_p) + R^M(X_p,Y_p,X_p,Y_p)\\
&= \langle(-\nabla_X \nabla_Y X + \nabla_Y \nabla_X X + \nabla_{[X,Y]} X\\
&+ \widetilde{\nabla}_{\widetilde{X}} \widetilde{\nabla}_{\widetilde{Y}} \widetilde{X} - \widetilde{\nabla}_{\widetilde{Y}} \widetilde{\nabla}_{\widetilde{X}} \widetilde{X} - \widetilde{\nabla}_{[\widetilde{X},\widetilde{Y}]} \widetilde{X})_p, Y_p\rangle\\
&= \langle(-\nabla_X \nabla_Y X + \nabla_Y \nabla_X X + \widetilde{\nabla}_{\widetilde{X}} \widetilde{\nabla}_{\widetilde{Y}} \widetilde{X} - \widetilde{\nabla}_{\widetilde{Y}} \widetilde{\nabla}_{\widetilde{X}} \widetilde{X})_p, Y_p\rangle,
\end{aligned}$$

where we have used the fact that $\widetilde{\nabla}_{[\widetilde{X},\widetilde{Y}]}\widetilde{X} - \nabla_{[X,Y]} X$ is normal to N (cf. Exercise 5.7.1). However, since on N

$$\begin{aligned}
&\widetilde{\nabla}_{\widetilde{Y}} \widetilde{\nabla}_{\widetilde{X}} \widetilde{X} = \widetilde{\nabla}_{\widetilde{Y}}(B(X,X) + \nabla_X X) =\\
&= \widetilde{\nabla}_{\widetilde{Y}} \left(\sum_{i=1}^{k} H_{E_{n+i}}(X,X)E_{n+i} + \nabla_X X\right)\\
&= \sum_{i=1}^{k}\Big(H_{E_{n+i}}(X,X)\widetilde{\nabla}_{\widetilde{Y}} E_{n+i} + \widetilde{Y}\cdot(H_{E_{n+i}}(X,X))E_{n+i}\Big) + \widetilde{\nabla}_{\widetilde{Y}} \nabla_X X,
\end{aligned}$$

we have

$$\langle\widetilde{\nabla}_{\widetilde{Y}} \widetilde{\nabla}_{\widetilde{X}} \widetilde{X}, Y\rangle = \sum_{i=1}^{k} H_{E_{n+i}}(X,X)\langle\widetilde{\nabla}_{\widetilde{Y}} E_{n+i}, Y\rangle + \langle\widetilde{\nabla}_{\widetilde{Y}} \nabla_X X, Y\rangle.$$

Moreover,

$$\begin{aligned} 0 &= \widetilde{Y} \cdot \langle E_{n+i}, Y \rangle = \langle \widetilde{\nabla}_{\widetilde{Y}} E_{n+i}, Y \rangle + \langle E_{n+i}, \widetilde{\nabla}_{\widetilde{Y}} Y \rangle \\ &= \langle \widetilde{\nabla}_{\widetilde{Y}} E_{n+i}, Y \rangle + \langle E_{n+i}, B(Y,Y) + \nabla_Y Y \rangle \\ &= \langle \widetilde{\nabla}_{\widetilde{Y}} E_{n+i}, Y \rangle + \langle E_{n+i}, B(Y,Y) \rangle \\ &= \langle \widetilde{\nabla}_{\widetilde{Y}} E_{n+i}, Y \rangle + H_{E_{n+i}}(Y,Y), \end{aligned}$$

and so

$$\begin{aligned} \langle \widetilde{\nabla}_{\widetilde{Y}} \widetilde{\nabla}_{\widetilde{X}} \widetilde{X}, Y \rangle &= -\sum_{i=1}^{k} H_{E_{n+i}}(X,X) H_{E_{n+i}}(Y,Y) + \langle \widetilde{\nabla}_{\widetilde{Y}} \nabla_X X, Y \rangle \\ &= -\sum_{i=1}^{k} H_{E_{n+i}}(X,X) H_{E_{n+i}}(Y,Y) + \langle \nabla_Y \nabla_X X, Y \rangle. \end{aligned}$$

Similarly, we can conclude that

$$\langle \widetilde{\nabla}_{\widetilde{X}} \widetilde{\nabla}_{\widetilde{Y}} \widetilde{X}, Y \rangle = -\sum_{i=1}^{k} H_{E_{n+i}}(X,Y) H_{E_{n+i}}(X,Y) + \langle \nabla_X \nabla_Y X, Y \rangle,$$

and then

$$\begin{aligned} & K^N(\Pi) - K^M(\Pi) = \\ &= \sum_{i=1}^{k} \left(-(H_{E_{n+i}}(X_p, Y_p))^2 + H_{E_{n+i}}(X_p, X_p) H_{E_{n+i}}(Y_p, Y_p) \right) \\ &= -\|B(X_P, Y_p)\|^2 + \langle B(X_p, X_p), B(Y_p, Y_p) \rangle. \end{aligned}$$

□

EXAMPLE 5.4. Again in the case of a hypersurface N, we choose an orthonormal basis $\{(E_1)_p, \dots, (E_n)_p\}$ of T_pN formed by eigenvectors of S_{n_p}, where $n_p \in (T_pN)^\perp$. Hence, considering a section Π of T_pN generated by two of these vectors $(E_i)_p, (E_j)_p$, and using $B(X_p, Y_p) = \langle\langle S_{n_p}(X_p), Y_p \rangle\rangle n_p$, we have

$$\begin{aligned} & K^N(\Pi) - K^M(\Pi) = \\ &= -\|B((E_i)_p, (E_j)_p)\|^2 + \langle B((E_i)_p, (E_i)_p), B((E_j)_p, (E_j)_p) \rangle \\ &= -\langle\langle S_{n_p}((E_i)_p), (E_j)_p \rangle\rangle^2 + \langle\langle S_{n_p}((E_i)_p), (E_i)_p \rangle\rangle \langle\langle S_{n_p}((E_j)_p), (E_j)_p \rangle\rangle \\ &= \lambda_i \lambda_j. \end{aligned}$$

EXAMPLE 5.5. In the special case where N is a 2-manifold, and $M = \mathbb{R}^3$ with the Euclidean metric, we have $K^M \equiv 0$ and hence $K^N(p) = \lambda_1 \lambda_2$, as promised in Example 5.1. Therefore, although λ_1 and λ_2 depend on the immersion, their **product** depends only on the intrinsic geometry of N. Gauss was so pleased by this discovery that he called it his **Theorema Egregium** ('Remarkable Theorem').

Let us now study in detail the particular case where N is a hypersurface in $M = \mathbb{R}^{n+1}$ with the Euclidean metric. Let $c : I \to N$ be a curve in N parametrized by arc length s and such that $c(0) = p$ and $\dot{c}(0) = X_p \in T_pN$. We will identify this curve c with the curve $f \circ c$ in $\mathbb{R}^{n+1}$. Considering the Gauss map $g : V \to S^n$ defined on a neighborhood V of p in N, we take the curve $n(s) := g \circ c(s)$ in S^n. Since $\widetilde{\nabla}$ is the Levi-Civita connection corresponding to the Euclidean metric in $\mathbb{R}^3$, we have $\langle \widetilde{\nabla}_{\dot{c}}\, \dot{c}, n\rangle = \langle \ddot{c}, n\rangle$. On the other hand,

$$\langle \widetilde{\nabla}_{\dot{c}}\, \dot{c}, n\rangle = \langle B(\dot{c}, \dot{c}) + \nabla_{\dot{c}}\, \dot{c}, n\rangle = \langle B(\dot{c}, \dot{c}), n\rangle = H_n(\dot{c}, \dot{c}) = \mathrm{II}_n(\dot{c}).$$

Hence, at $s = 0$, $\mathrm{II}_{g(p)}(X_p) = \langle \ddot{c}(0), n_p\rangle$. This value $k_{n_p} := \langle \ddot{c}(0), n_p\rangle$ is called the **normal curvature** of c at p. Since k_{n_p} is equal to $\mathrm{II}_{g(p)}(X_p)$, it does not depend on the curve, but only on its initial velocity. Because $\mathrm{II}_{g(p)}(X_p) = \langle\langle S_{g(p)}(X_p), X_p\rangle\rangle$, the critical values of these curvatures subject to $\|X_p\| = 1$ are equal to $\lambda_1, \ldots, \lambda_n$, and are called the **principal curvatures**. This is why in Example 5.1 we also called the eigenvalues of S_{n_p} principal curvatures. The Gauss curvature of f is then equal to the product of the principal curvatures, $K = \lambda_1 \ldots \lambda_n$. As the normal curvature does not depend on the choice of curve tangent to X_p at p, we can choose c to take values on a 2 containing n_p. Then $\ddot{c}(0)$ is parallel to the normal vector n_p, and

$$|k_n| = |\langle \ddot{c}(0), n\rangle| = \|\ddot{c}(0)\| = k_c,$$

where $k_c := \|\ddot{c}(0)\|$ is the so-called **curvature** of the curve c at $c(0)$.

EXAMPLE 5.6. Let us consider the following three surfaces: the 2-sphere, the cylinder and a saddle surface.

(1) Let p be any point on the sphere. Intuitively, all points of this surface are on the same side of the tangent plane at p, implying that both principal curvatures have the same sign (depending on the chosen orientation), and consequently that the Gauss curvature is positive at all points.
(2) If p is any point on the cylinder, one of the principal curvatures is zero (the maximum or the minimum, depending on the chosen orientation), and so the Gauss curvature is zero at all points.
(3) Finally, if p is a saddle point, the principal curvatures at p have opposite signs, and so the Gauss curvature is negative.

EXERCISES 5.7.

(1) Let M be a Riemannian manifold with Levi-Civita connection $\widetilde{\nabla}$, and let N be a submanifold endowed with the induced metric and Levi-Civita connection ∇. Let $\widetilde{X}, \widetilde{Y} \in \mathfrak{X}(M)$ be local extensions of $X, Y \in \mathfrak{X}(N)$. Recall that the second fundamental form of the inclusion of N in M is the map $B : T_pN \times T_pN \to (T_pN)^\perp$ defined at each point $p \in N$ by

$$B(X, Y) := \widetilde{\nabla}_{\widetilde{X}}\, \widetilde{Y} - \nabla_X\, Y.$$

Show that:

(a) $B(X,Y)$ does not depend on the choice of the extensions $\widetilde{X}, \widetilde{Y}$;

(b) $B(X,Y)$ is orthogonal to N;

(c) B is symmetric, i.e. $B(X,Y) = B(Y,X)$;

(d) $B(X,Y)_p$ depends only on the values of X_p and Y_p;

(e) $\widetilde{\nabla}_{[\widetilde{X},\widetilde{Y}]}\widetilde{X} - \nabla_{[X,Y]} X$ is orthogonal to N.

(2) Let $S^n(r) \subset \mathbb{R}^{n+1}$ be the n dimensional sphere of radius r.

a) Choosing at each point the outward pointing normal unit vector, what is the Gauss map of this inclusion?

b) What are the eigenvalues and eigenvectors of its derivative?

c) Show that all sectional curvatures are equal to $\frac{1}{r^2}$; conclude that $S^n(r)$ has constant curvature $\frac{1}{r^2}$.

(3) Let M be a Riemannian manifold. A submanifold $N \subset M$ is said to be **totally geodesic** if the image of any geodesic of M tangent to N at any point is contained in N. Show that:

(a) N is totally geodesic *iff* $B \equiv 0$, where B is the second fundamental form of N;

(b) if N is totally geodesic then the geodesics of N are geodesics of M;

(c) if N is the set of fixed points of an isometry then N is totally geodesic. Use this result to give examples of totally geodesic submanifolds of $\mathbb{R}^n$, S^n and H^n.

(4) Let N be a hypersurface in $\mathbb{R}^{n+1}$ and let p be a point in M. Show that

$$|K(p)| = \lim_{D\to p} \frac{\mathrm{vol}(g(D))}{\mathrm{vol}(D)}.$$

where D is a neighborhood of p and $g : V \subset N \to S^n$ is the Gauss map.

(5) Let M be a smooth Riemannian manifold, p a point in M and Π a section of T_pM. Considering a normal ball around p, $B_\varepsilon(p) := \exp_p(B_\varepsilon(0))$, take the set $N_p := \exp_p(B_\varepsilon(0) \cap \Pi)$. Show that:

a) The set N_p is a 2-dimensional submanifold of M formed by the segments of geodesics in $B_\varepsilon(p)$ which are tangent to Π at p;

b) If in N_p we use the metric induced by the metric in M, the sectional curvature $K^M(\Pi)$ is equal to the Gauss curvature of the 2-manifold N_p.

(6) Let M be a Riemannian manifold with Levi-Civita connection $\widetilde{\nabla}$ and let N be a hypersurface in M. Show that the absolute values of the principal curvatures are the geodesic curvatures (in M) of the geodesics of N tangent to the principal directions (the **geodesic curvature** of a curve $c : I \subset \mathbb{R} \to M$, parametrized by arclength,

is $k_g(s) = \|\widetilde{\nabla}_{\dot{c}(s)}\dot{c}(s)\|$; in the case of an oriented 2-dimensional Riemannian manifold, k_g is taken to be positive or negative according to the orientation of $\{\dot{c}(s), \widetilde{\nabla}_{\dot{c}(s)}\dot{c}(s)\}$, cf. Section 2).

(7) (*Surfaces of revolution*) Consider the map $f : \mathbb{R} \times (0, 2\pi) \to \mathbb{R}^3$ given by

$$f(s, \theta) = (h(s)\cos\theta, h(s)\sin\theta, g(s))$$

with $h > 0$ and g smooth maps such that

$$(h'(s))^2 + (g'(s))^2 = 1.$$

The image of f is the surface of revolution S with axis Oz, obtained by rotating the curve $\alpha(s) = (h(s), g(s))$, parametrized by the arclength s, around that axis.

(a) Show that f is an immersion.
(b) Show that $f_s := (df)\left(\frac{\partial}{\partial s}\right)$ and $f_\theta := (df)(\frac{\partial}{\partial \theta})$ are orthogonal.
(c) Determine the Gauss map and compute the matrix of the second fundamental form of S associated to the frame $\{E_s, E_\theta\}$, where $E_s := f_s$ and $E_\theta := \frac{1}{\|f_\theta\|} f_\theta$.
(d) Compute the mean curvature H and the Gauss curvature K of S.
(e) Using this result, give examples of surfaces of revolution with:
  (i) $K \equiv 0$;
  (ii) $K \equiv 1$;
  (iii) $K \equiv -1$;
  (iv) $H \equiv 0$ (not a plane).

(**Remark:** Surfaces with constant zero mean curvature are called minimal surfaces; it can be proved that if a compact surface with boundary has minimum area among all surfaces with the same boundary then it must be a minimal surface).

6. Notes on Chapter 4

6.1. Bibliographical notes. The material in this chapter can be found in most books on Riemannian geometry (e.g. [**Boo03, dC93, GHL04**]). The proof of The Gauss-Bonnet theorem (due to S. Chern) follows [**dC93**] closely. See [**KN96, Jos02**] to see how this theorem fits within the general theory of characteristic classes of fiber bundles. A more elementary discussion of isometric immersions of surfaces in $\mathbb{R}^3$ (including a proof of the Gauss-Bonnet Theorem) can be found in [**dC76, Mor98**].

CHAPTER 5

Geometric Mechanics

In this chapter we show how Riemannian Geometry can be used to give a geometric formulation of Newtonian Mechanics.

In Section 1 we define what is meant by an abstract **mechanical system**. Section 2 explains how **holonomic constraints** yield nontrivial examples of these, as for instance the **rigid body**, which is studied in detail in Section 3. **Non-holonomic constraints** are considered in Section 4.

Section 5 presents the **Lagrangian formulation** of mechanics, including **Noether's Theorem**, which relates symmetries to conservation laws. The dual **Hamiltonian formulation** is described in Section 6, and used to formulate the theory of **completely integrable systems** in Section 7.

1. Mechanical Systems

In Mechanics one studies the motions of particles or systems of particles subject to known forces.

Example 1.1. The motion of a single particle in n-dimensional space is described by a curve $x : I \subset \mathbb{R} \to \mathbb{R}^n$. It is generally assumed that the force acting on the particle depends only on its position and velocity. **Newton's Second Law** requires that the particle's motion satisfies the second order ordinary differential equation

$$m\ddot{x} = F(x, \dot{x}),$$

where $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is the force acting on the particle and $m > 0$ is the particle's mass. Therefore the solutions of this equation describe the possible motions of the particle.

It will prove advantageous to make the following generalization:

Definition 1.2. *A* **mechanical system** *is a triple* $(M, \langle \cdot, \cdot \rangle, \mathcal{F})$*, where:*

(i) M *is a differentiable manifold, called the* **configuration space***;*

(ii) $\langle \cdot, \cdot \rangle$ *is a Riemannian metric on* M *yielding the* **mass operator** $\mu : TM \to T^*M$*, defined by*

$$\mu(v)(w) = \langle v, w \rangle$$

for all $v, w \in T_pM$ *and* $p \in M$*;*

(iii) $\mathcal{F} : TM \to T^*M$ *is a differentiable map satisfying* $\mathcal{F}(T_pM) \subset T_p^*M$ *for all* $p \in M$*, called the* **external force**.

A **motion** *of the mechanical system is a solution $c : I \subset \mathbb{R} \to M$ of* **Newton's equation**

$$\mu\left(\frac{D\dot{c}}{dt}\right) = \mathcal{F}(\dot{c}).$$

REMARK 1.3. In particular, the geodesics of a Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ are the motions of the mechanical system $(M, \langle\cdot,\cdot\rangle, 0)$ (describing a **free particle** on M).

EXAMPLE 1.4. For the mechanical system comprising a single particle moving in n-dimensional space, the configuration space is clearly $\mathbb{R}^n$. If we set

$$\langle\langle v, w\rangle\rangle = m\,\langle v, w\rangle$$

for all $v, w \in \mathbb{R}^n$, where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product in $\mathbb{R}^n$, then the Levi-Civita connection of $\langle\langle\cdot,\cdot\rangle\rangle$ will still be the trivial connection, and

$$\frac{D\dot{x}}{dt} = \ddot{x}.$$

Setting

$$\mathcal{F}(x, v)(w) = \langle F(x, v), w\rangle$$

for all $v, w \in \mathbb{R}^n$, we see that

$$\begin{aligned}\mu\left(\frac{D\dot{x}}{dt}\right) = \mathcal{F}(x, \dot{x}) &\Leftrightarrow \mu\left(\frac{D\dot{x}}{dt}\right)(v) = \mathcal{F}(x, \dot{x})(v) \text{ for all } v \in \mathbb{R}^n \\ &\Leftrightarrow m\,\langle\ddot{x}, v\rangle = \langle F(x, \dot{x}), v\rangle \text{ for all } v \in \mathbb{R}^n \\ &\Leftrightarrow m\ddot{x} = F(x, \dot{x}).\end{aligned}$$

Hence the motions of the particle are the motions of the mechanical system $(\mathbb{R}^n, \langle\langle\cdot,\cdot\rangle\rangle, \mathcal{F})$.

DEFINITION 1.5. *Let $(M, \langle\cdot,\cdot\rangle, \mathcal{F})$ be a mechanical system. The external force $\mathcal{F}$ is said to be:*

(i) **positional** *if $\mathcal{F}(v)$ depends only on $\pi(v)$, where $\pi : TM \to M$ is the natural projection;*
(ii) **conservative** *if there exists $U : M \to \mathbb{R}$ such that $\mathcal{F}(v) = -(dU)_{\pi(v)}$ for all $v \in TM$ (the function U is called the* **potential energy***).*

REMARK 1.6. In particular any conservative force is positional.

DEFINITION 1.7. *Let $(M, \langle\cdot,\cdot\rangle, \mathcal{F})$ be a mechanical system. The* **kinetic energy** *is the differentiable map $K : TM \to \mathbb{R}$ given by*

$$K(v) = \frac{1}{2}\langle v, v\rangle$$

for all $v \in TM$.

EXAMPLE 1.8. For the mechanical system comprising a single particle moving in n-dimensional space, one has

$$K(v) = \frac{1}{2}m\,\langle v, v\rangle.$$

THEOREM 1.9. (Conservation of Energy) *In a conservative mechanical system* $(M, \langle\cdot,\cdot\rangle, -dU)$*, the* **mechanical energy** $E = K + U$ *is constant along any motion.*

PROOF.

$$\begin{aligned}\frac{dE}{dt} &= \frac{d}{dt}\left(\frac{1}{2}\langle\dot{c},\dot{c}\rangle + U(c(t))\right) = \left\langle\frac{D\dot{c}}{dt},\dot{c}\right\rangle + dU(\dot{c}) \\ &= \mu\left(\frac{D\dot{c}}{dt}\right)(\dot{c}) - \mathcal{F}(\dot{c}) = 0.\end{aligned}$$

□

A particularly simple example of a conservative mechanical system is $(M, \langle\cdot,\cdot\rangle, 0)$, whose motions are the geodesics of $(M, \langle\cdot,\cdot\rangle)$. In fact, the motions of **any** conservative system can be suitably reinterpreted as the geodesics of a certain metric.

DEFINITION 1.10. *Let* $(M, \langle\cdot,\cdot\rangle, -dU)$ *be a conservative mechanical system and* $h \in \mathbb{R}$ *such that*

$$M_h = \{p \in M \mid U(p) < h\} \neq \varnothing.$$

The **Jacobi metric** *on the manifold* M_h *is given by*

$$\langle\langle v, w\rangle\rangle = 2\,[h - U(p)]\,\langle v, w\rangle$$

for all $v, w \in T_pM$ *and* $p \in M$*.*

THEOREM 1.11. (Jacobi) *The motions of a conservative mechanical system* $(M, \langle\cdot,\cdot\rangle, -dU)$ *with mechanical energy* h *are, up to reparametrization, geodesics of the Jacobi metric on* M_h*.*

PROOF. We shall need the two following lemmas, whose proofs are left as exercises:

LEMMA 1.12. *Let* $(M, \langle\cdot,\cdot\rangle)$ *be a Riemannian manifold with Levi-Civita connection* ∇ *and* $\langle\langle\cdot,\cdot\rangle\rangle = e^{2\rho}\langle\cdot,\cdot\rangle$ *a metric conformally related to* $\langle\cdot,\cdot\rangle$ *(where* $\rho \in C^\infty(M)$*). Then the Levi-Civita connection* $\widetilde{\nabla}$ *of* $\langle\langle\cdot,\cdot\rangle\rangle$ *is given by*

$$\widetilde{\nabla}_X Y = \nabla_X Y + d\rho(X)Y + d\rho(Y)X - \langle X, Y\rangle \operatorname{grad}\rho$$

for all $X, Y \in \mathfrak{X}(M)$*, where* $\operatorname{grad}\rho$ *is defined through* $\langle \operatorname{grad}\rho, Z\rangle = d\rho(Z)$ *for all* $Z \in \mathfrak{X}(M)$*.*

LEMMA 1.13. *A curve* $c : I \subset \mathbb{R} \to M$ *is a reparametrized geodesic of the Riemannian manifold* $(M, \langle\cdot,\cdot\rangle)$ *if and only if it satisfies*

$$\frac{D\dot{c}}{dt} = f(t)\,\dot{c}$$

for some differentiable function $f : I \to \mathbb{R}$*.*

We now prove Jacobi's Theorem. Let $c : I \subset \mathbb{R} \to M$ be a motion of $(M, \langle\cdot,\cdot\rangle, -dU)$ with mechanical energy h. Then Lemma 1.12 yields

$$\frac{\widetilde{D}\dot{c}}{dt} = \frac{D\dot{c}}{dt} + 2d\rho(\dot{c})\,\dot{c} - \langle\dot{c},\dot{c}\rangle\,\operatorname{grad}\rho,$$

where $\frac{\widetilde{D}}{dt}$ is the covariant derivative along c with respect to the Jacobi metric and $e^{2\rho} = 2(h - U)$. Newton's equation yields

$$\frac{D\dot{c}}{dt} = -\operatorname{grad} U = -2e^{2\rho}\operatorname{grad}\rho,$$

and by conservation of energy

$$\langle\dot{c},\dot{c}\rangle = 2K = 2(h - U) = e^{2\rho}.$$

Consequently we have

$$\frac{\widetilde{D}\dot{c}}{dt} = 2d\rho(\dot{c})\,\dot{c},$$

which by Lemma 1.13 means that c is a reparametrized geodesic of the Jacobi metric. $\square$

A very useful expression for writing Newton's equation in local coordinates is the following:

Proposition 1.14. *Let $(M, \langle\cdot,\cdot\rangle, \mathcal{F})$ be a mechanical system. If $(x^1, \ldots, x^n)$ are local coordinates on M and $(x^1, \ldots, x^n, v^1, \ldots, v^n)$ are the local coordinates induced on TM then*

$$\mu\left(\frac{D\dot{c}}{dt}(t)\right) = \sum_{i=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial K}{\partial v^i}\left(x(t),\dot{x}(t)\right)\right) - \frac{\partial K}{\partial x^i}\left(x(t),\dot{x}(t)\right)\right]dx^i.$$

In particular, if $\mathcal{F} = -dU$ is conservative then the equations of motion are

$$\frac{d}{dt}\left(\frac{\partial K}{\partial v^i}\left(x(t),\dot{x}(t)\right)\right) - \frac{\partial K}{\partial x^i}\left(x(t),\dot{x}(t)\right) = -\frac{\partial U}{\partial x^i}(x(t))$$

$(i = 1, \ldots, n)$.

Proof. Recall that the local coordinates $(x^1, \ldots, x^n, v^1, \ldots, v^n)$ on TM label the vector

$$\sum_{i=1}^{n} v^i \frac{\partial}{\partial x^i}$$

which is tangent to M at the point with coordinates $(x^1, \ldots, x^n)$. Therefore, we have

$$K(x^1, \ldots, x^n, v^1, \ldots, v^n) = \frac{1}{2}\sum_{i,j=1}^{n} g_{ij}(x^1, \ldots, x^n)v^i v^j,$$

where

$$g_{ij} = \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle$$

are the components of the metric in this coordinate system. Consequently,

$$\frac{\partial K}{\partial v^i} = \sum_{j=1}^{n} g_{ij} v^j$$

and hence

$$\frac{\partial K}{\partial v^i}\,(x(t), \dot{x}(t)) = \sum_{j=1}^{n} g_{ij}(x(t))\dot{x}^j(t),$$

leading to

$$\frac{d}{dt}\left(\frac{\partial K}{\partial v^i}\,(x(t), \dot{x}(t))\right) = \sum_{j=1}^{n} g_{ij}(x(t))\ddot{x}^j(t) + \sum_{j,k=1}^{n} \frac{\partial g_{ij}}{\partial x^k}(x(t))\dot{x}^k(t)\dot{x}^j(t).$$

Moreover,

$$\frac{\partial K}{\partial x^i} = \frac{1}{2}\sum_{j,k=1}^{n} \frac{\partial g_{jk}}{\partial x^i} v^j v^k,$$

and hence

$$\frac{\partial K}{\partial x^i}\,(x(t), \dot{x}(t)) = \frac{1}{2}\sum_{j,k=1}^{n} \frac{\partial g_{jk}}{\partial x^i}(x(t))\dot{x}^j(t)\dot{x}^k(t).$$

We conclude that

$$\frac{d}{dt}\left(\frac{\partial K}{\partial v^i}\,(x(t), \dot{x}(t))\right) - \frac{\partial K}{\partial x^i}\,(x(t), \dot{x}(t)) =$$
$$\sum_{j=1}^{n} g_{ij}(x(t))\ddot{x}^j(t) + \sum_{j,k=1}^{n}\left(\frac{\partial g_{ij}}{\partial x^k}(x(t)) - \frac{1}{2}\frac{\partial g_{jk}}{\partial x^i}(x(t))\right)\dot{x}^j(t)\dot{x}^k(t).$$

On the other hand, if $v, w \in T_pM$ are written as

$$v = \sum_{i=1}^{n} v^i \frac{\partial}{\partial x^i}, \quad w = \sum_{i=1}^{n} w^i \frac{\partial}{\partial x^i}$$

then we have

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

and hence

$$\mu(v) = \sum_{i,j=1}^{n} g_{ij} v^i dx^j = \sum_{i,j=1}^{n} g_{ij} v^j dx^i.$$

Therefore

$$\mu\left(\frac{D\dot{c}}{dt}(t)\right) = \sum_{i,j=1}^{n} g_{ij}(x(t))\left(\ddot{x}^j(t) + \sum_{k,l=1}^{n}\Gamma^j_{kl}(x(t))\dot{x}^k(t)\dot{x}^l(t)\right)dx^i.$$

Since

$$\sum_{j=1}^{n} g_{ij}\Gamma^j_{kl} = \frac{1}{2}\sum_{j,m=1}^{n} g_{ij}\, g^{jm}\left(\frac{\partial g_{ml}}{\partial x^k} + \frac{\partial g_{mk}}{\partial x^l} - \frac{\partial g_{kl}}{\partial x^m}\right)$$
$$= \frac{1}{2}\left(\frac{\partial g_{il}}{\partial x^k} + \frac{\partial g_{ik}}{\partial x^l} - \frac{\partial g_{kl}}{\partial x^i}\right),$$

we have

$$\sum_{j,k,l=1}^{n} g_{ij}(x(t))\Gamma^j_{kl}(x(t))\dot{x}^k(t)\dot{x}^l(t)$$
$$= \frac{1}{2}\sum_{k,l=1}^{n}\left(\frac{\partial g_{il}}{\partial x^k}(x(t)) + \frac{\partial g_{ik}}{\partial x^l}(x(t)) - \frac{\partial g_{kl}}{\partial x^i}(x(t))\right)\dot{x}^k(t)\dot{x}^l(t)$$
$$= \frac{1}{2}\sum_{j,k=1}^{n}\left(\frac{\partial g_{ij}}{\partial x^k}(x(t)) + \frac{\partial g_{ik}}{\partial x^j}(x(t)) - \frac{\partial g_{jk}}{\partial x^i}(x(t))\right)\dot{x}^j(t)\dot{x}^k(t)$$
$$= \sum_{j,k=1}^{n}\left(\frac{\partial g_{ij}}{\partial x^k}(x(t)) - \frac{1}{2}\frac{\partial g_{jk}}{\partial x^i}(x(t))\right)\dot{x}^j(t)\dot{x}^k(t),$$

which completes the proof. □

EXAMPLE 1.15.

(1) (*Particle in a central field*) Consider a particle of mass $m > 0$ moving in $\mathbb{R}^2$ under the influence of a conservative force whose potential energy U depends only on the distance $r = \sqrt{x^2+y^2}$ to the origin, $U = u(r)$. The equations of motion are most easily solved when written in **polar coordinates** (r, θ), defined by

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$$

Since

$$dx = \cos\theta dr - r\sin\theta d\theta;$$
$$dy = \sin\theta dr + r\cos\theta d\theta,$$

it is easily seen that the Euclidean metric is written in these coordinates as

$$\langle\cdot,\cdot\rangle = dx\otimes dx + dy\otimes dy = dr\otimes dr + r^2 d\theta\otimes d\theta,$$

and hence

$$K\left(r,\theta,v^r,v^\theta\right) = \frac{1}{2}m\left[(v^r)^2 + r^2\left(v^\theta\right)^2\right]$$

Therefore we have

$$\frac{\partial K}{\partial v^r} = mv^r, \quad \frac{\partial K}{\partial v^\theta} = mr^2 v^\theta, \quad \frac{\partial K}{\partial r} = mr\left(v^\theta\right)^2, \quad \frac{\partial K}{\partial \theta} = 0,$$

and consequently Newton's equations are written

$$\frac{d}{dt}\left(m\dot{r}\right) - mr\dot{\theta}^2 = -u'(r);$$
$$\frac{d}{dt}\left(mr^2\dot{\theta}\right) = 0.$$

Notice that the **angular momentum**

$$p_\theta = mr^2\dot{\theta}$$

is constant along the motion. This conservation law can be traced to the fact that neither K nor U depend on θ.

(2) (*Christoffel symbols for the* 2*-sphere*) The metric for the 2-sphere $S^2 \subset \mathbb{R}^3$ is written as

$$\langle\cdot,\cdot\rangle = d\theta \otimes d\theta + \sin^2\theta \, d\varphi \otimes d\varphi$$

in the usual local coordinates (θ,φ) defined by the parametrization

$$\phi(\theta,\varphi) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

(cf. Exercise 3.3.4 in Chapter 3). A quick way to obtain the Christoffel symbols in this coordinate system is to write out Newton's equations for a free particle (of mass $m = 1$, say) on S^2. We have

$$K\left(\theta,\varphi,v^\theta,v^\varphi\right) = \frac{1}{2}\left[\left(v^\theta\right)^2 + \sin^2\theta\,(v^\varphi)^2\right]$$

and hence

$$\frac{\partial K}{\partial v^\theta} = v^\theta, \quad \frac{\partial K}{\partial v^\varphi} = \sin^2\theta\, v^\varphi, \quad \frac{\partial K}{\partial \theta} = \sin\theta\cos\theta\,(v^\varphi)^2, \quad \frac{\partial K}{\partial \varphi} = 0.$$

Consequently Newton's equation are written

$$\frac{d}{dt}\left(\dot{\theta}\right) - \sin\theta\cos\theta\,\dot{\varphi}^2 = 0 \Leftrightarrow \ddot{\theta} - \sin\theta\cos\theta\,\dot{\varphi}^2 = 0$$
$$\frac{d}{dt}\left(\sin^2\dot{\varphi}\right) = 0 \Leftrightarrow \ddot{\varphi} + 2\cot\theta\,\dot{\theta}\,\dot{\varphi} = 0$$

Since these must be the equations for a geodesic on S^2, by comparing with the geodesic equations

$$\ddot{x}^i + \sum_{j,k=1}^{n} \dot{x}^j\dot{x}^k = 0 \qquad (i = 1,\dots,n)$$

one immediately reads off the nonvanishing Christoffel symbols:

$$\Gamma^\theta_{\varphi\varphi} = -\sin\theta\cos\theta, \qquad \Gamma^\theta_{\theta\varphi} = \Gamma^\theta_{\varphi\theta} = \cot\theta.$$

EXERCISES 1.16.

(1) Generalize Examples 1.1, 1.4 and 1.8 to a system of N particles moving in $\mathbb{R}^n$.

(2) Let $(M, \langle\cdot,\cdot\rangle, \mathcal{F})$ be a mechanical system. Show that Newton's equation defines a flow on TM, generated by the vector field $X \in \mathfrak{X}(TM)$ whose local expression is

$$X = v^i \frac{\partial}{\partial x^i} + \left(\sum_{j=1}^{n} g^{ij} F_j(x,v) - \sum_{j,k=1}^{n} \Gamma^i_{jk}(x) v^j v^k \right) \frac{\partial}{\partial v^i},$$

where $(x^1, \ldots, x^n)$ are local coordinates on M, $(x^1, \ldots, x^n, v^1, \ldots, v^n)$ are the local coordinates induced on TM, and

$$\mathcal{F} = \sum_{i=1}^{n} F_i(x,v) dx^i$$

on these coordinates. What are the fixed points of the flow?

(3) (*Harmonic oscillator*) The **harmonic oscillator** (in appropriate units) is the conservative mechanical system $(\mathbb{R}, dx \otimes dx, -dU)$, where $U : \mathbb{R} \to \mathbb{R}$ is given by

$$U(x) = \frac{1}{2}\omega^2 x^2.$$

(a) Write the equation of motion and its general solution.

(b) Friction can be included in this model by consedering the external force

$$\mathcal{F}\left(u \frac{d}{dx}\right) = -dU - ku\, dx$$

(where $k > 0$ is a constant). Write the equation of motion of this new mechanical system and its general solution.

(c) Generalize the results above to the n-dimensional harmonic oscillator, whose potential energy $U : \mathbb{R}^n \to \mathbb{R}$ is given by

$$U(x^1, \ldots, x^n) = \frac{1}{2}\omega^2 \left(\left(x^1\right)^2 + \ldots + \left(x^1\right)^2 \right).$$

(4) Consider the conservative mechanical system $(\mathbb{R}, dx \otimes dx, -dU)$. Show that:

(a) The flow determined by Newton's equation on $T\mathbb{R} \cong \mathbb{R}^2$ is generated by the vector field

$$X = v\frac{\partial}{\partial x} - U'(x)\frac{\partial}{\partial v} \in \mathfrak{X}(\mathbb{R}^2).$$

(b) The fixed points of the flow are the points of the form $(x_0, 0)$, where x_0 is a critical point of U.

(c) If x_0 is a maximum of U with $U''(x_0) < 0$ then $(x_0, 0)$ is an unstable fixed point.

(d) If x_0 is a minimum of U with $U''(x_0) > 0$ then $(x_0, 0)$ is a stable fixed point, with arbitrarily small neighborhoods formed by periodic orbits.

(e) The periods of these orbits converge to $2\pi U''(x_0)^{-\frac{1}{2}}$ as they approach $(x_0, 0)$.
(f) Locally, any conservative mechanical system $(M, \langle\cdot,\cdot\rangle, -dU)$ with $\dim M = 1$ is of the form above.

(5) Prove Lemma 1.12. (**Hint:** Use the Koszul formula).
(6) Prove Lemma 1.13.
(7) If $(M, \langle\cdot,\cdot\rangle)$ is a compact Riemannian manifold, it is known that there exists a nontrivial periodic geodesic. Use this fact to show that if M is compact then any conservative mechanical system $(M, \langle\cdot,\cdot\rangle, -dU)$ admits a nontrivial periodic motion.
(8) Recall that the hyperbolic plane is the upper half plane

$$H = \left\{(x, y) \in \mathbb{R}^2 \mid y > 0\right\}$$

with the Riemannian metric

$$\langle\cdot,\cdot\rangle = \frac{1}{y^2}\left(dx \otimes dx + dy \otimes dy\right)$$

(cf. Exercise 3.3.5 in Chapter 3). Use Proposition 1.14 to compute the Christoffel symbols for the Levi-Civita connection of $(H, \langle\cdot,\cdot\rangle)$ in the coordinates (x, y).
(9) (*Kepler problem*) The **Kepler problem** (in appropriate units) consists in determining the motion of a particle of mass $m = 1$ in the central potential

$$U(r) = -\frac{1}{r}.$$

(a) Show that the equations of motion can be integrated to

$$r^2\dot{\theta} = p_\theta;$$
$$\frac{\dot{r}^2}{2} + \frac{{p_\theta}^2}{2r^2} - \frac{1}{r} = E,$$

where E and p_θ are integration constants.
(b) Use these equations to show that $u = \frac{1}{r}$ satisfies the linear ODE

$$\frac{d^2u}{d\theta^2} + u = \frac{1}{{p_\theta}^2}.$$

(c) Assuming that the **pericenter** (i.e. the point in the particle's orbit closer to the center of attraction $r = 0$) occurs at $\theta = 0$, show that the equation of the particle's trajectory is

$$r = \frac{{p_\theta}^2}{1 + \varepsilon\cos\theta},$$

where

$$\varepsilon = \sqrt{1 + 2{p_\theta}^2 E}.$$

(**Remark:** This is the equation of a conic section with eccentricity ε in polar coordinates).

(d) Characterize all geodesics of $\mathbb{R}^2 \setminus \{(0,0)\}$ with the Riemannian metric
$$\langle \cdot, \cdot \rangle = \frac{1}{\sqrt{x^2+y^2}} \left(dx \otimes dx + dy \otimes dy \right).$$
Show that this manifold is isometric to the surface of a cone with aperture $\frac{\pi}{3}$.

2. Holonomic Constraints

Many mechanical systems involve particles or systems of particles whose positions are constrained (for example, a simple pendulum, a particle moving on a given surface, or a rigid system of particles connected by massless rods). To account for these we introduce the following definition:

Definition 2.1. *A* **holonomic constraint** *on a mechanical system* $(M, \langle \cdot, \cdot \rangle, \mathcal{F})$ *is a submanifold* $N \subset M$ *such that* $\dim N < \dim M$*. A curve* $c : I \subset \mathbb{R} \to M$ *is said to be* **compatible** *with* N *if* $c(t) \in N$ *for all* $t \in I$*.*

Example 2.2.

(1) A particle of mass $m > 0$ moving in $\mathbb{R}^2$ subject to a constant gravitational acceleration g is modelled by the mechanical system $(\mathbb{R}^2, \langle\langle \cdot, \cdot \rangle\rangle, -mg\, dy)$, where
$$\langle\langle v, w \rangle\rangle = m \langle v, w \rangle$$
($\langle \cdot, \cdot \rangle$ being the Euclidean inner product on $\mathbb{R}^2$). A **simple pendulum** is obtained by connecting the particle to a fixed pivoting point by an ideal massless rod of length $l > 0$. Assuming the pivoting point to be the origin, this corresponds to the holonomic constraint
$$N = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = l^2\}$$
(diffeomorphic to S^1).

(2) Similarly, a particle of mass $m > 0$ moving in $\mathbb{R}^3$ subject to a constant gravitational acceleration g is modelled by the mechanical system $(\mathbb{R}^3, \langle\langle \cdot, \cdot \rangle\rangle, -mg\, dz)$, where
$$\langle\langle v, w \rangle\rangle = m \langle v, w \rangle$$
($\langle \cdot, \cdot \rangle$ being the Euclidean inner product on $\mathbb{R}^3$). Requiring the particle to move on a surface of equation $z = f(x,y)$ yields the holonomic constraint
$$N = \{(x,y,z) \in \mathbb{R}^3 \mid z = f(x,y)\}.$$

(3) A system of k particles of masses $m_1, \ldots, m_k$ moving freely in $\mathbb{R}^3$ is modelled by the mechanical system $(\mathbb{R}^{3k}, \langle\langle \cdot, \cdot \rangle\rangle, 0)$, where
$$\langle\langle (v_1, \ldots, v_k), (w_1, \ldots, w_k) \rangle\rangle = \sum_{i=1}^{k} m_i \langle v_i, w_i \rangle$$

($\langle \cdot, \cdot \rangle$ being the Euclidean inner product on $\mathbb{R}^3$). A **rigid body** is obtained by connecting all particles by ideal massless rods, and correspond to the holonomic constraint

$$N = \left\{ (x_1, \dots, x_k) \in \mathbb{R}^{3k} \mid \|x_i - x_j\| = d_{ij} \text{ for } 1 \leq i < j \leq k \right\}.$$

If at least three particles are not collinear, N is easily seen to be diffeomorphic to $\mathbb{R}^3 \times O(3)$.

Keeping the particles on the holonomic constraint requires an additional external force (provided by the rods or by the surface in the examples above).

DEFINITION 2.3. *A* **reaction force** *in a mechanical system with holonomic constraint* $(M, \langle \cdot, \cdot \rangle, \mathcal{F}, N)$ *is a map* $\mathcal{R} : TN \to T^*M$ *with* $\mathcal{R}(T_pN) \subset T_p^*M$ *for all* $p \in N$ *such that the* **generalized Newton equation**

$$\mu \left(\frac{D\dot{c}}{dt} \right) = (\mathcal{F} + \mathcal{R})(\dot{c})$$

has solutions for any initial condition in TN.

REMARK 2.4. Since the reaction force is defined only on vectors tangent to the holonomic constraint N, any solution of the generalized Newton equation is necessarily compatible with N.

For any holonomic constraint there exist in general infinite possible choices of reaction forces. The following definition yields a particularly useful criterion for selecting reaction forces.

DEFINITION 2.5. *A reaction force in a mechanical system with holonomic constraint* $(M, \langle \cdot, \cdot \rangle, \mathcal{F}, N)$ *is said to be* **perfect**, *or to satisfy* **D'Alembert's principle**, *if*

$$\mu^{-1}\mathcal{R}(v) \in (T_pN)^\perp$$

for all $v \in T_pN$ *and* $p \in N$.

REMARK 2.6. The variation of the kinetic energy of a solution of the generalized Newton equation is

$$\frac{dK}{dt} = \left\langle \frac{D\dot{c}}{dt}, \dot{c} \right\rangle = \mathcal{F}(\dot{c}) + \mathcal{R}(\dot{c}) = \mathcal{F}(\dot{c}) + \left\langle \mu^{-1}\mathcal{R}, \dot{c} \right\rangle.$$

Therefore, a reaction force is perfect if and only if it does not creates nor dissipates energy along any motion compatible with the constraint.

EXAMPLE 2.7. In each of the examples above, requiring the reaction force to be perfect amounts to requiring that:

(1) **Simple pendulum:** The force transmitted by the rod is purely radial (i.e. there is no damping);
(2) **Particle on a surface:** The force exerted by the surface is orthogonal to it (i.e. the surface is frictionless);
(3) **Rigid body:** The cohesive forces do not dissipate energy.

The next result establishes the existence and unicity of perfect reaction forces.

THEOREM 2.8. *Given any mechanical system with holonomic constraint* $(M, \langle\cdot,\cdot\rangle, \mathcal{F}, N)$*, there exists a unique reaction force* $\mathcal{R} : TN \to T^*M$ *satisfying D'Alembert's principle. The solutions of the generalized Newton Law*

$$\mu\left(\frac{D\dot{c}}{dt}\right) = (\mathcal{F} + \mathcal{R})(\dot{c})$$

are exactly the motions of the mechanical system $(N, \langle\langle\cdot,\cdot\rangle\rangle, \mathcal{F}_N)$*, where* $\langle\langle\cdot,\cdot\rangle\rangle$ *is the metric induced on* N *by* $\langle\cdot,\cdot\rangle$ *and* $\mathcal{F}_N$ *is the constraint of* $\mathcal{F}$ *to* N*. In particular, if* $\mathcal{F} = -dU$ *is conservative then* $\mathcal{F}_N = -d\left(U|_N\right)$*.*

PROOF. Recall from Chapter 4, Section 5 that if $\widetilde{\nabla}$ is the Levi-Civita connection of $(M, \langle\cdot,\cdot\rangle)$ and ∇ is the Levi-Civita connection of $(N, \langle\langle\cdot,\cdot\rangle\rangle)$

$$\nabla_X Y = \left(\widetilde{\nabla}_{\widetilde{X}}\widetilde{Y}\right)^\top$$

for all $X, Y \in \mathfrak{X}(N)$, where $\widetilde{X}, \widetilde{Y}$ are any extensions of X, Y to $\mathfrak{X}(M)$ (as usual $v = v^\top + v^\perp$ designates the unique decomposition arising from the splitting $T_pM = T_pN \oplus (T_pN)^\perp$ for each $p \in N$). Moreover, the second fundamental form of N,

$$B(X, Y) = \widetilde{\nabla}_{\widetilde{X}}\widetilde{Y} - \nabla_X Y = \left(\widetilde{\nabla}_{\widetilde{X}}\widetilde{Y}\right)^\perp,$$

is well defined, and $B(X, Y)_p \in (T_pN)^\perp$ is a symmetric bilinear function of X_p, Y_p for all $p \in N$.

Assume that a perfect reaction force $\mathcal{R}$ exists; then the solutions of the generalized Newton equation satisfy

$$\widetilde{\nabla}_{\dot{c}}\,\dot{c} = \mu^{-1}\mathcal{F} + \mu^{-1}\mathcal{R}.$$

Since by hypothesis $\mu^{-1}\mathcal{R}$ is orthogonal to N, the component of this equation tangent to N yields

$$\nabla_{\dot{c}}\,\dot{c} = \mu^{-1}\mathcal{F}_N,$$

as for any $v \in TN$ one has

$$\langle\langle(\mu^{-1}\mathcal{F})^\top, v\rangle\rangle = \langle\mu^{-1}\mathcal{F}, v\rangle = \mathcal{F}(v) = \mathcal{F}_N(v) = \langle\langle\mu^{-1}\mathcal{F}_N, v\rangle\rangle.$$

Hence c is a motion of $(N, \langle\langle\cdot,\cdot\rangle\rangle, \mathcal{F}_N)$. The component of the generalized Newton equation orthogonal to N yields

$$B(\dot{c}, \dot{c}) = (\mu^{-1}\mathcal{F})^\perp + \mu^{-1}\mathcal{R}.$$

Therefore, if $\mathcal{R}$ exists then it must satisfy

$$\mathcal{R}(v) = \mu(B(v, v)) - \mu\left[\left(\mu^{-1}\mathcal{F}(v)\right)^\perp\right] \tag{23}$$

for all $v \in TN$. This proves uniqueness.

To prove existence, **define** $\mathcal{R}$ through (23), which certainly guarantees that $\mu^{-1}\mathcal{R}(v) \in (T_pN)^\perp$ for all $v \in T_pN$ and $p \in N$. Given $v \in TN$, let

$c : I \subset \mathbb{R} \to N$ be the motion of the mechanical system $(N, \langle\langle \cdot, \cdot \rangle\rangle, \mathcal{F}_N)$ with initial condition v. Then

$$\begin{aligned}\widetilde{\nabla}_{\dot{c}}\, \dot{c} &= \nabla_{\dot{c}}\, \dot{c} + B(\dot{c}, \dot{c}) = \mu^{-1}\mathcal{F}_N + (\mu^{-1}\mathcal{F})^{\perp} + \mu^{-1}\mathcal{R} \\ &= (\mu^{-1}\mathcal{F})^{\top} + (\mu^{-1}\mathcal{F})^{\perp} + \mu^{-1}\mathcal{R} = \mu^{-1}\mathcal{F} + \mu^{-1}\mathcal{R}.\end{aligned}$$

□

EXAMPLE 2.9. To write the equation of motion of a simple pendulum with a perfect reaction force, we parametrize the holonomic constraint N using the map $\varphi : (-\pi, \pi) \to \mathbb{R}^2$ defined by

$$\varphi(\theta) = (l \sin \theta, -l \cos \theta)$$

(so that $\theta = 0$ labels the stable equilibrium position). We have

$$\frac{d}{d\theta} = \frac{dx}{d\theta}\frac{\partial}{\partial x} + \frac{dy}{d\theta}\frac{\partial}{\partial y} = l \cos \theta \frac{\partial}{\partial x} + l \sin \theta \frac{\partial}{\partial y},$$

and hence the kinetic energy of the pendulum is

$$\begin{aligned}K\left(v\frac{d}{d\theta}\right) &= \frac{1}{2}m\left\langle vl\cos\theta\frac{\partial}{\partial x} + vl\sin\theta\frac{\partial}{\partial y}, vl\cos\theta\frac{\partial}{\partial x} + vl\sin\theta\frac{\partial}{\partial y}\right\rangle \\ &= \frac{1}{2}ml^2v^2.\end{aligned}$$

On the other hand, the potential energy is written

$$U(x, y) = -mgy,$$

and hence its constraint to N has the local expression

$$U(\theta) = -mgl \cos \theta.$$

Consequently the equation of motion is

$$\begin{aligned}&\frac{d}{dt}\left(\frac{\partial K}{\partial v}\left(\theta, \dot{\theta}\right)\right) - \frac{\partial K}{\partial \theta}\left(\theta, \dot{\theta}\right) = -\frac{\partial U}{\partial \theta}(\theta) \\ &\Leftrightarrow \frac{d}{dt}\left(ml^2\dot{\theta}\right) = -mgl \sin \theta \\ &\Leftrightarrow \ddot{\theta} = -\frac{g}{l}\sin \theta.\end{aligned}$$

EXERCISES 2.10.

(1) Use spherical coordinates to write the equations of motion for the **spherical pendulum** of length l, i.e. a particle of mass $m > 0$ moving in $\mathbb{R}^3$ subject to a constant gravitational acceleration g and the the holonomic constraint

$$N = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = l^2\}.$$

Which parallels of N are possible trajectories of the particle?

(2) Write the equations of motion for a particle moving on a frictionless surface of revolution with equation $z = f(r)$ (where $r = \sqrt{x^2 + y^2}$) under a constant gravitational acceleration g.

(3) Write and solve the equations of motion for a free dumbell, i.e. a system of two particles of masses m_1 and m_2 connected by a massless rod of length l, moving in:
   (a) $\mathbb{R}^2$;
   (b) $\mathbb{R}^3$.

   (**Hint:** Use the coordinates of the **center of mass**, i.e. the point along the rod at a distance $\frac{m_2}{m_1+m_2}l$ from m_1).

(4) The **double pendulum** of lengths l_1, l_2 is the mechanical system defined by two particles moving in $\mathbb{R}^2$ subject to a constant gravitational acceleration g and the the holonomic constraint

$$N = \{(x_1, x_2) \in \mathbb{R}^4 \mid \|x_1\| = l_1 \text{ and } \|x_1 - x_2\| = l_2\}.$$

   (diffeomorphic to the 2-torus T^2).
   (a) Write the equations of motion for the double pendulum using the parametrization $\phi : (-\pi, \pi) \times (-\pi, \pi) \to N$ given by

$$\phi(\theta, \varphi) = (l_1 \sin\theta, -l_1 \cos\theta, l_1 \sin\theta + l_2 \sin\varphi, -l_1 \cos\theta - l_2 \cos\varphi).$$

   (b) Linearize the equations of motion around $\theta = \varphi = 0$. Look for solutions of the linearized equations satisfying $\theta = k\varphi$, with $k \in \mathbb{R}$ constant (**normal modes**). What are the periods of the ensuing oscillations?

3. Rigid Body

Recall that a rigid body is a system of k particles of masses $m_1, \ldots, m_k$ connected by massless rods in such a way that their mutual distances remain constant. If in addition we assume that a given particle is fixed (at the origin, say) then we obtain the holonomic constraint

$$N = \left\{(x_1, \ldots, x_k) \in \mathbb{R}^{3k} \mid x_1 = 0 \text{ and } \|x_i - x_j\| = d_{ij} \text{ for } 1 \leq i < j \leq k\right\}.$$

If at least three particles are not collinear, this manifold is diffeomorphic to $O(3)$. In fact, if $(\xi_1, \ldots, \xi_k)$ is a point in N then any other point in N is of the form $(S\xi_1, \ldots, S\xi_k)$ for a unique $S \in O(3)$. A motion in N can therefore be specified by a curve $S : I \subset \mathbb{R} \to O(3)$. The trajectory in $\mathbb{R}^3$ of the particle with mass m_i will be given by the curve $S\xi_i : I \subset \mathbb{R} \to \mathbb{R}^3$, whose velocity is $\dot{S}\xi_i$ (where we use $O(3) \subset \mathcal{M}_{3\times3}(\mathbb{R}) \cong \mathbb{R}^9$ to identify $T_SO(3)$ with an appropriate subspace of $\mathcal{M}_{3\times3}(\mathbb{R})$). Therefore the kinetic energy of the system along the motion will be

$$K = \frac{1}{2}\sum_{i=1}^{n} m_i \langle \dot{S}\xi_i, \dot{S}\xi_i \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product on $\mathbb{R}^3$.

Now $O(3)$, and hence N, has two diffeomorphic connected components. Since any motion must necessarily occurs in one connected component, we can take our configuration space to be simply $SO(3)$. To account for continuum rigid bodies, we make the following generalization:

DEFINITION 3.1. *A* **rigid body with a fixed point** *is any mechanical system of the form* $(SO(3), \langle\langle\cdot,\cdot\rangle\rangle, \mathcal{F})$, *with*

$$\langle\langle V, W\rangle\rangle = \int_{\mathbb{R}^3} \langle V\xi, W\xi\rangle \, dm$$

for all $V, W \in T_S SO(3)$ *and all* $S \in SO(3)$, *where* $\langle\cdot,\cdot\rangle$ *is the usual Euclidean inner product on* $\mathbb{R}^3$ *and* m *(called the* **mass distribution of the reference configuration***) is a positive finite measure on* $\mathbb{R}^3$ *not supported on any straight line through the origin and satisfying* $\int_{\mathbb{R}^3} \|\xi\|^2 dm < +\infty$.

EXAMPLE 3.2.

(1) The rigid body composed by k particles of masses $m_1, \ldots, m_k$ corresponds to the measure

$$m = \sum_{i=1}^{k} m_i \delta_{\xi_i},$$

where δ_{ξ_i} is the Dirac delta centered at the point $\xi_i \in \mathbb{R}^3$.

(2) A continuum rigid body with (say) compactly supported integrable density function $\rho : \mathbb{R}^3 \to [0, +\infty)$ is described by the measure m defined on the Lebesgue sigma-algebra by

$$m(A) = \int_A \rho(\xi) d^3\xi.$$

REMARK 3.3. The rotational motion of a general rigid body can in many cases be reduced to the motion of a rigid body with a fixed point (cf. Exercise 3.20.2). Unless otherwise stated, from this point onwards we will take "rigid body" to mean "rigid body with a fixed point".

PROPOSITION 3.4. *The metric* $\langle\langle\cdot,\cdot\rangle\rangle$ *defined on* $SO(3)$ *by a rigid body is left-invariant.*

PROOF. Since left multiplication by a fixed matrix $R \in SO(3)$ is a linear map $L_R : \mathcal{M}_{3\times3}(\mathbb{R}) \to \mathcal{M}_{3\times3}(\mathbb{R})$, we have $(dL_R)_S V = RV \in T_{RS}SO(3)$ for any $V \in T_S SO(3)$. Consequently,

$$\begin{aligned}\langle\langle (dL_R)_S V, (dL_R)_S W\rangle\rangle = \langle\langle RV, RW\rangle\rangle &= \int_{\mathbb{R}^3} \langle RV\xi, RW\xi\rangle \, dm \\ &= \int_{\mathbb{R}^3} \langle V\xi, W\xi\rangle \, dm = \langle\langle V, W\rangle\rangle\end{aligned}$$

(as $R \in SO(3)$ preserves the Euclidean inner product). □

Therefore there exist at most as many rigid bodies as inner products on $\mathfrak{so}(3) \cong \mathbb{R}^3$, i.e., as real symmetric positive definite 3×3 matrices. In fact, we shall see that any rigid body can be specified by 3 positive numbers.

PROPOSITION 3.5. *The metric* $\langle\langle\cdot,\cdot\rangle\rangle$ *defined on* $SO(3)$ *by a rigid body is given by*

$$\langle\langle V, W\rangle\rangle = \operatorname{tr}\left(VJW^t\right),$$

where

$$J_{ij} = \int_{\mathbb{R}^3} \xi^i \xi^j dm.$$

PROOF. We just have to notice that

$$\begin{aligned}\langle\langle V, W\rangle\rangle &= \int_{\mathbb{R}^3} \sum_{i=1}^{3} \left(\sum_{j=1}^{3} V_{ij}\xi^j\right)\left(\sum_{k=1}^{3} W_{ik}\xi^k\right) dm \\ &= \sum_{i,j,k=1}^{3} V_{ij}W_{ik} \int_{\mathbb{R}^3} \xi^j \xi^k \, dm = \sum_{i,j,k=1}^{3} V_{ij}J_{jk}W_{ik}.\end{aligned}$$

□

PROPOSITION 3.6. *If* $S : I \subset \mathbb{R} \to SO(3)$ *is a curve and* ∇ *is the Levi-Civita connection on* $(SO(3), \langle\langle\cdot,\cdot\rangle\rangle)$ *then*

$$\langle\langle \nabla_{\dot{S}}\dot{S}, V\rangle\rangle = \int_{\mathbb{R}^3} \langle \ddot{S}\xi, V\xi\rangle dm$$

for any $V \in T_S SO(3)$.

PROOF. We consider first the case in which the rigid body is **non planar**, i.e. m is not supported in any plane through the origin. In this case, the metric $\langle\langle\cdot,\cdot\rangle\rangle$ can be extended to a flat metric on $\mathcal{M}_{3\times3}(\mathbb{R}) \cong \mathbb{R}^9$ by the same formula

$$\langle\langle\langle V, W\rangle\rangle\rangle = \int_{\mathbb{R}^3} \langle V\xi, W\xi\rangle \, dm$$

for all $V, W \in T_S\mathcal{M}_{3\times3}(\mathbb{R})$ and all $S \in \mathcal{M}_{3\times3}(\mathbb{R})$. Indeed, this formula clearly defines a symmetric 2-tensor on $\mathcal{M}_{3\times3}(\mathbb{R})$. To check positive definiteness, we notice that if $V \in T_S\mathcal{M}_{3\times3}(\mathbb{R})$ in nonzero then its kernel is contained on a plane through the origin. Therefore, the continuous function $\langle V\xi, V\xi\rangle$ is positive on a set of positive measure, and hence

$$\langle\langle\langle V, V\rangle\rangle\rangle = \int_{\mathbb{R}^3} \langle V\xi, V\xi\rangle \, dm > 0.$$

This metric is easily seen to be flat, as the components of the metric on the natural coordinates of $\mathcal{M}_{3\times3}(\mathbb{R})$ are the constant coefficients J_{ij}. Therefore all Christoffel symbols vanish on these coordinates, and the corresponding Levi-Civita connection $\widetilde{\nabla}$ is the trivial connection. If $S : I \subset \mathbb{R} \to \mathcal{M}_{3\times3}(\mathbb{R})$ is a curve then

$$\widetilde{\nabla}_{\dot{S}}\dot{S} = \ddot{S}.$$

Since $\langle\langle\cdot,\cdot\rangle\rangle$ is the metric induced on $SO(3)$ by $\langle\langle\langle\cdot,\cdot\rangle\rangle\rangle$, we see that for any curve $S : I \subset \mathbb{R} \to SO(3)$ one has

$$\nabla_{\dot{S}}\dot{S} = \left(\widetilde{\nabla}_{\dot{S}}\dot{S}\right)^\top = \ddot{S}^\top,$$

and hence

$$\langle\langle \nabla_{\dot S}\dot S, V\rangle\rangle = \langle\langle \ddot S^{\top}, V\rangle\rangle = \langle\langle\langle \ddot S^{\top}, V\rangle\rangle\rangle = \int_{\mathbb{R}^3} \langle \ddot S\xi, V\xi\rangle dm$$

for any $V \in T_S SO(3)$.

For planar rigid bodies the formula can by obtained by a limiting procedure (cf. Exercise 3.20.3). □

We can use this result to determine the geodesics of $(SO(3), \langle\langle\cdot,\cdot\rangle\rangle)$. A remarkable shortcut (whose precise nature will be discussed in Section 5) can be obtained by introducing the following quantity:

DEFINITION 3.7. *The* **angular moment** *of a rigid body whose motion is described by $S : I \subset \mathbb{R} \to SO(3)$ is the vector*

$$p(t) = \int_{\mathbb{R}^3} \left[(S(t)\xi) \times (\dot S(t)\xi) \right] dm$$

(where $\times$ is the usual cross product on $\mathbb{R}^3$).

THEOREM 3.8. *If $S : I \subset \mathbb{R} \to SO(3)$ is a geodesic of $(SO(3), \langle\langle\cdot,\cdot\rangle\rangle)$ then $p(t)$ is constant.*

PROOF. We have

$$\dot p = \int_{\mathbb{R}^3} \left[(\dot S\xi) \times (\dot S\xi) + (S\xi) \times (\ddot S\xi) \right] dm = \int_{\mathbb{R}^3} \left[(S\xi) \times (\ddot S\xi) \right] dm.$$

Take any $v \in \mathbb{R}^3$. Then

$$\begin{aligned}\langle Sv, \dot p\rangle &= \int_{\mathbb{R}^3} \left\langle Sv, (S\xi) \times (\ddot S\xi) \right\rangle dm = \int_{\mathbb{R}^3} \left\langle \ddot S\xi, (Sv) \times (S\xi) \right\rangle dm \\ &= \int_{\mathbb{R}^3} \left\langle \ddot S\xi, S(v \times \xi) \right\rangle dm,\end{aligned}$$

where we have used the invariance of $\langle\cdot, \cdot\times\cdot\rangle \equiv \det(\cdot,\cdot,\cdot)$ under even permutations its arguments.

To complete the proof we will need the following lemma, whose proof is left as an exercise:

LEMMA 3.9. *There exists a linear isomorphism $\Omega : \mathfrak{so}(3) \to \mathbb{R}^3$ such that*

$$A\xi = \Omega(A) \times \xi$$

for all $\xi \in \mathbb{R}^3$ and $A \in \mathfrak{so}(3)$.

Let $V \in \mathfrak{so}(3)$ be such that $\Omega(V) = v$. Then $SV \in T_S SO(3)$ and

$$\langle Sv, \dot p\rangle = \int_{\mathbb{R}^3} \left\langle \ddot S\xi, SV\xi \right\rangle dm = \langle\langle \nabla_{\dot S}\dot S, SV\rangle\rangle = 0$$

(as $S : I \subset \mathbb{R} \to SO(3)$ is a geodesic). Since $v \in \mathbb{R}^3$ is arbitrary, we see that $\dot p = 0$ along the motion. □

If $S : I \subset \mathbb{R} \to SO(3)$ is a curve then $\dot{S} = SA$ for some $A \in \mathfrak{so}(3)$. Let us define $\Omega := \Omega(A)$. Since multiplication by $S \in SO(3)$ preserves the cross product, we have

$$
\begin{aligned}
p &= \int_{\mathbb{R}^3} [(S\xi) \times (SA\xi)]\, dm = \int_{\mathbb{R}^3} S\, [\xi \times (A\xi)]\, dm \\
&= S \int_{\mathbb{R}^3} [\xi \times (\Omega \times \xi)]\, dm.
\end{aligned}
$$

This suggests the following definition.

Definition 3.10. *The linear operator $I : \mathbb{R}^3 \to \mathbb{R}^3$ defined as*

$$
I(v) = \int_{\mathbb{R}^3} [\xi \times (v \times \xi)] dm
$$

is called the rigid body's **moment of inertia tensor**.

Proposition 3.11. *The moment of inertia tensor of any given rigid body is a symmetric positive definite linear operator, and the corresponding kinetic energy map $K : TSO(3) \to \mathbb{R}$ is given by*

$$
K(V) = \frac{1}{2}\langle\langle V, V\rangle\rangle = \frac{1}{2}\langle\langle SA, SA\rangle\rangle = \frac{1}{2}\langle I\Omega, \Omega\rangle,
$$

for all $V \in T_S SO(3)$ and all $S \in SO(3)$, where $V = SA$ and $\Omega = \Omega(A)$.

Proof. We start by checking that I is symmetric:

$$
\begin{aligned}
\langle Iv, w\rangle &= \langle \int_{\mathbb{R}^3} [\xi \times (v \times \xi)]\, dm, w\rangle = \int_{\mathbb{R}^3} \langle \xi \times (v \times \xi), w\rangle\, dm \\
&= \int_{\mathbb{R}^3} \langle v \times \xi, w \times \xi\rangle\, dm = \langle v, Iw\rangle.
\end{aligned}
$$

In particular we have

$$
\begin{aligned}
\langle I\Omega, \Omega\rangle &= \int_{\mathbb{R}^3} \langle \Omega \times \xi, \Omega \times \xi\rangle\, dm = \int_{\mathbb{R}^3} \langle A\xi, A\xi\rangle\, dm \\
&= \int_{\mathbb{R}^3} \langle SA\xi, SA\xi\rangle\, dm = 2K(V).
\end{aligned}
$$

The positive definiteness of I is an immediate consequence of this formula. □

Corollary 3.12. *Given any rigid body there exist positive numbers I_1, I_2, I_3 (***principal moments of inertia***) and an ortonormal basis of $\mathbb{R}^3$ $\{e_1, e_2, e_3\}$ (***principal axes***) such that $Ie_i = I_i e_i$ $(i = 1, 2, 3)$.*

The principal moments of inertia are the three positive numbers which completely specify the rigid body (as they determine the inertia tensor, which in turn yields the kinetic energy). To compute these numbers we must compute the eigenvalues of a matrix representation of the inertia tensor.

PROPOSITION 3.13. *The matrix representation of the inertia tensor in the canonical basis of* $\mathbb{R}^3$ *is*

$$\begin{pmatrix} \int_{\mathbb{R}^3}(y^2+z^2)\,dm & -\int_{\mathbb{R}^3} xy\,dm & -\int_{\mathbb{R}^3} xz\,dm \\ -\int_{\mathbb{R}^3} xy\,dm & \int_{\mathbb{R}^3}(x^2+z^2)\,dm & -\int_{\mathbb{R}^3} yz\,dm \\ -\int_{\mathbb{R}^3} xz\,dm & -\int_{\mathbb{R}^3} yz\,dm & \int_{\mathbb{R}^3}(x^2+y^2)\,dm \end{pmatrix}$$

PROOF. Let $\{u_1, u_2, u_3\}$ be the canonical basis of $\mathbb{R}^3$. Then

$$I_{ij} = \langle Iu_i, u_j\rangle = \int_{\mathbb{R}^3} \langle \xi \times (u_i \times \xi), u_j\rangle\, dm.$$

Using the vector identity

$$u \times (v \times w) = \langle u, w\rangle v - \langle u, v\rangle w$$

for all $u, v, w \in \mathbb{R}^3$, we have

$$I_{ij} = \int_{\mathbb{R}^3} \left\langle \|\xi\|^2 u_i - \langle \xi, u_i\rangle \xi, u_j \right\rangle\, dm = \int_{\mathbb{R}^3} \left(\|\xi\|^2 \delta_{ij} - \xi^i \xi^j \right)\, dm.$$

□

PROPOSITION 3.14. *The equations of motion of a rigid body in the absence of external forces are given by the* **Euler equations**

$$I\dot{\Omega} = (I\Omega) \times \Omega.$$

PROOF. We just have to notice that

$$p = SI\Omega.$$

Therefore

$$0 = \dot{p} = \dot{S}I\Omega + SI\dot{\Omega} = SAI\Omega + SI\dot{\Omega} = S\left(\Omega \times (I\Omega) + I\dot{\Omega}\right).$$

□

REMARK 3.15. Any point $\xi \in \mathbb{R}^3$ in the rigid body traverses a curve $x(t) = S(t)\xi$ with velocity

$$\dot{x} = \dot{S}\xi = SA\xi = S(\Omega \times \xi) = (S\Omega) \times (S\xi).$$

Therefore $\omega = S\Omega$ is the rigid body's *instantaneous angular velocity*: at each instant, the rigid body rotates about the axis determined by ω with angular speed $\|\omega\|$. Consequently, Ω is the angular speed as seen in the (accelerated) rigid body's rest frame.

In the basis $\{e_1, e_2, e_3\}$ of the principal axes, the Euler equations are written

$$\begin{cases} I_1\dot{\Omega}^1 = (I_2 - I_3)\Omega^2\Omega^3 \\ I_2\dot{\Omega}^2 = (I_3 - I_1)\Omega^3\Omega^1 \\ I_3\dot{\Omega}^3 = (I_1 - I_2)\Omega^1\Omega^2 \end{cases}$$

Since I is positive definite (hence invertible), we can change variables to $P = I\Omega$. Notice that $p = SP$, i.e. P is the (constant) angular momentum vector as seen in rigid body's rest frame. In these new variables, the Euler equations are written

$$\dot{P} = P \times \left(I^{-1}P\right).$$

In the basis $\{e_1, e_2, e_3\}$ of the principal axes, these are

$$\begin{cases} \dot{P}^1 = \left(\dfrac{1}{I_3} - \dfrac{1}{I_2}\right) P^2 P^3 \\ \dot{P}^2 = \left(\dfrac{1}{I_1} - \dfrac{1}{I_3}\right) P^3 P^1 \\ \dot{P}^3 = \left(\dfrac{1}{I_2} - \dfrac{1}{I_1}\right) P^1 P^2 \end{cases}$$

Proposition 3.16. *If $I_1 > I_2 > I_3$, the stationary points of the Euler equations are given by*

$$P = \lambda e_i \qquad (i = 1, 2, 3),$$

and are stable for $i = 1, 3$ and unstable for $i = 2$.

Proof. Since there are no external forces, the kinetic energy is conserved:

$$2K = \langle I\Omega, \Omega \rangle = \left\langle P, I^{-1}P \right\rangle = \frac{\left(P^1\right)^2}{I_1} + \frac{\left(P^2\right)^2}{I_2} + \frac{\left(P^3\right)^2}{I_3}.$$

This means that the flow defined by the Euler equations is along ellipsoids with semiaxes $\sqrt{\frac{I_1}{2K}} > \sqrt{\frac{I_2}{2K}} > \sqrt{\frac{I_3}{2K}}$. On the other hand, since p is constant along the motion, we have a second conserved quantity:

$$\|p\|^2 = \|P\|^2 = \left(P^1\right)^2 + \left(P^2\right)^2 + \left(P^3\right)^2.$$

Therefore the flow is along spheres. The integral curves on a particular sphere can be found by intersecting it with the ellipsoids correspondig to different values of K, as shown in Figure 2. □

Remark 3.17. Since $\Omega = I^{-1}P$, Proposition 3.16 is still true if we replace P with Ω. The equilibrium points represent rotations about the principal axes with constant angular speed, as they satisfy $\Omega = I_i P$, and hence $\omega = I_i p$ is constant. If the rigid body is placed in a rotation state close to a rotation about the axes e_1 or e_3, P will remain close to these axes, and hence Se_1 or Se_3 will remain close to the fixed vector p. On the other hand, if the rigid body is placed in a rotation state close to a rotation about the axis e_2, then P will drift away from e_2 (approaching $-e_2$ before returning to e_2), and hence Se_2 will drift away from the fixed vector p (approaching $-p$ before returning to p). This can be illustrated by throwing a rigid body (say a brick) in the air, as its rotational motion about the center of mass is that of a rigid body with a fixed point (cf. Exercise 3.20.2). When rotating

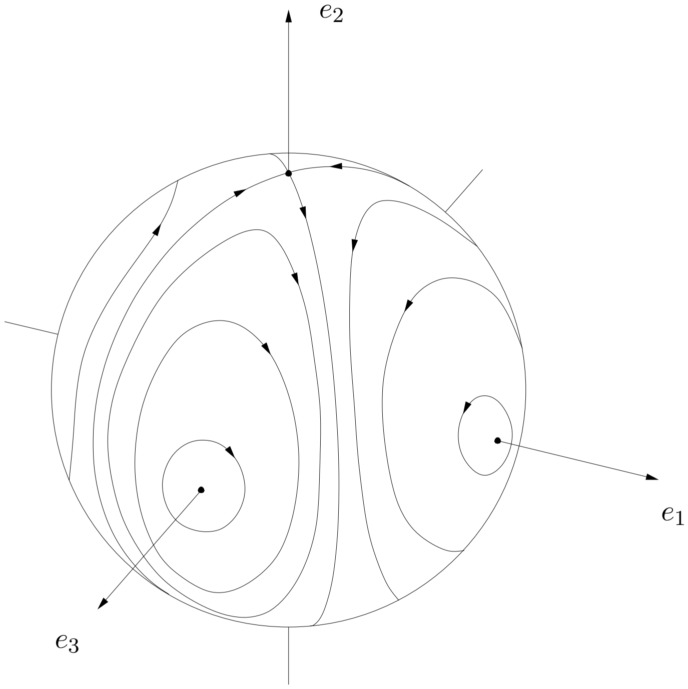

FIGURE 1. Integral curves of the Euler equations.

about the smaller or the larger axis (cf. Exercise 3.20.6) it performs a stable rotation, but when rotating about the middle axis it flips in midair.

If the rigid body is not free, one must use parametrizations of $SO(3)$.

DEFINITION 3.18. *The* **Euler angles** *correspond to the local coordinates* $(\theta, \varphi, \psi) : SO(3) \to (0, \pi) \times (0, 2\pi) \times (0, 2\pi)$ *defined by*

$$S(\theta, \varphi, \psi) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The geometric interpretation of the Euler angles is sketched in Figure 2: if the rotation carries the canonical basis $\{e_x, e_y, e_z\}$ to a new orthonormal basis $\{e_1, e_2, e_3\}$, then θ is the angle between e_3 and e_z, φ is the angle between the line of intersection of the planes spanned by $\{e_1, e_2\}$ and $\{e_x, e_y\}$ (called the **nodal line**) and the x-axis, and ψ is the angle between e_1 and the nodal line.

The general expression of the kinetic energy in the local coordinates of $TSO(3)$ associated to the Euler angles is quite complicated; here we present it only in the simpler case $I_1 = I_2$.

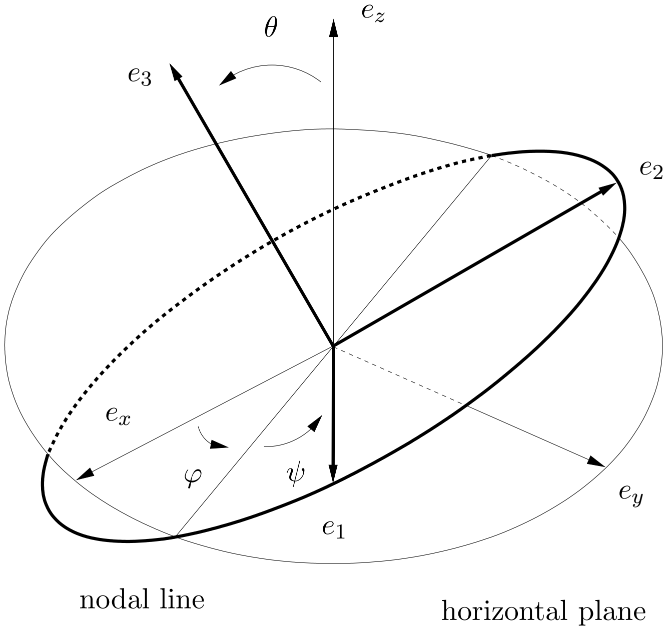

FIGURE 2. Euler angles.

PROPOSITION 3.19. *If $I_1 = I_2$ then the kinetic energy of a rigid body in the local coordinates $(\theta, \varphi, \psi, v^\theta, v^\varphi, v^\psi)$ of $TSO(3)$ is given by*

$$K = \frac{I_1}{2}\left(\left(v^\theta\right)^2 + (v^\varphi)^2 \sin^2\theta\right) + \frac{I_3}{2}\left(v^\psi + v^\varphi \cos\theta\right)^2.$$

A famous model which can be studied using this expression is the so-called **Lagrange top**, corresponding to a symmetric rigid body in a constant gravity field g. The potential energy for the corresponding mechanical system is

$$U = g\int_{\mathbb{R}^3} \langle S\xi, e_z\rangle \, dm = Mg\langle S\overline{\xi}, e_z\rangle,$$

where $M = m(\mathbb{R}^3)$ is the total mass and

$$\overline{\xi} = \frac{1}{M}\int_{\mathbb{R}^3} \xi \, dm$$

is the position of the **center of mass** in the rigid body's frame. If the center of mass satisfies $\overline{\xi} = le_3$ then

$$U = Mgl\cos\theta.$$

EXERCISES 3.20.

(1) Show that the bilinear form $\langle\langle\cdot,\cdot\rangle\rangle$ defined on $SO(3)$ by a rigid body is indeed a Riemannian metric.

(2) A general rigid body (i.e. with no fixed points) is any mechanical system of the form $(\mathbb{R}^3 \times SO(3), \langle\langle\langle\cdot,\cdot\rangle\rangle\rangle, \mathcal{F})$, with
$$\langle\langle\langle(v,V),(w,W)\rangle\rangle\rangle = \int_{\mathbb{R}^3} \langle v + V\xi, w + W\xi\rangle \, dm$$
for all $(v,V),(w,W) \in T_{(x,S)}\mathbb{R}^3 \times SO(3)$ and $(x,S) \in \mathbb{R}^3 \times SO(3)$, where $\langle\cdot,\cdot\rangle$ is the usual Euclidean inner product on $\mathbb{R}^3$ and m is a positive finite measure on $\mathbb{R}^3$ not supported on any straight line and satisfying $\int_{\mathbb{R}^3} \|\xi\|^2 dm < +\infty$.
  (a) Show that one can always translate m in such a way that
$$\int_{\mathbb{R}^3} \xi \, dm = 0$$
  (i.e. the center of mass of the reference configuration is placed at the origin).
  (b) Show that for this choice the kinetic energy of the rigid body is
$$K(v,V) = \frac{1}{2}M\langle v,v\rangle + \frac{1}{2}\langle\langle V,V\rangle\rangle,$$
  where $M = m(\mathbb{R}^3)$ is the total mass of the rigid body and $\langle\langle\cdot,\cdot\rangle\rangle$ is the metric for the rigid body with a fixed point determined by m.
  (c) Assume that there exists a differentiable function $F : \mathbb{R}^3 \to \mathbb{R}^3$ such that
$$\mathcal{F}(x,S,v,V)(w,W) = \int_{\mathbb{R}^3} \langle F(x+S\xi), w + W\xi\rangle \, dm.$$
  Show that if
$$\int_{\mathbb{R}^3} (S\xi) \times F(x+S\xi) \, dm = 0$$
  for all $(x,S) \in \mathbb{R}^3 \times SO(3)$ then the projection of any motion on $SO(3)$ is a geodesic of $(SO(3), \langle\langle\cdot,\cdot\rangle\rangle)$.
  (d) Describe the motion of a rigid body falling in a constant gravitational field, for which $F = (0,0,-g)$ is constant.

(3) Prove Proposition 3.6 for a planar rigid body. (**Hint:** Include the planar rigid body in a smooth one-parameter family of non planar rigid bodies).

(4) Prove Lemma 3.9.

(5) Show that $I_1 \leq I_2 + I_3$ (and cyclic permutations). When is $I_1 = I_2 + I_3$?

(6) Determine the principal axes and the corresponding principal moments of inertia of:
  (a) a homogeneous rectangular parallelepiped with mass M, sides $2a, 2b, 2c \in \mathbb{R}^+$ and centered at the origin;
  (b) a homogeneous (solid) ellipsoid with mass M, semiaxes $a,b,c \in \mathbb{R}^+$ and centered at the origin. (**Hint:** Use the coordinate change $(x,y,z) = (au,bv,cw)$).

(7) A **symmetry** of a rigid body is an isometry $S \in O(3)$ which preserves the mass distribution (i.e. $m(SA) = m(A)$ for any measurable set $A \subset \mathbb{R}^3$). Show that:
   (a) $SIS^t = I$, where I is the matrix representation of the inertia tensor;
   (b) if S is a reflection in a plane then there exists a principal axis orthogonal to the reflection plane;
   (c) if S is a nontrivial rotation about an axis then that axis is principal;
   (d) if moreover the rotation is not by π then all axes orthogonal to the rotation axis are principal.

(8) Consider a rigid body satisfying $I_1 = I_2$. Use the Euler equations to show that:
   (a) the angular velocity satisfies
$$\dot{\omega} = \frac{1}{I_1} p \times \omega;$$
   (b) if $I_1 = I_2 = I_3$ then the rigid body rotates about a fixed axis with constant angular speed (i.e. ω is constant);
   (c) if $I_1 = I_2 \neq I_3$ then ω **precesses** (i.e. rotates) about p with angular velocity
$$\omega_{\text{pr}} = \frac{p}{I_1}.$$

(9) Many asteroids have irregular shapes, and hence satisfy $I_1 < I_2 < I_3$. To a very good approximation, their rotational motion about the center of mass is described by the Euler equations. Over very long periods of time, however, their small interactions with the Sun and other planetary bodies tend to decrease their kinetic energy while conserving their angular momentum. Which rotation state do asteroids approach?

(10) Due to its rotation, Earth is not a perfect sphere, but an oblate ellipsoid; therefore its moments of inertia are not quite equal, satisfying approximately
$$\begin{aligned} &I_1 = I_2 \neq I_3; \\ &\frac{I_3 - I_1}{I_1} \simeq \frac{1}{306}. \end{aligned}$$
Earth's rotation axis is very close to e_3, but precesses around it (**Chandler precession**). Find the period of this precession (in Earth's frame).

(11) Consider a rigid body whose motion is described by the curve $S : \mathbb{R} \to SO(3)$, and let Ω be the corresponding angular velocity. Consider a particle with mass m whose motion **in the rigid body's frame** is given by the curve $\xi : \mathbb{R} \to \mathbb{R}^3$. Let f be the

external force on the particle, so that its motion equation is

$$m\frac{d^2}{dt^2}(S\xi) = f.$$

(a) Show that the motion equation can be written as

$$m\ddot{\xi} = F - m\Omega \times (\Omega \times \xi) - 2m\Omega \times \dot{\xi} - m\dot{\Omega} \times \xi$$

where $f = SF$. (The terms following F are the so-called **inertial forces**, and are known, respectively, as the **centrifugal force**, the **Coriolis force** and the **Euler force**).

(b) Show that if the rigid body is a homogeneous sphere rotating freely (like Earth, for instance) then the Euler force vanishes. Why must a long range gun in the Northern hemisphere be aimed at the **left** of the target?

(c) Compute the force necessary to keep the particle motionless on the surface of a rigid body satisfying $I_1 = I_2 \neq I_3$ which rotates freely in space.

(12) Prove Proposition 3.19. (**Hint:** Notice that symmetry demands that the expression for K must not depend neither on φ nor on ψ).

(13) Consider the Lagrange top.

(a) Write the equations of motion and determine the equilibrium points.

(b) Show that there exist solutions such that θ, $\dot{\varphi}$ and $\dot{\psi}$ are constant, which in the limit $|\dot{\varphi}| << |\dot{\psi}|$ (**fast top**) satisfy

$$\dot{\varphi} \simeq \frac{Mgl}{I_3\dot{\psi}}.$$

(14) (*Precession of the equinoxes*) Due to its rotation, Earth is not a perfect sphere, but an oblate ellipsoid; therefore its moments of inertia are not quite equal, satisfying approximately

$$I_1 = I_2 \neq I_3;$$
$$\frac{I_3 - I_1}{I_1} \simeq \frac{1}{306}$$

(cf. Exercise 10). As a consequence, the combined gravitational attraction of the Moon and the Sun disturbs the Earth's rotation motion. This perturbation can be modelled by the potential energy $U : SO(3) \to \mathbb{R}$ given in the Euler angles (θ, φ, ψ) by

$$U = -\frac{\Omega^2}{2}(I_3 - I_1)\cos^2\theta,$$

where $\dfrac{2\pi}{\Omega} \simeq 168$ days.

(a) Write the equations of motion and determine the equilibrium points.

(b) Show that there exist solutions such that θ, $\dot{\varphi}$ and $\dot{\psi}$ are constant, which in the limit $|\dot{\varphi}| << |\dot{\psi}|$ (as is the case with Earth) satisfy

$$\dot{\varphi} \simeq -\frac{(I_3 - I_1)\Omega^2 \cos\theta}{I_3 \dot{\psi}}.$$

Given that for Earth $\theta \simeq 23°$, determine the approximate value of the period of $\varphi(t)$.

(15) (*Pseudo-rigid body*) Recall that the (non planar) rigid body metric is the constraint to $SO(3)$ of the flat metric on $GL(3)$ given by

$$\langle\langle V, W \rangle\rangle = \operatorname{tr}(VJW^t),$$

where

$$J_{ij} = \int_{\mathbb{R}^3} \xi^i \xi^j \, dm.$$

(a) What are the geodesics of the Levi-Civita connection for this metric? Is $(GL(3), \langle\langle \cdot, \cdot \rangle\rangle)$ geodesically complete?

(b) The **Euler equation** and the **continuity equation** for an incompressible fluid with velocity field $u : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ and pressure $p : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$ are

$$\begin{aligned} &\frac{\partial u}{\partial t} + (u \cdot \nabla) u = -\nabla p; \\ &\nabla \cdot u = 0, \end{aligned}$$

where

$$\nabla = \left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right)$$

is the usual operator of vector calculus.

Given a geodesic $S : \mathbb{R} \to GL(3)$, we define

$$\begin{aligned} &x(t, \xi) = S(t)\xi \\ &u(t, x) = \dot{S}(t)\xi = \dot{S}(t) S^{-1}(t) x. \end{aligned}$$

Show that the velocity field u satisfies the Euler equation (with $p = 0$), but not the continuity equation.

(c) Let $f : GL(3) \to \mathbb{R}$ be given by $f(S) = \det S$. Show that

$$\frac{\partial f}{\partial S_{ij}} = \operatorname{cof}(S)_{ij}$$

(where $\operatorname{cof}(S)$ is the matrix of the cofactors of S), and consequentemently

$$\frac{df}{dt} = (\det S) \operatorname{tr}(\dot{S} S^{-1}).$$

Therefore if we impose the constraint $\det S(t) = 1$ then the continuity equation is satisfied.

(d) Show that the holonomic constraint $SL(3) \subset GL(3)$ satisfies D'Alembert's Principle if and only if

$$\begin{cases} \mu\left(\ddot{S}\right) = \lambda(t) df \\ \det S = 1 \end{cases}$$

Show that the motion equation can be rewritten as

$$\ddot{S} = \lambda \left(S^{-1}\right)^t J^{-1}.$$

(e) Show that the geodesics of $(SL(3), \langle\langle \cdot, \cdot \rangle\rangle)$ yield solutions of the Euler equation with

$$p = -\frac{\lambda}{2} x^t \left(S^{-1}\right)^t J^{-1} S^{-1} x$$

differentiable which also satisfy the continuity equation. (**Remark:** More generally, it is possible to interpret the Euler equation on an open set $U \subset \mathbb{R}^n$ as a mechanical system on the group of diffeomorphisms of U (which is an infinite dimensional Lie group); the continuity equation imposes the holonomic constraint corresponding to the subgroup of volume-preserving diffeomorphisms, and the pressure is the perfect reaction force associated to this constraint).

4. Non-Holonomic Constraints

Some mechanical systems are subject to constaints which force the motions to proceed in certain admissible directions. To handle such constraints we must first introduce the corresponding geometric concept.

Definition 4.1. *A* **distribution** Σ *of dimension m on a differentiable manifold M is a choice of an m-dimensional subspace $\Sigma_p \subset T_pM$ for each $p \in M$. The distribution is said to be* **differentiable** *if for all $p \in M$ there exists a neighborhood $U \ni p$ and vector fields $X_1, \ldots, X_m \in \mathfrak{X}(U)$ such that*

$$\Sigma_q = \operatorname{span}\left\{ (X_1)_q, \ldots, (X_m)_q \right\}$$

for all $q \in U$.

Equivalently, Σ is differentiable if for all $p \in M$ there exists a neighborhood $U \ni p$ and 1-forms $\omega^1, \ldots, \omega^{n-m} \in \Omega^1(U)$ such that

$$\Sigma_q = \ker\left(\omega^1\right)_q \cap \ldots \cap \ker\left(\omega^{n-m}\right)_q$$

for all $p \in U$ (cf. Exercise 4.15.1). We will assume from this point on that all distributions are differentiable.

Definition 4.2. *A* **non-holonomic constraint** *on a mechanical system $(M, \langle \cdot, \cdot \rangle, \mathcal{F})$ is a distribution Σ on M. A curve $c : I \subset \mathbb{R} \to M$ is said to be* **compatible** *with Σ if $\dot{c}(t) \in \Sigma_{c(t)}$ for all $t \in I$.*

Example 4.3.

(1) (*Wheel rolling without slipping)* Consider a vertical wheel of radius R rolling without slipping on a plane. Assuming that the motion takes place along a straight line, we can parametrize any position of the wheel by the position x of the contact point and the angle θ between a fixed radius of the wheel and the radius containing the contact point (cf. Figure 3); hence the configuration space is $\mathbb{R} \times S^1$.

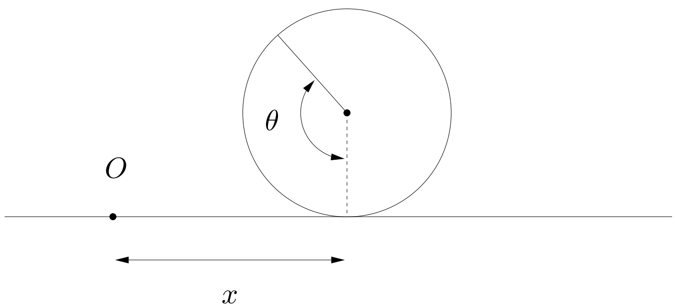

FIGURE 3. Wheel rolling without slipping.

If the wheel is to rotate without slipping, we must require that $\dot{x} = R\dot{\theta}$ along any motion; this is equivalent to requiring that the motion be compatible with the distribution defined on $\mathbb{R} \times S^1$ by the vector field

$$X = R\frac{\partial}{\partial x} + \frac{\partial}{\partial \theta},$$

or, equivalently, by the kernel of the 1-form

$$\omega = dx - Rd\theta.$$

(2) (*Ice skate*) A simple model for an ice skate is provided by a line segment which can either move along itself or rotate about its middle point. The position of the skate can be specified by the Cartesian coordinates (x, y) of the middle point and the angle θ between the skate and the x-axis (cf. Figure 4); hence the configuration space is $\mathbb{R}^2 \times S^1$.

If the skate can only move along itslef, we must require that $(\dot{x}, \dot{y})$ be proportional to $(\cos\theta, \sin\theta)$; this is equivalent to requiring that the motion be compatible with the distribution defined on $\mathbb{R}^2 \times S^1$ by the vector fields

$$X = \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y}, \qquad Y = \frac{\partial}{\partial \theta}$$

or, equivalently, by the kernel of the 1-form

$$\omega = -\sin\theta dx + \cos\theta dy.$$

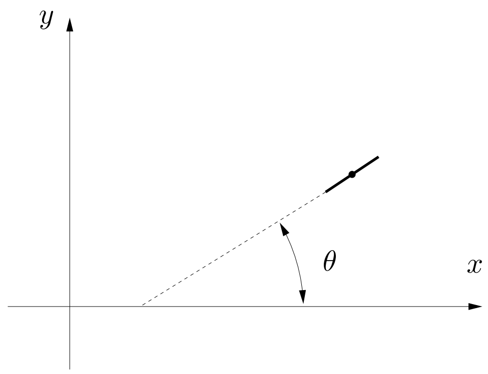

FIGURE 4. Ice skate.

One may wonder whether there exists any connection between holonomic and non-holonomic constraints. To answer this question, we must make a small digression.

DEFINITION 4.4. *A* **foliation** *of dimension m on a differentiable manifold M is a family $\mathcal{F} = \{L_\alpha\}_{\alpha\in A}$ of subsets of M (called* **leafs***) satisfying:*

(1) $M = \cup_{\alpha\in A} L_\alpha$;
(2) $L_\alpha \cap L_\beta = \varnothing$ *if* $\alpha \neq \beta$;
(3) each leaf L_α is **pathwise connected**, *i.e. if $p, q \in L_\alpha$ then there exists a continuous curve $c : [0,1] \to L_\alpha$ such that $c(0) = p$ and $c(1) = q$;*
(4) for each point $p \in M$ there exists an open set $U \ni p$ and local coordinates $(x^1, \dots, x^n) : U \to \mathbb{R}^n$ such that connected components of the intersections of the leafs with U are the level sets of $(x^{m+1}, \dots, x^n) : U \to \mathbb{R}^{n-m}$.

REMARK 4.5. The coordinates $(x^1, \dots, x^m)$ provide local coordinates on the leafs, which are therefore images of injective immersions. In particular, the leafs have well defined m-dimensional tangent spaces at each point, and consequently any foliation of dimension m defines an m-dimensional distribution.

DEFINITION 4.6. *An m-dimensional distribution Σ on a differential manifold M is said to be* **integrable** *if there exists an m-dimensional foliation $\mathcal{F} = \{L_\alpha\}_{\alpha\in A}$ on M such that*

$$\Sigma_p = T_p L_p$$

for all $p \in M$, where L_p is the leaf containing p. The leafs of $\mathcal{F}$ are called the **integral submanifolds** *of the distribution.*

Integrable distributions are particularly simple. For instance, the set of points $q \in M$ which are accessible from a given point $p \in M$ by a curve compatible with the distribution is simply the leaf L_p through p. If the leafs are **embedded** submanifolds, then an integrable non-holonomic restriction reduces to a family of holonomic restrictions. For this reason, an integrable distribution is sometimes called a **semi-holonomic constraint**, whereas a non-integrable distribution is called a **true non-holonomic constraint** .

It is therefore important to have a criterion for identifying integrable distributions.

DEFINITION 4.7. *Let* Σ *be a distribution on a differentiable manifold* M*. A vector field* $X \in \mathfrak{X}(M)$ *is said to be* **compatible** *with* Σ *if* $X_p \in \Sigma_p$ *for all* $p \in M$*. We denote by* $\mathfrak{X}(\Sigma)$ *the linear subspace of* $\mathfrak{X}(M)$ *formed by all vector fields which are compatible with* Σ*.*

THEOREM 4.8. (Frobenius) *A distribution* Σ *is integrable if and only if* $X, Y \in \mathfrak{X}(\Sigma) \Rightarrow [X, Y] \in \mathfrak{X}(\Sigma)$*.*

The proof of this theorem can be found in [**War83**] (see also Exercise 4.15.2). If Σ is locally given by m vector fields $X_1, \ldots, X_m$, then to check integrability it suffices to check if $[X_i, X_j] = \sum_{k=1}^{n} C_{ij}^k X_k$ for locally defined functions C_{ij}^k (cf. Exercise 4.15.3). The next proposition (whose proof is left as an exercise) provides an alternative criterion.

PROPOSITION 4.9. *An* m*-dimensional distribution* Σ *on an* n*-manifold* M *is integrable if and only if*

$$d\omega^i \wedge \omega^1 \wedge \ldots \wedge \omega^{n-m} = 0 \qquad (i = 1, \ldots, n-m)$$

for all locally defined sets of differential forms $\{\omega^1, \ldots, \omega^{n-m}\}$ *whose kernels determine* Σ*.*

This condition needs only be checked on an open cover of M.

EXAMPLE 4.10.

(1) (*Wheel rolling without slipping)* Recall that in this case the constraint is given by the kernel of the 1-form

$$\omega = dx - R d\theta.$$

Since $d\omega = 0$, we see that this is a semi-holonomic constraint, corresponding to an integrable distribution. The leafs of the distribution are the submanifolds with equation $x = x_0 + R\theta$.

(2) (*Ice skate*) Recall that in this case the constraint is given by the kernel of the 1-form

$$\omega = -\sin\theta dx + \cos\theta dy.$$

Since

$$\begin{aligned} d\omega \wedge \omega &= (-\cos\theta d\theta \wedge dx - \sin\theta d\theta \wedge dy) \wedge (-\sin\theta dx + \cos\theta dy) \\ &= -d\theta \wedge dx \wedge dy \neq 0, \end{aligned}$$

we see that this is a true non-holonomic constraint.

In a Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$, any distribution Σ determines an **orthogonal distribution** $\Sigma^\perp$, given by

$$\Sigma_p^\perp = (\Sigma_p)^\perp \subset T_pM.$$

Hence we have two orthogonal projections $^\top : TM \to \Sigma$ and $^\perp : TM \to \Sigma^\perp$. The set of external forces $\mathcal{F} : TM \to T^*M$ such that

$$\mathcal{F}(v) = \mathcal{F}\left(v^\top\right)$$

for all $v \in TM$ is denoted by F_Σ.

Definition 4.11. *A* **reaction force** *on a mechanical system with non-holonomic constraints $(M, \langle\cdot,\cdot\rangle, \mathcal{F}, \Sigma)$ is a force $\mathcal{R} \in F_\Sigma$ such that the solutions of the* **generalized Newton equation**

$$\mu\left(\frac{D\dot{c}}{dt}\right) = (\mathcal{F} + \mathcal{R})(\dot{c})$$

with initial condition in Σ are compatible with Σ. The reaction force is said to be **perfect**, *or to satisfy* **D'Alembert's principle**, *if*

$$\mu^{-1}\mathcal{R}(v) \in \Sigma_p^\perp$$

for all $v \in T_pM, p \in M$.

Just like in the holonomic case, a reaction force is perfect if and only if it does not creates nor dissipates energy along any motion compatible with the constraint.

Theorem 4.12. *Given a mechanical system with non-holonomic constraints $(M, \langle\cdot,\cdot\rangle, \mathcal{F}, \Sigma)$, there exists a unique reaction force $\mathcal{R} \in F_\Sigma$ satisfying D'Alembert's principle.*

Proof. We define the **second fundamental form** of the distribution Σ at a point $p \in M$ as the map $B : T_pM \times \Sigma_p \to \Sigma_p^\perp$ given by

$$B(v, w) = (\nabla_X Y)^\perp,$$

where $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(\Sigma)$ satisfy $X_p = v$ and $Y_p = w$. To check the validity of this definition, let $\{Z_1, \ldots, Z_n\}$ is a local orthonormal frame such that $\{Z_1, \ldots, Z_m\}$ is a basis for Σ and $\{Z_{m+1}, \ldots, Z_n\}$ is a basis for $\Sigma^\perp$. Then

$$\nabla_X Y = \nabla_X\left(\sum_{i=1}^m Y^i Z_i\right) = \sum_{i=1}^m\left((X\cdot Y^i)Z_i + \sum_{j,k=1}^n \Gamma_{ji}^k X^j Y^i Z_k\right),$$

where the functions Γ_{ij}^k are defined by

$$\nabla_{Z_i} Z_j = \sum_{k=1}^n \Gamma_{ij}^k Z_k,$$

and consequently

$$(\nabla_X Y)^\perp = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=m+1}^{n}\Gamma_{ij}^k X^i Y^j Z_k$$

depends indeed only on $v = X_p$ and $w = Y_p$. Moreover, we see that B is a bilinear map. Incidentally, the restriction of B to $\Sigma_p \times \Sigma_p$ is symmetric for all $p \in M$ if only

$$\Gamma_{ij}^k = \Gamma_{ji}^k \Leftrightarrow \langle \nabla_{Z_i} Z_j, Z_k\rangle = \langle \nabla_{Z_i} Z_j, Z_k\rangle \Leftrightarrow \langle [Z_i, Z_j], Z_k\rangle = 0$$

for all $i, j = 1, \ldots, m$ and all $k = m+1, \ldots, n$, i.e. if and only if Σ is integrable. In this case, B is, of course, the second fundamental form of the leaves.

Let us assume that $\mathcal{R}$ exists. Then any motion $c : I \subset \mathbb{R} \to M$ with initial condition on Σ is compatible with Σ and satisfies

$$\frac{D\dot{c}}{dt} = \mu^{-1}\mathcal{F} + \mu^{-1}\mathcal{R}.$$

The projection of this equation on $\Sigma^\perp$ yields

$$B(\dot{c}, \dot{c}) = \left(\mu^{-1}\mathcal{F}\right)^\perp + \mu^{-1}\mathcal{R}$$

(recall that $\frac{D\dot{c}}{dt} = \nabla_{\dot{c}}\dot{c}$). Therefore, if $\mathcal{R}$ exists then it must be given by

$$\mathcal{R}(v) = \mu\left(B(v, v)\right) - \mu\left(\left(\mu^{-1}\mathcal{F}\right)^\perp\right)$$

for any $v \in \Sigma$, and by $\mathcal{R}(v) = \mathcal{R}\left(v^\top\right)$ for any $v \in TM$ (as $\mathcal{R} \in F_\Sigma$). This proves unicity of $\mathcal{R}$.

To prove existence, we just have to show that for this choice of $\mathcal{R}$ the solutions of the generalized Newton equation with initial condition on Σ are compatible with Σ. Consider the system

$$\begin{cases} \dot{c} = \sum_{i=1}^m v^i Z_i \\ \frac{D\dot{c}}{dt} = \mu^{-1}\mathcal{F} - \left(\mu^{-1}\mathcal{F}\right)^\perp + B(\dot{c}, \dot{c}) \end{cases} \tag{24}$$

When written in local coordinates, this is a system of first order ODEs with $n+m$ unknowns $x^1(t), \ldots, x^n(t), v^1(t), \ldots, v^m(t)$. Since the second equation is just

$$\frac{D\dot{c}}{dt} = \left(\mu^{-1}\mathcal{F}\right)^\top + \left(\frac{D\dot{c}}{dt}\right)^\perp \Leftrightarrow \left(\frac{D\dot{c}}{dt}\right)^\top = \left(\mu^{-1}\mathcal{F}\right)^\top,$$

we see that this equation has only m nonvanishing components in the local frame $\{Z_1, \ldots, Z_n\}$. Therefore, (24) is a system of $(n+m)$ first order ODEs on $n+m$ unknowns, and has a unique local solution for any initial condition. If $\dot{c}(0) \in \Sigma_{c(0)}$, we can always choose $v^1(0), \ldots, v^m(0)$ such that

$$\dot{c}(0) = \sum_{i=1}^m v^i(0)\left(Z_i\right)_{c(0)}.$$

The solution of (24) with initial condition $(x^1(0),\dots,x^n(0),v^1(0),\dots,v^m(0))$ must then, by unicity, be the solution of

$$\frac{D\dot{c}}{dt} = \mu^{-1}\mathcal{F} + \mu^{-1}\mathcal{R}$$

with initial condition $\dot{c}(0)$. On the other hand, it is, by construction, compatible with Σ. $\square$

EXAMPLE 4.13. (*Wheel rolling without slipping)* Recall that in this case the constraint is given by the kernel of the 1-form

$$\omega = dx - Rd\theta.$$

Since $\mu^{-1}\mathcal{R}$ is orthogonal to the constraint for perfect reaction force $\mathcal{R}$, the constraint must be in the kernel of $\mathcal{R}$, and hence $\mathcal{R} = \lambda\omega$ for some smooth function $\lambda : TM \to \mathbb{R}$.

If kinetic energy of the wheel is

$$K = \frac{M}{2}\left(v^x\right)^2 + \frac{I}{2}\left(v^\theta\right)^2$$

then

$$\mu\left(\frac{D\dot{c}}{dt}\right) = M\ddot{x}dx + I\ddot{\theta}d\theta.$$

Just to make things more interesting, consider a constant gravitational acceleration g and suppose that the plane on which the wheel rolls is inclined by an angle α with respect to the horizontal, so that there exists a conservative force with potential energy

$$U = Mg\sin\alpha\, x.$$

The equation of motion is therefore

$$\mu\left(\frac{D\dot{c}}{dt}\right) = -dU + \mathcal{R} \Leftrightarrow M\ddot{x}dx + I\ddot{\theta}d\theta = -Mg\sin\alpha dx + \lambda dx - \lambda R d\theta.$$

The motion of the wheel will be given by a solution of this equation which also satisfies the constraint equation, i.e. a solution of the system of ODEs

$$\begin{cases} M\ddot{x} = -Mg\sin\alpha + \lambda \\ I\ddot{\theta} = -R\lambda \\ \dot{x} = R\dot{\theta} \end{cases}$$

Thys system is easily solved to yield

$$\begin{cases} x(t) = x_0 + v_0 t - \frac{\gamma}{2}t^2 \\ \theta(t) = \theta_0 + \frac{v_0}{R}t - \frac{\gamma}{2R}t^2 \\ \lambda = -\frac{I\gamma}{R^2} \end{cases}$$

where

$$\gamma = \frac{g\sin\alpha}{1 + \frac{I}{MR^2}}$$

and x_0, v_0, θ_0 are integration constants.

Physically, the reaction force must be interpreted as a friction force exerted by the plane on the wheel. This force opposes the translational motion of the wheel but accelerates its spinning motion. Therefore, contrary to intuition, there is no dissipation of energy: all the translational kinetic energy lost by the wheel is restored as rotational kinetic energy.

A perfect reaction force guarantees, as one would expect, conservation of energy.

THEOREM 4.14. *If in a conservative mechanical system with constraints* $(M, \langle\cdot,\cdot\rangle, -dU, \Sigma)$ *the reaction force satisfies D'Alembert's principle then the mechanical energy* $E_m = K + U$ *is constant along any motion with initial condition in* Σ.

EXERCISES 4.15.

(1) Show that an m-dimensional distribution Σ on an n-manifold M is differentiable if and only if for all $p \in M$ there exists a neighborhood $U \ni p$ and 1-forms $\omega^1, \ldots, \omega^{n-m} \in \Omega^1(U)$ such that
$$\Sigma_q = \ker\left(\omega^1\right)_q \cap \ldots \cap \ker\left(\omega^{n-m}\right)_q$$
for all $q \in U$.

(2) Let Σ be an integrable distribution. Shwo that $X, Y \in \mathfrak{X}(\Sigma) \Rightarrow [X, Y] \in \mathfrak{X}(\Sigma)$.

(3) Show that an m-dimensional distribution Σ is integrable if and only if each local basis of vector fields $\{X_1, \ldots, X_m\}$ satisfies $[X_i, X_j] = \sum_{k=1}^{n} C_{ij}^k X_k$ for locally defined functions C_{ij}^k. (**Remark:** Obviously this condition needs only be checked for an open cover).

(4) Recall that our model for an ice skate is given by the non-holonomic constraint Σ defined on $\mathbb{R}^2 \times S^1$ by the kernel of the 1-form $\omega = -\sin\theta dx + \cos\theta dy$.
   (a) Show that the ice skate can access all points in the configuration space: given two points $p, q \in \mathbb{R}^2 \times S^1$ there exists a curve $c : [0,1] \to \mathbb{R}^2 \times S^1$ compatible with Σ such that $c(0) = p$ and $c(1) = q$. Why does this show that Σ is non-integrable?
   (b) Assuming that the kinetic energy of the skate is
   $$K = \frac{M}{2}\left((v^x)^2 + (v^y)^2\right) + \frac{I}{2}\left(v^\theta\right)^2$$
   and that the reaction force is perfect, show that the skate moves with constant speed along straight lines or circles. What is the physical interpretation of the reaction force?
   (c) Determine the motion of the skate moving on an inclined plane, i.e. subject to a potential energy $U = Mg\sin\alpha\, x$.

(5) Prove Proposition 4.9. (**Hint:** Recall from Exercise 3.8.2 in Chapter 2 that $d\omega(X,Y) = X \cdot \omega(Y) - Y \cdot \omega(X) - \omega([X,Y])$ for any $\omega \in \Omega^1(M)$ and $X, Y \in \mathfrak{X}(M)$).

(6) Prove Theorem 4.14.

(7) Consider a vertical wheel of radius R moving on a plane.

(a) Show that the non-holonomic constraint corresponding to the condition of rolling without slipping or sliding is the distribution determined on the configuration space $\mathbb{R}^2 \times S^1 \times S^1$ by the 1-forms

$$\omega^1 = dx - R\cos\varphi\, d\psi, \qquad \omega^2 = dy - R\sin\varphi\, d\psi,$$

where (x, y, φ, ψ) are the local coordinates indicated in Figure 5.

(b) Assuming that the kinetic energy of the wheel is

$$K = \frac{M}{2}\left((v^x)^2 + (v^y)^2\right) + \frac{I}{2}(v^\varphi)^2 + \frac{J}{2}(v^\varphi)^2$$

and that the reaction force is perfect, show that the wheel moves with constant speed along straight lines or circles. What is the physical interpretation of the reaction force?

(c) Determine the motion of the vertical wheel moving on an inclined plane, i.e. subject to a potential energy $U = Mg\sin\alpha\, x$.

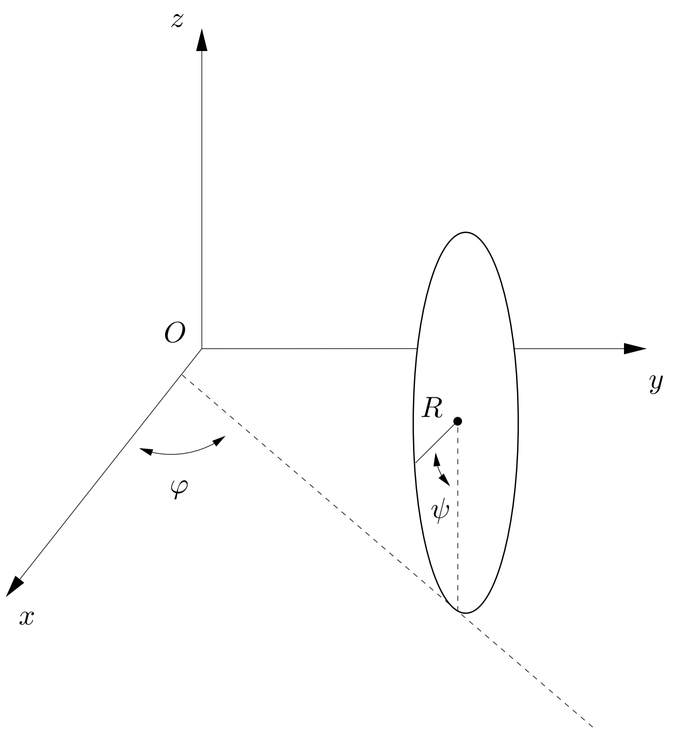

FIGURE 5. Vertical wheel on a plane.

(8) Consider a sphere of radius R and mass M rolling without slipping on a plane.

(a) Show that the condition of rolling without slipping is

$$\dot{x} = R\omega^2, \qquad \dot{y} = -R\omega^1,$$

where (x, y) are the Cartesian coodinates of the contact point on the plane and ω is the angular velocity of the sphere.

(b) Show that if the sphere's mass is symmetrically distributed then its kinetic energy is

$$K = \frac{M}{2}\left(\dot{x}^2 + \dot{y}^2\right) + \frac{I}{2}\langle \omega, \omega \rangle,$$

where I is the sphere's moment of inertia and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product.

(c) Using ω as coordinates on the fibers of $TSO(3)$, show that

$$\frac{D\dot{c}}{dt} = \ddot{x}\frac{\partial}{\partial x} + \ddot{y}\frac{\partial}{\partial y} + \dot{\omega}$$

(**Hint:** Recall from Exercise 4.8.3 in Chapter 3 that if ∇ is the Levi-Civita connection for a bi-invariant metric on a Lie group and X, Y are left-invariant vector fields then $\nabla_X Y = \frac{1}{2}[X, Y]$).

(d) Since we are identifying the fibers of $TSO(3)$ with $\mathbb{R}^3$, we can use the Euclidean inner product to also identify the fibers of $T^*SO(3)$ with $\mathbb{R}^3$. Show that under this identification the non-holonomic constraint yielding the condition of rolling without slipping is the distribution determined by the kernels of the 1-forms

$$\theta^1 = dx - R\,e_2, \qquad \theta^2 = dx - R\,e_1$$

(where $\{e_1, e_2, e_3\}$ is the canonical basis of $\mathbb{R}^3$). Is this distribution integrable?

(e) Show that the sphere moves along straight lines with constant speed and constant angular velocity orthogonal to its motion.

(f) Determine the motion of the sphere moving on an inclined plane, i.e. subject to a potential energy $U = Mg \sin\alpha\, x$.

5. Lagrangian Mechanics

Let M be a differentiable manifold, $p, q \in M$ and $a, b \in \mathbb{R}$ such that $a < b$. Let us denote by $\mathcal{C}$ the set of differentiable curves $c : [a, b] \to M$ such that $c(a) = p$ and $c(b) = q$.

Definition 5.1. *A* **Lagrangian function** *on M is a differentiable map $L : TM \to \mathbb{R}$. The* **action** *determined by L on $\mathcal{C}$ is the map $S : \mathcal{C} \to \mathbb{R}$ given by*

$$S(c) = \int_a^b L(\dot{c}(t))dt.$$

We can look for the global minima (or maxima) of the action by considering curves on $\mathcal{C}$.

Definition 5.2. *A* **variation** *of $c \in \mathcal{C}$ is a map $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{C}$ (for some $\varepsilon > 0$) such that $\gamma(0) = c$ and the map $\gamma : (-\varepsilon, \varepsilon) \times [a, b] \to M$ given by $\gamma(s, t) = \gamma(s)(t)$ is differentiable. The curve c is said to be a* **critical point** *of the action if*

$$\left.\frac{d}{ds}\right|_{s=0} S(\gamma(s)) = 0$$

for any variation γ of c.

Notice that the global minima (or maxima) of S must certainly be critical points. However, as is usually the case, a critical point is not necessarily a minimum (or a maximum).

It turns out that the critical points of the action are solutions of second order ODE's.

Theorem 5.3. *The curve $c \in \mathcal{C}$ is a critical point of the action determined by the Lagrangian $L : TM \to \mathbb{R}$ if and only if it satisfies the* **Euler-Lagrange equations**

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v^i}\left(x(t), \dot{x}(t)\right)\right) - \frac{\partial L}{\partial x^i}\left(x(t), \dot{x}(t)\right) = 0 \qquad (i = 1, \ldots, n)$$

for any local chart $(x^1, \ldots, x^n)$.

Proof. Assume first that the image of c is contained on the domain of a local chart $(x^1, \ldots, x^n)$. Let $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{C}$ be a variation of c. Setting $x(s, t) = x \circ \gamma(s, t)$, we have

$$S(\gamma(s)) = \int_a^b L\left(x(s,t), \frac{\partial x}{\partial t}(s,t)\right) dt,$$

and hence

$$\begin{aligned}\left.\frac{d}{ds}\right|_{s=0} S(\gamma(s)) &= \int_a^b \sum_{i=1}^n \frac{\partial L}{\partial x^i}\left(x(0,t), \frac{\partial x}{\partial t}(0,t)\right) \frac{\partial x^i}{\partial s}(0,t)\, dt \\ &+ \int_a^b \sum_{i=1}^n \frac{\partial L}{\partial v^i}\left(x(0,t), \frac{\partial x}{\partial t}(0,t)\right) \frac{\partial^2 x^i}{\partial s \partial t}(0,t)\, dt.\end{aligned}$$

Differentiating the relations $x(s, a) = x(p)$, $x(s, b) = x(q)$ with respect to s one obtains

$$\frac{\partial x}{\partial s}(0, a) = \frac{\partial x}{\partial s}(0, b) = 0.$$

Consequently, the second integral above can be integrated by parts to yield

$$-\int_a^b \sum_{i=1}^n \frac{d}{dt}\left(\frac{\partial L}{\partial v^i}\left(x(0,t), \frac{\partial x}{\partial t}(0,t)\right)\right) \frac{\partial x^i}{\partial s}(0,t)\, dt,$$

and hence

$$\left.\frac{d}{ds}\right|_{s=0} S(\gamma(s)) = \int_a^b \sum_{i=1}^n \left(\frac{\partial L}{\partial x^i}\left(x(t), \dot{x}(t)\right) - \frac{d}{dt}\left(\frac{\partial L}{\partial v^i}\left(x(t), \dot{x}(t)\right)\right)\right) w^i(t)\, dt,$$

where we have set $x(t) = x \circ c(t)$ and $w(t) = \frac{\partial x}{\partial s}(0,t)$. This shows that if c satisfies the Euler-Lagrange equations then c is a critical point of the action. To show the converse, we notice that any function $w : [a,b] \to \mathbb{R}^n$ satisfying $w(a) = w(b) = 0$ arises from the variation γ determined by $x(s,t) = x(t) + sw(t)$. In particular, if $\rho : [a,b] \to \mathbb{R}$ is a smooth positive function with $\rho(a) = \rho(b) = 0$, we can take

$$w^i(t) = \rho(t)\left(\frac{\partial L}{\partial x^i}\left(x(t),\dot{x}(t)\right) - \frac{d}{dt}\left(\frac{\partial L}{\partial v^i}\left(x(t),\dot{x}(t)\right)\right)\right).$$

Therefore if c is a critical point of the action then we must have

$$\int_a^b \sum_{i=1}^n \left(\frac{\partial L}{\partial x^i}\left(x(t),\dot{x}(t)\right) - \frac{d}{dt}\left(\frac{\partial L}{\partial v^i}\left(x(t),\dot{x}(t)\right)\right)\right)^2 \rho(t)\,dt = 0,$$

and hence c must satisfy the Euler-Lagrange equations.

The general case (in which the image of c is not contained in the domain of the local chart) is left as an exercise. □

COROLLARY 5.4. *The motions of any conservative mechanical system $(M, \langle\cdot,\cdot\rangle, -dU)$ are the critical points of the action determined by the Lagrangian $L = K - U$.*

Therefore we can find motions of conservative systems by looking for minima, say, of the action. This variational approach is often very useful.

The energy conservation in a conservative system is in fact a particular case of a more general conservation law, which holds for any Lagrangian.

DEFINITION 5.5. *The* **fiber derivative** *of a Lagrangian function $L : TM \to \mathbb{R}$ at $v \in T_pM$ is the linear map $(\mathbb{F}L)_v : T_pM \to \mathbb{R}$ given by*

$$(\mathbb{F}L)_v(w) = \left.\frac{d}{dt}\right|_{t=0} L(v + tw)$$

for all $w \in T_pM$.

DEFINITION 5.6. *If $L : TM \to \mathbb{R}$ is a Lagrangian function then its associated* **Hamiltonian function** *$H : TM \to \mathbb{R}$ is defined as*

$$H(v) = (\mathbb{F}L)_v(v) - L(v).$$

THEOREM 5.7. *The Hamiltonian function is constant along the solutions of the Euler-Lagrange equations.*

PROOF. In local coordinates, we have

$$H(x,v) = \sum_{i=1}^n v^i \frac{\partial L}{\partial v^i}(x,v) - L(x,v).$$

Consequently, if $c : I \subset \mathbb{R} \to M$ is a solution of the Euler-Lagrange equations, given in local coordinates by $x = x(t)$, we have

$$\begin{aligned}
\frac{d}{dt}\left(H(\dot{c}(t))\right) &= \frac{d}{dt}\left(\sum_{i=1}^{n} \dot{x}^i(t)\frac{\partial L}{\partial v^i}(x(t),\dot{x}(t)) - L(x(t),\dot{x}(t))\right) \\
&= \sum_{i,j=1}^{n} \ddot{x}^i(t)\frac{\partial L}{\partial v^i}(x(t),\dot{x}(t)) + \sum_{i=1}^{n} \dot{x}^i(t)\frac{d}{dt}\left(\frac{\partial L}{\partial v^i}(x(t),\dot{x}(t))\right) \\
&- \sum_{i=1}^{n} \dot{x}^i(t)\frac{\partial L}{\partial x^i}(x(t),\dot{x}(t)) - \sum_{i=1}^{n} \ddot{x}^i(t)\frac{\partial L}{\partial v^i}(x(t),\dot{x}(t)) = 0.
\end{aligned}$$

□

EXAMPLE 5.8. If $(M, \langle\cdot,\cdot\rangle, -dU)$ is a conservative mechanical system then its motions are the solutions of the Euler-Lagrange equations for the Lagrangean $L : TM \to \mathbb{R}$ given by

$$L(v) = \frac{1}{2}\langle v, v\rangle - U(\pi(v))$$

(where $\pi : TM \to M$ is the canonical projection). Clearly,

$$(\mathbb{F}L)_v(w) = \left.\frac{d}{dt}\right|_{t=0} \frac{1}{2}\langle v + tw, v + tw\rangle = \langle v, w\rangle,$$

and hence

$$H(v) = \langle v, v\rangle - \frac{1}{2}\langle v, v\rangle + U(\pi(v)) = \frac{1}{2}\langle v, v\rangle + U(\pi(v))$$

is the mechanical energy.

The Lagrangian formulation is particularly useful for exploring the relation between symmetry and conservation laws.

DEFINITION 5.9. *Let G be a Lie group acting on a manifold M. The Lagrangian $L : TM \to \mathbb{R}$ is said to be* **G-invariant** *if*

$$L\left((dg)_p v\right) = L(v)$$

for all $v \in T_pM$, $p \in M$ and $g \in G$ (where $g : M \to M$ is the map $p \mapsto g \cdot p$).

We will now show that if a Lagrangian is G-invariant then to each element $v \in \mathfrak{g}$ there corresponds a conserved quantity. To do so, we need the following definitions.

DEFINITION 5.10. *Let G be a Lie group acting on a manifold M. The* **infinitesimal action** *of $V \in \mathfrak{g}$ on M is the vector field $X^V \in \mathfrak{X}(M)$ defined as*

$$X^V_p = \left.\frac{d}{dt}\right|_{t=0} (\exp(tV) \cdot p) = (dA_p)_e V,$$

where $A_p : G \to M$ is the map $A_p(g) = g \cdot p$.

THEOREM 5.11. (Noether) *Let G be a Lie group acting on a manifold M. If $L : TM \to \mathbb{R}$ is G-invariant then $J^V : TM \to \mathbb{R}$ defined as $J^V(v) = (\mathbb{F}L)_v\left(X^V\right)$ is constant along the solutions of the Euler-Lagrange equations for all $V \in \mathfrak{g}$.*

PROOF. Choose local coordinates $(x^1, \ldots, x^n)$ on M and let $(y^1, \ldots, y^m)$ be local coordinates centered on $e \in G$. Let $A : G \times M \to M$ be the action of G on M, written in these local coordinates as

$$(A^1(x^1, \ldots, x^n, y^1, \ldots, y^m), \ldots, A^n(x^1, \ldots, x^n, y^1, \ldots, y^m)).$$

Then the infinitesimal action of $V = \sum_{i=1}^m V^a \frac{\partial}{\partial y^a}$ has components

$$X^i(x) = \sum_{a=1}^m \frac{\partial A^i}{\partial y^a}(x, 0) V^a.$$

Since L is G-invariant, we have

$$L\left(A^1(x,y), \ldots, A^n(x,y), \sum_{i=1}^n \frac{\partial A^1}{\partial x^i}(x,y)v^i, \ldots, \sum_{i=1}^n \frac{\partial A^n}{\partial x^i}(x,y)v^i\right) = L(x^1, \ldots, x^n, v^1, \ldots, v^n).$$

Setting $y = y(t)$ in the above identity, where $(y^1(t), \ldots, y^m(t))$ is the expression of the curve $\exp(tV)$ in local coordinates, and differentiating with respect to t at $t = 0$, one obtains

$$\sum_{i=1}^n \sum_{a=1}^m \frac{\partial L}{\partial x^i}(x,v) \frac{\partial A^i}{\partial y^a}(x,0)V^a + \sum_{i,j=1}^n \sum_{a=1}^m \frac{\partial L}{\partial v^i}(x,v) \frac{\partial^2 A^i}{\partial y^a \partial x^j}(x,0) v^j V^a = 0$$
$$\Leftrightarrow \sum_{i=1}^n \frac{\partial L}{\partial x^i}(x,v) X^i(x) + \sum_{i,j=1}^n \frac{\partial L}{\partial v^i}(x,v) \frac{\partial X^i}{\partial x^j}(x) v^j = 0,$$

where we have used the fact that $A(x, 0) = x$ (and hence $\frac{\partial A^i}{\partial x^j}(x, 0) = \delta_{ij}$).

In these coordinates,

$$J^V(x) = \sum_{i=1}^n \frac{\partial L}{\partial v^i}(x,v) X^i(x).$$

Therefore, if $c : I \subset \mathbb{R} \to M$ is a solution of the Euler-Lagrange equations, given in local coordinates by $x = x(t)$, we have

$$\frac{d}{dt}\left(J^V(\dot{c}(t))\right) = \frac{d}{dt}\left(\sum_{i=1}^n \frac{\partial L}{\partial v^i}(x(t), \dot{x}(t)) X^i(x(t))\right)$$
$$= \sum_{i=1}^n \frac{d}{dt}\left(\frac{\partial L}{\partial v^i}(x(t), \dot{x}(t))\right) X^i(x(t)) + \sum_{i,j=1}^n \frac{\partial L}{\partial v^i}(x(t), \dot{x}(t)) \frac{\partial X^i}{\partial x^j}(x(t)) \dot{x}^j(t)$$
$$= \sum_{i=1}^n \frac{d}{dt}\left(\frac{\partial L}{\partial v^i}(x(t), \dot{x}(t))\right) X^i(x(t)) - \sum_{i=1}^n \frac{\partial L}{\partial x^i}(x(t), \dot{x}(t)) X^i(x(t)) = 0.$$

□

REMARK 5.12. Notice that the map $\mathfrak{g} \ni V \mapsto X^V \in \mathfrak{X}(M)$ is linear. Since $\mathbb{F}L(v)$ is also linear, we can see J^V as a linear map $\mathfrak{g} \ni V \mapsto J^V \in C^\infty(M)$. Therefore Noether's theorem yields $m = \dim \mathfrak{g}$ independent conserved quantities.

EXAMPLE 5.13. Consider a conservative mechanical system consisting of k particles with masses $m_1, \ldots, m_k$ moving in $\mathbb{R}^3$ under a potential energy $U : \mathbb{R}^{3N} \to \mathbb{R}$ which depends only on the distances between them. The motions of the system are the solutions of the Euler-Lagrange equations obtained from the Lagrangian $L : T\mathbb{R}^{3k} \to \mathbb{R}$ given by

$$L(x_1, \ldots, x_k, v_1, \ldots, v_k) = \frac{1}{2} \sum_{i=1}^{k} m_i \langle v_i, v_i \rangle + U(x_1, \ldots, x_k).$$

This Lagrangian is clearly $SO(3)$-invariant, where the action of $SO(3)$ on $\mathbb{R}^{3k}$ is defined through

$$S \cdot (x_1, \ldots, x_k) = (Sx_1, \ldots, Sx_k).$$

The infinitesimal action of $V \in \mathfrak{so}(3)$ is the vector field

$$X^V_{(x_1,\ldots,x_k)} = (Vx_1, \ldots, Vx_k) = (\Omega(V) \times x_1, \ldots, \Omega(V) \times x_k),$$

where $\Omega : \mathfrak{so}(3) \to \mathbb{R}^3$ is the isomorphism in Lemma 3.9. On the other hand,

$$(\mathbb{F}L)_{(v_1,\ldots,v_k)}(w_1, \ldots, w_k) = \sum_{i=1}^{k} m_i \langle v_i, w_i \rangle.$$

Therefore, Noether's Theorem guarantees that the quantity

$$J^V = \sum_{i=1}^{k} m_i \langle \dot{x}_i, \Omega(V) \times x_i \rangle = \sum_{i=1}^{k} m_i \langle \Omega(V), x_i \times \dot{x}_i \rangle = \left\langle \Omega(V), \sum_{i=1}^{k} m_i x_i \times \dot{x}_i \right\rangle$$

is conserved along the motion of the system for any $V \in \mathfrak{so}(3)$. In other words, the system's **total angular momentum**

$$Q = \sum_{i=1}^{k} m_i x_i \times \dot{x}_i$$

is conserved.

EXERCISES 5.14.

(1) Complete the proof of Theorem 5.3.
(2) Show that the geodesics of a Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$ are, up to reparametrization, critical points of the arclength, i.e., of the action determined by the Lagrangian $L : TM \to \mathbb{R}$ given by

$$L(v) = \langle v, v \rangle^{\frac{1}{2}}$$

(where we must restrict the action to curves with nonvanishing velocity).

(3) (*Brachistochrone curve*) A particle with mass m moves on a curve $y = y(x)$ under the action of a constant gravitational field, corresponding to the potential energy $U = mgy$. The curve satisfies $y(0) = y(d) = 0$ and $y(x) < 0$ for $0 < x < d$.

(a) Asuming that the particle is set free at the origin with zero velocity, show that its speed at each point is

$$v = \sqrt{-2gy},$$

and that therefore the travel time between the origin and point $(d, 0)$ is

$$S = (2g)^{-\frac{1}{2}} \int_0^d (1 + y'^2)^{\frac{1}{2}} (-y)^{-\frac{1}{2}} dx,$$

where $y' = \frac{dy}{dx}$.

(b) Write a differential equation for the curve $y = y(x)$ which corresponds to the minimum travel time, and show that it can be integrated to

$$\frac{d}{dx} \left[\left(1 + y'^2\right) y \right] = 0.$$

(c) Check that the solution of this equation satisfying $y(0) = y(d) = 0$ is given parametrically by

$$\begin{cases} x = R\theta - R\sin\theta \\ y = -R + R\cos\theta \end{cases}$$

where $d = 2\pi R$. (This curve is called a **cycloid**, because it is the curved traced out by a point on a circle which rolls without slipping on the xx-axis).

(4) (*Charged particle in a stationary electromagnetic field*) The motion of a particle with mass $m > 0$ and charge $e \in \mathbb{R}$ in a stationary electromagnetic field is determined by the Lagrangian $L : T\mathbb{R}^3 \to \mathbb{R}$ given by

$$L = \frac{1}{2} m \langle \dot{x}, \dot{x} \rangle + e \langle A, \dot{x} \rangle - e\Phi,$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product, $\Phi \in C^\infty(\mathbb{R})$ is the **electric potencial** and $A \in \mathfrak{X}(\mathbb{R}^3)$ is the **magnetic vector potential**.

(a) Show that the motion equations are

$$m\ddot{x} = eE + \dot{x} \times B,$$

where $E = -\operatorname{grad} \Phi$ is the **electric field** and $B = \operatorname{curl} A$ is the **magnetic field**.

(b) Write an expression for the Hamiltonian function and use the motion equations to check that it is constant along any motion.

(5) (*Restricted 3-body problem*) Consider two particles with masses $\mu \in (0,1)$ and $1-\mu$, set in circular orbit around each other. We identify the plane of the orbit with $\mathbb{R}^2$ and place the center of mass at the origin. In the rotating frame where the particles are at rest they are placed at, say, $p_1 = (1-\mu, 0)$ and $p_2 = (-\mu, 0)$. The motion of a third particle with negligible mass in this frame is determined by the Lagrangian $L : T\left(\mathbb{R}^2 \setminus \{p_1, p_2\}\right) \to \mathbb{R}$ given by

$$L(x, y, \dot{x}, \dot{y}) = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2\right) + x\dot{y} - y\dot{x} + \frac{1}{2}\left(x^2 + y^2\right) + \frac{1-\mu}{r_1} + \frac{\mu}{r_2},$$

where $r_1, r_2 : \mathbb{R}^2 \to \mathbb{R}$ are the Euclidean distances to p_1, p_2.

(a) Determine the equations of motion and the Hamiltonian function.

(b) Show that (r_1, r_2) are local coordinates for each of the half-planes $\{y > 0\}$ and $\{y < 0\}$, and that

$$(1-\mu){r_1}^2 + \mu{r_2}^2 = x^2 + y^2 + \mu(1-\mu).$$

Use this result to compute the equilibrium points on these half-planes.

(6) Consider the mechanical system in Example 5.13.

(a) Use Noether's Theorem to prove that the **total linear momentum**

$$P = \sum_{i=1}^{N} m_i \dot{x}_i$$

is conserved along the motion.

(b) Show that the system's **center of mass**, defined as the point

$$X = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i},$$

moves with constant velocity.

(7) Generalize Example 5.13 to the case in which the particles move in an arbitrary Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$, by showing that given any Killing vector field $X \in \mathfrak{X}(M)$ (cf. Exercise 3.3.8 in Chapter 3) the quantity

$$J^X = \sum_{i=1}^{N} m_i \langle \dot{c}_i, X\rangle$$

is conserved, where $c_i : I \subset \mathbb{R} \to M$ is the motion of the particle with mass m_i.

(8) Consider the action of $SO(3)$ on itself by left multiplication.

(a) Show that the infinitesimal action of $B \in \mathfrak{so}(3)$ is the **right-invariant** vector field determined by B.

(b) Use Noether's Theorem to show that the angular momentum of the free rigid body is constant.

(9) Consider a satellite equiped with a small rotor, i.e. a cylinder which can spin freely about its axis. When the rotor is locked the satellite can be modelled by a free rigid body with inertia tensor I. The rotor's axis passes through the satellite's center of mass, and its direction is given by the unit vector e. The rotor's mass is symmetrically distributed around the axis, producing a moment of inertia J.

(a) Show that the configuration space for the satellite with unlocked rotor is the Lie group $SO(3) \times S^1$, and that its motion is a geodesic of the left-invariant metric corresponding to the kinetic energy

$$K = \frac{1}{2}\langle I\Omega, \Omega\rangle + \frac{1}{2}J\varpi^2 + J\varpi\langle \Omega, e\rangle,$$

where the $\Omega \in \mathbb{R}^3$ is the satellite's angular velocity as seen on the satellite's frame and $\varpi \in \mathbb{R}$ is the rotor's angular speed around its axis.

(b) Use Noether's Theorem to show that $l = J(\varpi + \langle \Omega, e\rangle) \in \mathbb{R}$ and $p = S(I\Omega + J\varpi e) \in \mathbb{R}^3$ are conserved along the motion of the satellite with unlocked rotor, where $S : \mathbb{R} \to SO(3)$ describes the satellite's orientation.

6. Hamiltonian Mechanics

We will now see that under certain conditions it is possible to study the Euler-Lagrange equations as a flow on the cotangent bundle with special geometric properties.

Let M be an n-dimensional manifold. The set

$$TM \oplus T^*M := \bigcup_{p \in M} T_pM \times T_p^*M$$

has an obvious differentiable structure: if $(x^1, \ldots, x^n)$ are local coordinates on M then $(x^1, \ldots, x^n, v^1, \ldots, v^n, p_1, \ldots, p_n)$ are the local coordinates on $TM \oplus T^*M$ which label the pair $(v, \omega) \in T_pM \times T_p^*M$, where

$$v = \sum_{i=1}^{n} v^i \frac{\partial}{\partial x^i}, \qquad \omega = \sum_{i=1}^{n} p_i dx^i,$$

and $p \in M$ is the point with coordinates $(x^1, \ldots, x^n)$. For this differentiable structure, the maps $\pi_1 : TM \oplus T^*M \to TM$ and $\pi_2 : TM \oplus T^*M \to T^*M$ given by $\pi_1(v, \omega) = v$ and $\pi_2(v, \omega) = \omega$ are submersions.

Definition 6.1. *The* **extended Hamiltonian function** *corresponding to a Lagrangian* $L : TM \to \mathbb{R}$ *is the map* $\widetilde{H} : TM \oplus T^*M \to \mathbb{R}$ *given by*

$$\widetilde{H}(v, \omega) = \omega(v) - L(v).$$

In local coordinates, we have

$$\widetilde{H}(x^1,\ldots,x^n,v^1,\ldots,v^n,p_1,\ldots,p_n)=\sum_{i=1}^n p_iv^i-L(x^1,\ldots,x^n,v^1,\ldots,v^n),$$

and hence

$$d\widetilde{H}=\sum_{i=1}^n\left(p_i-\frac{\partial L}{\partial v^i}\right)dv^i+\sum_{i=1}^n v^i dp_i-\sum_{i=1}^n\frac{\partial L}{\partial x^i}dx^i.$$

Thus any critical point of any restriction of $\widetilde{H}$ to a submanifold of the form $\{\omega\}\times T_pM$ must satisfy

$$p_i-\frac{\partial L}{\partial v^i}(x^1,\ldots,x^n,v^1,\ldots,v^n)=0 \qquad (i=1,\ldots,n).$$

It follows that the set of all such critical points is naturally a $2n$-dimensional submanifold $S\subset TM\oplus T^*M$ such that $\pi_1|_S:S\to TM$ is a diffeomorphism. If $\pi_2|_S:S\to T^*M$ is also a diffeomorphism then the Lagrangian is said to be **hyper-regular**. In this case, $\pi_2|_S\circ\pi_1|_S{}^{-1}:TM\to T^*M$ is a fiber-preserving diffeomorphism, called the **Legendre transformation**.

Given a hyper-regular Lagrangian, we can use the maps $\pi_1|_S$ and $\pi_2|_S$ to make the identifications $TM\cong S\cong T^*M$. Since the Hamiltonian function $H:TM\to\mathbb{R}$ is clearly related to the extended Hamiltonian function through $H=\widetilde{H}\circ\pi_1|_S{}^{-1}$, we can under these identifications simply write $H=\widetilde{H}|_S$. Therefore

$$dH=\sum_{i=1}^n v^i dp_i-\sum_{i=1}^n\frac{\partial L}{\partial x^i}dx^i$$

(here we must think of $(x^1,\ldots,x^n,v^1,\ldots,v^n,p_1,\ldots,p_n)$ as local functions on S such that **both** $(x^1,\ldots,x^n,v^1,\ldots,v^n)$ and $(x^1,\ldots,x^n,p_1,\ldots,p_n)$ are local coordinates). On the other hand, thinking of H as a function on the cotangent bundle, we must have

$$dH=\sum_{i=1}^n\frac{\partial H}{\partial x^i}dx^i+\sum_{i=1}^n\frac{\partial H}{\partial p_i}dp_i$$

Therefore we must have

$$\begin{cases}\dfrac{\partial H}{\partial x^i}=-\dfrac{\partial L}{\partial x^i}\\[2ex] \dfrac{\partial H}{\partial p_i}=v^i\end{cases} \qquad (i=1,\ldots,n),$$

where the partial derivatives of the Hamiltionian must be computed with respect to the local coordinates $(x^1,\ldots,x^n,p_1,\ldots,p_n)$ and the partial derivatives of the Lagrangian must be computed with respect to the local coordinates $(x^1,\ldots,x^n,v^1,\ldots,v^n)$.

PROPOSITION 6.2. *The Euler-Lagrange equations for a hyper-regular Lagrangian $L : TM \to \mathbb{R}$ define a flow on TM. This flow is carried by the Legendre tranformation to the flow defined on T^*M by* **Hamilton's equations**

$$\begin{cases} \dot{x}^i = \dfrac{\partial H}{\partial p_i} \\ \\ \dot{p}_i = -\dfrac{\partial H}{\partial x^i} \end{cases} \qquad (i = 1, \ldots, n).$$

PROOF. The Euler-Lagrange equations can be cast as a system of first order ordinary differential equations on TM as follows:

$$\begin{cases} \dot{x}^i = v^i \\ \\ \dfrac{d}{dt}\left(\dfrac{\partial L}{\partial v^i}\right) = \dfrac{\partial L}{\partial x^i} \end{cases} \qquad (i = 1, \ldots, n).$$

Since on S one has

$$p_i = \frac{\partial L}{\partial v^i}, \qquad v^i = \frac{\partial H}{\partial p_i}, \qquad \frac{\partial L}{\partial x^i} = -\frac{\partial H}{\partial x^i},$$

we see that this system reduces to Hamilton's equations in the local coordinates $(x^1, \ldots, x^n, p_1, \ldots, p_n)$. Since Hamilton's equations clearly define a flow on T^*M, the Euler-Lagrange equations must define a flow on TM. $\square$

EXAMPLE 6.3. The Lagrangian for a conservative mechanical system $(M, \langle \cdot, \cdot \rangle, -dU)$ is written is local coordinates as

$$L(x^1, \ldots, x^n, v^1, \ldots, v^n) = \frac{1}{2} \sum_{i,j=1}^{n} g_{ij}(x^1, \ldots, x^n) v^i v^j - U(x^1, \ldots, x^n).$$

The Legendre transformation is given in these coordinates by

$$p_i = \frac{\partial L}{\partial v^i} = \sum_{j=1}^{n} g_{ij} v^j \qquad (i = 1, \ldots, n),$$

and is indeed a fiber-preserving diffeomorphism, whose inverse is given by

$$v^i = \sum_{j=1}^{n} g^{ij} p_j \qquad (i = 1, \ldots, n).$$

As a function on the tangent bundle, the Hamiltonian is (cf. Example 5.8)

$$H = \frac{1}{2} \sum_{i,j=1}^{n} g_{ij} v^i v^j + U.$$

Using the Legendre transformation, we can see the Hamiltonian as the following function on the cotangent bundle:

$$H = \frac{1}{2} \sum_{i,j,k,l=1}^{n} g_{ij} g^{ik} p_k g^{jl} p_l + U = \frac{1}{2} \sum_{k,l=1}^{n} g^{kl} p_k p_l + U.$$

Therefore Hamilton's equations for a conservative mechanical system are

$$\begin{cases} \dot{x}^i = \displaystyle\sum_{j=1}^{n} g^{ij} p_j \\ \\ \dot{p}_i = \displaystyle\frac{1}{2} \sum_{k,l=1}^{n} \frac{\partial g^{kl}}{\partial x^i} p_k p_l \end{cases} \qquad (i = 1, \ldots, n).$$

The flow defined by Hamilton's equations has remarkable geometric properties, which are better understood by introducing the following definition.

Definition 6.4. *The* **canonical symplectic potential** *is the* 1*-form* $\theta \in \Omega^1(T^*M)$ *given by*

$$\theta_\alpha(v) = \alpha\left((d\pi)_\alpha(v)\right)$$

for all $v \in T_\alpha(T^*M)$ *and all* $\alpha \in T^*M$, *where* $\pi : T^*M \to M$ *is the natural projection. The* **canonical symplectic form** *is the* 2*-form* $\omega \in \Omega^2(T^*M)$ *given by* $\omega = d\theta$.

In local coordinates, we have

$$\pi(x^1, \ldots, x^n, p_1, \ldots, p_n) = (x^1, \ldots, x^n)$$

and

$$v = \sum_{i=1}^{n} dx^i(v) \frac{\partial}{\partial x^i} + \sum_{i=1}^{n} dp_i(v) \frac{\partial}{\partial p_i}.$$

Consequently,

$$(d\pi)_\alpha(v) = \sum_{i=1}^{n} dx^i(v) \frac{\partial}{\partial x^i},$$

and hence

$$\theta_\alpha(v) = \alpha\left((d\pi)_\alpha(v)\right) = \sum_{i=1}^{n} p_i dx^i \left(\sum_{j=1}^{n} dx^j(v) \frac{\partial}{\partial x^j} \right) = \sum_{i=1}^{n} p_i dx^i(v).$$

We conclude that

$$\theta = \sum_{i=1}^{n} p_i dx^i,$$

and consequently

$$\omega = \sum_{i=1}^{n} dp_i \wedge dx^i$$

Proposition 6.5. *The canonical symplectic form ω is closed ($d\omega = 0$) and non-degenerate. Moreover, $\omega^n = \omega \wedge \ldots \wedge \omega$ is a volume form (in particular T^*M is always orientable, even if M itself is not).*

We leave the proof of this proposition as an exercise. Recall from Exercise 1.14.8 in Chapter 2 that if $v \in T_pM$ then $\iota(v)\omega \in T_p^*M$ is the covector given by

$$(\iota(v)\omega)\,(w) = \omega(v, w)$$

for all $w \in T_pM$. Therefore the first statement in Proposition 6.5 is equivalent to saying that the map $T_pM \ni v \mapsto \iota(v)\omega \in T_p^*M$ is a linear isomorphism for all $p \in M$.

The key to the geometric meaning of Hamilton's equations is contained in the following result.

Proposition 6.6. *The Hamilton equations are the equations for the flow of the vector field X_H satisfying*

$$\iota(X_H)\omega = -dH.$$

Proof. Hamilton's equations yield the flow of the vector field

$$X_H = \sum_{i=1}^{n} \left(\frac{\partial H}{\partial p_i} \frac{\partial}{\partial x^i} - \frac{\partial H}{\partial x^i} \frac{\partial}{\partial p_i} \right).$$

Therefore

$$\begin{aligned} \iota(X_H)\omega &= \iota(X_H) \sum_{i=1}^{n} (dp_i \otimes dx^i - dx^i \otimes dp_i) \\ &= \sum_{i=1}^{n} \left(-\frac{\partial H}{\partial x^i} dx^i - \frac{\partial H}{\partial p_i} dp_i \right) = -dH \end{aligned}$$

□

Remark 6.7. Notice that H completely determines X_H, as ω is nondegenerate. By analogy with the Riemannian case, $-X_H$ is sometimes called the **symplectic gradient** of H.

Definition 6.8. *The* **Hamiltonian flow** *generated by $F \in C^\infty(T^*M)$ is the flow of the unique vector field $X_F \in \mathfrak{X}(T^*M)$ such that*

$$\iota(X_F)\omega = -dF.$$

The flow determined on T^*M by a hyper-regular Lagrangian is therefore a particular case of a Hamiltonian flow (in which the generating function is the Hamiltonian function). We will now dicuss the geometric properties of general Hamiltonian flows.

Proposition 6.9. *Hamiltonian flows preserve their generating functions.*

Proof. $X_F \cdot F = dF(X_F) = (-\iota(X_F)\omega)\,(X_F) = -\omega(X_F, X_F) = 0$, as ω is alternating. □

PROPOSITION 6.10. *Hamiltonian flows preserve the canonical symplectic form: if* $\psi_t : T^*M \to T^*M$ *is a Hamiltonian flow then* ${\psi_t}^*\omega = \omega$.

PROOF. Let $F \in C^\infty(T^*M)$ be the function whose Hamiltonian flow is ψ_t. Recall from Exercise 3.8.8 in Chapter 2 that the Lie derivative of ω along $X_F \in \mathfrak{X}(T^*M)$,

$$L_{X_F}\omega = \left.\frac{d}{dt}\right|_{t=0} {\psi_t}^*\omega,$$

can be computed by the Cartan formula:

$$L_{X_F}\omega = \iota(X_F)d\omega + d(\iota(X_F)\omega) = d(-dF) = 0.$$

Therefore

$$\begin{aligned}\frac{d}{dt}{\psi_t}^*\omega = \left.\frac{d}{ds}\right|_{s=0}(\psi_{t+s})^*\omega = \left.\frac{d}{ds}\right|_{s=0}(\psi_s \circ \psi_t)^*\omega = \left.\frac{d}{ds}\right|_{s=0}{\psi_t}^*(\psi_s)^*\omega \\ = {\psi_t}^*\left.\frac{d}{ds}\right|_{s=0}(\psi_s)^*\omega = {\psi_t}^*L_X\omega = 0.\end{aligned}$$

We conclude that

$${\psi_t}^*\omega = (\psi_0)^*\omega = \omega.$$

□

THEOREM 6.11. (Liouville) *Hamiltonian flows preserve integral with respect to the symplectic volume form: if* $\psi_t : T^*M \to T^*M$ *is a Hamiltonian flow and* $F \in C^\infty(T^*M)$ *is a compactly supported function then*

$$\int_{T^*M} F \circ \psi_t = \int_{T^*M} F.$$

PROOF. This is a simple consequence of the fact that ψ_t preserves the symplectic volume form:

$${\psi_t}^*(\omega^n) = ({\psi_t}^*\omega)^n = \omega^n.$$

Therefore

$$\begin{aligned}\int_{T^*M} F \circ \psi_t = \int_{T^*M}(F \circ \psi_t)\omega^n = \int_{T^*M}(F \circ \psi_t){\psi_t}^*(\omega^n) \\ = \int_{T^*M}{\psi_t}^*(F\omega^n) = \int_{T^*M} F\omega^n = \int_{T^*M} F\end{aligned}$$

(cf. Exercise 4.2.3 in Chapter 2). □

COROLLARY 6.12. (Poincaré Recurrence) *Let* $\psi_t : T^*M \to T^*M$ *be a Hamiltonian flow and* $K \subset T^*M$ *a compact set invariant for* ψ_t. *Then for each open set* $U \subset K$ *and each* $T > 0$ *there exist* $p \in U$ *and* $t \geq T$ *such that* $\psi_t(p) \in U$.

PROOF. Let $F \in C^\infty(T^*M)$ be non-negative function with support contained in $\overline{U} \subset K$ such that

$$\int_{T^*M} F > 0.$$

Consider the open sets $U_n = \psi_{nT}(U)$. If these sets were all disjoint then one could define a function $\widetilde{F} \in C^\infty(M)$ through

$$\widetilde{F}(p) = \begin{cases} (F \circ \psi_{nT})(p) \text{ if } p \in U_n \\ 0 \text{ otherwise} \end{cases}$$

The support of $\widetilde{F}$ would be a closed subset of K, hence compact. On the other hand, one would have

$$\int_{T^*M} \widetilde{F} \geq \sum_{n=1}^{N} \int_{T^*M} F \circ \psi_{nT} = N \int_{T^*M} F$$

for all $N \in \mathbb{N}$, which is absurd. We conclude that there must exist $m, n \in \mathbb{N}$ (with, say, $n > m$) such that

$$U_m \cap U_n \neq \varnothing \Leftrightarrow \psi_{mT}(U) \cap \psi_{nT}(U) \neq \varnothing \Leftrightarrow U \cap \psi_{(n-m)T}(U) \neq \varnothing.$$

Choosing $t = (n-m)T$ and $p \in U \cap \psi_t(U)$ yields the result. $\square$

We can use the symplectic structure of the contangent bundle to define a new binary operation on the set of differentiable functions on T^*M.

Definition 6.13. *The* **Poisson bracket** *of two differentiable functions $F, G \in C^\infty(T^*M)$ is $\{F, G\} := X_F(G)$.*

Proposition 6.14. *$(C^\infty(T^*M), \{\cdot, \cdot\})$ is a Lie algebra, and the map that associates to a function $F \in C^\infty(T^*M)$ its Hamiltonian vector field $X_F \in \mathfrak{X}(T^*M)$ is a Lie algebra homomorphism:*

(i) $\{F, G\} = -\{G, F\}$;
(ii) $\{\alpha F + \beta G, H\} = \alpha\{F, H\} + \beta\{F, H\}$;
(iii) $\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\} = 0$;
(iv) $X_{\{F,G\}} = [X_F, X_G]$

(for any $F, G, H \in C^\infty(T^*M)$ and any $\alpha, \beta \in \mathbb{R}$).

Proof. We have

$$\begin{aligned} \{F, G\} &= X_F(G) = dG(X_F) = (-\iota(X_G)\omega)(X_F) \\ &= -\omega(X_G, X_F) = \omega(X_F, X_G), \end{aligned}$$

which proves the anti-symmetry and bilinearity of the Poisson bracket. On the other hand,

$$\begin{aligned} \iota(X_{\{F,G\}}\omega) &= -d\{F, G\} = -d(X_F \cdot G) = -d(\iota(X_F)dG) = -L_{X_F}dG \\ &= L_{X_F}(\iota(X_G)\omega) = \iota(L_{X_F}X_G)\omega + \iota(X_G)L_{X_F}\omega \\ &= \iota([X_F, X_G])\omega \end{aligned}$$

(cf. Exercise 3.8.8 in Chapter 2). Since ω is non-degenerate, we have

$$X_{\{F,G\}} = [X_F, X_G].$$

Finally,

$$\begin{aligned}
&\{F,\{G,H\}\}+\{G,\{H,F\}\}+\{H,\{F,G\}\}\\
&=\{F,X_G\cdot H\}-\{G,X_F\cdot H\}+X_{\{F,G\}}\cdot H\\
&=X_F\cdot(X_G\cdot H)-X_G\cdot(X_F\cdot H)-[X_F,X_G]\cdot H=0.
\end{aligned}$$

□

REMARK 6.15. In general, we can define a **symplectic manifold** as a pair (M,ω), where M is a differentiable manifold and $\omega\in\Omega^2(M)$ is closed and nondegenerate (hence the dimension of M is necessarily even). All definitions and results above are readily extended to arbitrary symplectic manifolds.

Darboux's Theorem guarantees that around each point of a symplectic manifold (M,ω) there exist local coordinates $(x^1,\ldots,x^n,p_1,\ldots,p_n)$ such that

$$\omega=\sum_{i=1}^n dp_i\wedge dx^i.$$

Therefore all symplectic manifolds are locally the same (i.e. there is no symplectic analogue of the curvature).

EXERCISES 6.16.

(1) Prove Proposition 6.5
(2) Let $(M,\langle\cdot,\cdot\rangle)$ be a compact Riemannian manifold. Show that for each normal ball $B\subset M$ and each $T>0$ there exist geodesics $c:\mathbb{R}\to M$ with $\|\dot c(t)\|=1$ such that $c(0)\in B$ and $c(t)\in B$ for some $t\geq T$.
(3) Let $(x^1,\ldots,x^n,p_1,\ldots,p_n)$ be the usual local coordinates on T^*M. Compute X_{x^i}, X_{p_i}, $\{x^i,x^j\}$, $\{p_i,p_j\}$ and $\{p_i,x^j\}$.
(4) Show that the Poisson bracket satisfies the **Leibnitz rule**

$$\{F,GH\}=\{F,G\}H+\{F,H\}G$$

for all $F,G,H\in C^\infty(T^*M)$.
(5) Let $(M,\langle\cdot,\cdot\rangle)$ be a Riemannian manifold, $\alpha\in\Omega^1(M)$ a 1-form and $U\in C^\infty(M)$ a differentiable function.
  (a) Show that the Euler-Lagrange equations for the Lagrangian $L:TM\to\mathbb{R}$ given by

$$L(v)=\frac{1}{2}\langle v,v\rangle+\iota(v)\alpha_p-U(p)$$

for $v\in T_pM$ yield the motions of the mechanical system $(M,\langle\cdot,\cdot\rangle,\mathcal{F})$, where

$$\mathcal{F}(v)=-(dU)_p-\iota(v)(d\alpha)_p$$

for $v\in T_pM$.

(b) Show that the mechanical energy $E = K + U$ is conserved along the motions of $(M, \langle\cdot,\cdot\rangle, \mathcal{F})$ (which is therefore called a **conservative mechanical system with magnetic term**).
(c) Show that L is hyper-regular and compute the Legendre transformation.
(d) Find the Hamiltonian $H : T^*M \to \mathbb{R}$ and write Hamilton's equations.
(e) Show that $\tilde{\omega} = \omega + \pi^* d\alpha$ is a symplectic form on T^*M, where ω is the canonical symplectic form and $\pi : T^*M \to M$ is the natural projection ($\tilde{\omega}$ is called a **canonical symplectic form with magnetic term**).
(f) Show that the Hamiltonian flow generated by the function $\tilde{H} \in C^\infty(T^*M)$ with respect to the symplectic form $\tilde{\omega}$ is given by the equations

$$\begin{cases} \dot{x}^i = \dfrac{\partial \tilde{H}}{\partial p_i} \\ \\ \dot{p}_i = -\dfrac{\partial \tilde{H}}{\partial x^i} + \displaystyle\sum_{j=1}^n \left(\frac{\partial \alpha_j}{\partial x^i} - \frac{\partial \alpha_i}{\partial x^j} \right) \dot{x}^j \end{cases}$$

(g) The map $F : T^*M \to T^*M$ given by

$$F(\xi) = \xi - \alpha_p$$

for $\xi \in T_p^*M$ is a fiber-preserving diffeomorfism. Show that F carries the Hamiltonian flow of H with respect to the canonical symplectic form ω to the Hamiltonian flow of $\tilde{H}$ with respect to the symplectic form $\tilde{\omega}$, where

$$\tilde{H}(\xi) = \frac{1}{2}\langle \xi, \xi \rangle + U(p)$$

for $\xi \in T_p^*M$. (**Remark:** Since the projections of the two flows on M coincide, we see that the motion of a conservative mechanical system with magnetic term can be obtained by changing either the Lagrangian or the symplectic form.)

(6) Show that:
(a) symplectic manifolds are even-dimensional and orientable;
(b) any orientable 2-manifold admits a symplectic structure;
(c) S^2 is the only sphere which admits a symplectic structure. (**Hint:** Use the fact that if $n > 2$ then any closed 2-form $\omega \in \Omega^2(S^n)$ is exact).

(7) Consider the symplectic structure on

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$$

determined by the usual volume form. Compute the Hamiltonian flow generated by the function $H(x, y, z) = z$.

7. Completely Integrable Systems

We now concentrate on studying the Hamiltonian flow of the Hamiltonian function $H \in C^\infty(T^*M)$. As in the previous section, all definitions and results can be easily extended to arbitrary symplectic manifolds

We already know that H is constant along its Hamiltonian flow, so that it suffices to study this flow along the level sets of H. This can be further simplified if there exist additional nontrivial functions $F \in C^\infty(T^*M)$ such that

$$X_H \cdot F = 0 \Leftrightarrow \{H, F\} = 0.$$

DEFINITION 7.1. *A function $F \in C^\infty(T^*M)$ is said to be a* **first integral** *of H if $\{H, F\} = 0$.*

In general, there is no reason to expect that there should exist nontrivial first integrals othen than H itself. In the special cases when these exist, many times they satisfy additional conditions.

DEFINITION 7.2. *The functions $F_1, \ldots, F_m \in C^\infty(M)$ are said to be*

(i) **in involution** *if $\{F_i, F_j\} = 0 \qquad (i, j = 1, \ldots, m)$;*
(ii) **independent** *at $\alpha \in T^*M$ if $(dF_1)_\alpha, \ldots, (dF_m)_\alpha \in T^*_\alpha(T^*M)$ are linearly independent covectors.*

PROPOSITION 7.3. *If $F_1, \ldots, F_m \in C^\infty(T^*M)$ are in involution and are independent at some point $\alpha \in T^*M$ then $m \leq n$.*

We leave the proof of this proposition as an exercise. The maximal case $m = n$ is especially interesting.

DEFINITION 7.4. *The Hamiltonian H is said to be* **completely integrable** *if there exist n first integrals $F_1, \ldots, F_n$ in involution which are independent on an dense open set $U \subset T^*M$.*

EXAMPLE 7.5.

(1) If M is 1-dimensional and $dH \neq 0$ on a dense open set of T^*M then H is completely integrable.
(2) (*Particle in a central field*) Recall from Example 1.15 that a particle of mass $m > 0$ moving in a central field is described by the Lagrangian function

$$L\left(r, \theta, v^r, v^\theta\right) = \frac{1}{2} m \left[(v^r)^2 + r^2 \left(v^\theta\right)^2 \right] + u(r).$$

corresponding to the Hamiltonian

$$H\left(r, \theta, p_r, p_\theta\right) = \frac{{p_r}^2}{2m} + \frac{{p_\theta}^2}{2mr^2} + u(r).$$

By Hamilton's equations,

$$\dot{p_\theta} = -\frac{\partial H}{\partial \theta} = 0,$$

and hence p_θ is a first integral. Since

$$dH = \left(-\frac{{p_\theta}^2}{mr^3} + u'(r)\right) dr + \frac{p_r}{m} dp_r + \frac{p_\theta}{mr^2} dp_\theta,$$

we see that dH and dp_θ are independent on the dense open set of $T^*\mathbb{R}^2$ formed by the points whose polar coordinates $(r, \theta, p_r, p_\theta)$ are well defined and do not verify

$$u'(r) - \frac{{p_\theta}^2}{mr^3} = p_r = 0$$

(i.e. are not on a circular orbit – cf. Exercise 7.17.4). Therefore this Hamiltonian is completely integrable.

PROPOSITION 7.6. *Let H be a completely integrable Hamiltonian with first integrals $F_1, \ldots, F_n$ in involution, independent in the dense open set $U \subset T^*M$, and such that $X_{F_1}, \ldots, X_{F_n}$ are complete on U. Then each nonempty level set*

$$L_f = \{p \in U : F_1(p) = f_1, \ldots, F_n(p) = f_n\}$$

is a submanifold of dimension n, invariant for the Hamiltonian flow of H, admitting a locally free action of $\mathbb{R}^n$ which is transitive on each connected component.

PROOF. All points in U are regular points of the map $F : U \to \mathbb{R}^n$ given by $F(\alpha) = (F_1(\alpha), \ldots, F_n(\alpha))$; therefore all nonempty level sets $L_f = F^{-1}(f)$ are submanifolds of dimension n.

Since $X_H \cdot F_i = 0$ for $i = 1, \ldots, n$, the level sets L_f are invariant for the flow of X_H. In fact, we have $X_{F_i} \cdot F_j = \{F_i, F_j\} = 0$, and hence these level sets are invariant for the flow of X_{F_i}. Moreover, these flows commute, as $[X_{F_i}, X_{F_j}] = X_{\{F_i,F_j\}} = 0$.

Consider the map $A : \mathbb{R}^n \times L_f \to L_f$ given by

$$A(t_1, \ldots, t_n, \alpha) = \psi_{1,t_1} \circ \ldots \circ \psi_{n,t_n}(\alpha),$$

where $\psi_{i,t} : L_f \to L_f$ is the flow of X_{F_i}. Since these flows commute, this map defines an action of $\mathbb{R}^n$ on L_f. On the other hand, for each $\alpha \in L_f$ the map $A_\alpha : \mathbb{R}^n \to L_f$ given by $A_\alpha(t_1, \ldots, t_n) = A(t_1, \ldots, t_n, \alpha)$ is a local diffeomorphism at the origin, as

$$(dA_\alpha)_0 (e_i) = \left.\frac{d}{dt}\right|_{t=0} \psi_{i,t}(\alpha) = (X_{F_i})_\alpha$$

and the vector fields X_{F_i} are linearly independent. Therefore the action is locally free, meaning that the isotropy groups are discrete. Also, the action is locally transitive, and hence transitive on each connected component. □

PROPOSITION 7.7. *Let Γ be a discrete subgroup of $\mathbb{R}^n$. Then there exist $k \in \{0, 1, \ldots, n\}$ linearly independ vectors $e_1, \ldots, e_k$ such that $\Gamma = \operatorname{span}_{\mathbb{Z}}\{e_1, \ldots, e_k\}$.*

PROOF. If $\Gamma = \{0\}$ then we are done. If not, let $e \in \Gamma \setminus \{0\}$. Since Γ is discrete, the set

$$\Gamma \cap \{\lambda e \mid 0 < \lambda \leq 1\}$$

is finite (and nonempty). Let e_1 be the element in this set which is closest to 0. Then

$$\Gamma \cap \operatorname{span}_{\mathbb{R}}\{e\} = \operatorname{span}_{\mathbb{Z}}\{e_1\}.$$

If $\Gamma = \operatorname{span}_{\mathbb{Z}}\{e_1\}$ then we are done. If not, let $e \in \Gamma \setminus \operatorname{span}_{\mathbb{Z}}\{e_1\}$. Then the set

$$\Gamma \cap \{\lambda e + \lambda_1 e_1 \mid 0 < \lambda, \lambda_1 \leq 1\}$$

is finite (and nonempty). Let e_2 be the element in this set which is closest to $\operatorname{span}_{\mathbb{R}}\{e_1\}$. Then

$$\Gamma \cap \operatorname{span}_{\mathbb{R}}\{e, e_1\} = \operatorname{span}_{\mathbb{Z}}\{e_1, e_2\}.$$

Iterating this procedure yields the result. □

PROPOSITION 7.8. *Let L_f^α be the connected component of $\alpha \in L_f$. Then L_f^p is diffeomorphic to $T^k \times \mathbb{R}^{n-k}$, where k is the number of generators of the isotropy subgroup Γ_α. In particular, if L_f^α is compact then it is diffeomorphic to the n-dimensional torus T^n.*

PROOF. Since the action $A : \mathbb{R}^n \times L_f^\alpha \to L_f^\alpha$ is transitive, the local diffeomorphism $A_\alpha : \mathbb{R}^n \to L_\alpha^p$ is surjective. On the other hand, because Γ_α is discrete, the action of Γ_α on $\mathbb{R}^n$ by translation is free and proper, and we can form the quotient $\mathbb{R}^n/\Gamma_\alpha$, which is clearly diffeomorphic to $T^k \times \mathbb{R}^{n-k}$. Finally, it is easily seen that A_α induces a diffeomorphism $\mathbb{R}^n/\Gamma_\alpha \cong L_f^\alpha$. □

DEFINITION 7.9. *A* **linear flow** *on the torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$ is the projection of the flow $\psi_t : \mathbb{R}^n \to \mathbb{R}^n$ given by*

$$\psi_t(x) = x + \nu t.$$

The **frequencies** *of the linear flow are the components $\nu^1, \ldots, \nu^n$ of ν.*

THEOREM 7.10. (Arnold-Liouville) *Let H be a completely integrable Hamiltonian with n first integrals $F_1, \ldots, F_n \in C^\infty(T^*M)$ in involution, independent on the dense open set $U \subset T^*M$. If the connected components of the level sets of the map $(F_1, \ldots, F_n) : U \to \mathbb{R}^n$ are compact then they are n-dimensional tori, invariant for the flow of X_H. The flow of X_H on these tori is a linear flow.*

PROOF. All that remains to be seen is that the flow of X_H on the invariant tori is a linear flow. It is clear that the flow of each X_{F_i} is linear. Since X_H is tangent to the invariant tori, we have $X_H = \sum_{i=1}^n f^i X_{F_i}$ for certain functions f^i. Now

$$0 = X_{\{F_i,H\}} = [X_{F_i}, X_H] = \sum_{j=1}^n (X_{F_i} \cdot f^j) X_{F_j},$$

and hence each function f^i is constant on the invariant torus. We conclude that the flow of X_H is linear. □

DEFINITION 7.11. *Let $\psi_t : T^n \to T^n$ be a linear flow. The* **time average** *of a function $f \in C^\infty(T^n)$ along ψ_y is the map*

$$\overline{f}(x) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T f(\psi_t(x))dt$$

(defined on the set of points $x \in T^n$ where the limit exists).

DEFINITION 7.12. *The frequencies $\nu \in \mathbb{R}^n$ of a linear flow $\psi_t : T^n \to T^n$ are said to be* **independent** *if they are linearly independent over $\mathbb{Q}$, i.e. if $\langle k, \nu \rangle \neq 0$ for all $k \in \mathbb{Z}^n \setminus \{0\}$.*

THEOREM 7.13. (Birkhoff) *If the frequencies $\nu \in \mathbb{R}^n$ of a linear flow $\psi_t : T^n \to T^n$ are independent then the time average of any function $f \in C^\infty(T^n)$ exists for all $x \in T^n$ and*

$$\overline{f}(x) = \int_{T^n} f.$$

PROOF. Since $T^n = \mathbb{R}^n/\mathbb{Z}^n$, the differentiable functions on the torus arise from periodic differentiable functions on $\mathbb{R}^n$, which can be expanded as Fourier series. Therefore it suffices to show that the theorem holds for $f(x) = e^{2\pi i \langle k, x \rangle}$ with $k \in \mathbb{Z}^n$.

If $k = 0$ then both sides of the equality are 1, and the theorem holds.

If $k \neq 0$, the right-hand side of the equality is zero, and the left-hand side is

$$\begin{aligned}\overline{f}(x) &= \lim_{T \to +\infty} \frac{1}{T} \int_0^T e^{2\pi i \langle k, x + \nu t \rangle} dt \\ &= \lim_{T \to +\infty} \frac{1}{T} e^{2\pi i \langle k, x \rangle} \frac{e^{2\pi i \langle k, \nu \rangle T} - 1}{2\pi i \langle k, \nu \rangle} = 0\end{aligned}$$

(where we used the fact that $\langle k, \nu \rangle \neq 0$). □

COROLLARY 7.14. *If the frequencies of a linear flow $\psi_t : T^n \to T^n$ are independent then $\{\psi_t(x) \mid t \geq 0\}$ is dense on the torus for all $x \in T^n$.*

PROOF. If $\{\psi_t(x) \mid t \geq 0\}$ were not dense then it would not intersect an open set $U \subset T^n$. Therefore any nonnegative function $f \in C^\infty(T^n)$ with nonempty support contained in U would satisfy $\overline{f}(x) = 0$ and $\int_{T^n} f > 0$, contradicting Birkhoff's Theorem. □

COROLLARY 7.15. *If the frequencies of a linear flow $\psi_t : T^n \to T^n$ are independent and $n \geq 2$ then $\psi_t(x)$ is not periodic.*

REMARK 7.16. The qualitative behaviour of the Hamiltonian flow generated by completely integrable Hamiltonians is completely understood. Complete integrability is however a very strong condition, not satisfied by generic Hamiltonians. The **Komolgorov-Arnold-Moser (KAM) Theorem** guarantees a small measure of genericity by establishing that a large fraction of the invariant tori of a completely integrable Hamiltonians survives under perturbation, the flow on these tori remaining linear with the same frequencies. On the other hand, many invariant tori, including those whose frequencies are not independ (**resonant tori**), are tipically destroyed.

EXERCISES 7.17.

(1) Show that if $F, G \in C^\infty(T^*M)$ are first integrals, then $\{F, G\}$ is also a first integral.
(2) Prove Proposition 7.3
(3) Consider a revolution surface $M \subset \mathbb{R}^3$ given in cylindrical coordinates (r, θ, z) by
$$r = f(z),$$
where $f : (a, b) \to (0, +\infty)$ is differentiable.
  (a) Show that the geodesics of M are the critical points of the action determined by the Lagrangian $L : TM \to \mathbb{R}$ given in local cordinates by
  $$L(\theta, z, \dot{\theta}, \dot{z}) = \frac{1}{2}\left((f(z))^2\dot{\theta}^2 + \left((f'(z))^2 + 1\right)\dot{z}^2\right).$$
  (b) Show that the curves given in local coordinates by $\theta =$ constant or $f'(z) = 0$ are images of geodesics.
  (c) Compute the Legendre tranformation, show that L is hyperregular and write an expression in local coordinates for the Hamiltonian $H : T^*M \to \mathbb{R}$.
  (d) Show that H is completely integrable.
  (e) Show that the projection on M of the invariant set
  $$L_{(E,l)} = H^{-1}(E) \cap {p_\theta}^{-1}(l)$$
  $(E, l > 0)$ is given in local coordinates by
  $$f(z) \geq \frac{l}{\sqrt{2E}}.$$
  Use this fact to conclude that if f has a strict local maximum at $z = z_0$ then the geodesic whose image is $z = z_0$ is **stable**, i.e. geodesic with initial condition close to $(\theta_0, z_0, 1, 0) \in TM$ stay close to $z = z_0$.
(4) Recall from Example 7.5 that a particle of mass $m > 0$ moving in a central field is described by the completely integrable Hamiltonian function
$$H\left(r, \theta, p_r, p_\theta\right) = \frac{{p_r}^2}{2m} + \frac{{p_\theta}^2}{2mr^2} + u(r).$$
  (a) Show that there exist circular orbits of radius r_0 whenever $u'(r_0) > 0$.

(b) Show that the set of points where dH and dp_θ are not independent is the unions of these circular orbits.

(c) Show that the projection of the invariant set

$$L_{(E,l)} = H^{-1}(E) \cap {p_\theta}^{-1}(l)$$

on $\mathbb{R}^2$ is given in local coordinates by

$$u(r) + \frac{l^2}{2r^2} \leq E.$$

(d) Conclude that if $u'(r_0) > 0$ and

$$u''(r_0) + \frac{3u'(r_0)}{r_0} > 0$$

the the circular orbit of radius r_0 is stable.

(5) In General Relativity, the motion of a particle in the gravitational field of a point mass $M > 0$ is given by the Lagrangian $L : TU \to \mathbb{R}$ written in cylindrical coordinates (u, r, θ) as

$$L(u, r, \theta, \dot{u}, \dot{r}, \dot{\theta}) = -\frac{1}{2}\left(1 - \frac{2M}{r}\right)\dot{u}^2 + \frac{1}{2}\left(1 - \frac{2M}{r}\right)^{-1}\dot{r}^2 + \frac{1}{2}r^2\dot{\theta}^2,$$

where $U \subset \mathbb{R}^3$ is the open set given by $r > 2M$ (the coordinate u is called the **time coordinate**, and in general is different from the **proper time** of the particle, i.e. the parameter t of the curve).

(a) Show that L is hyper-regular and compute the corresponding Hamiltoninan $H : T^*U \to \mathbb{R}$.

(b) Show that H is completely integrable.

(c) Show that there exist circular orbits of radius r_0 for any $r_0 > 2M$, with $H < 0$ for $r_0 > 3M$ (speed lower than the speed of light), $H = 0$ for $r_0 = 3M$ (speed equal to the speed of light) and $H > 0$ for $r_0 < 3M$ (speed higher than the speed of light).

(d) Show that the set of points where dH, dp_u and dp_θ are not independent is the unions of these circular orbits.

(e) Show that the projection of the invariant cylinder

$$L_{(E,k,l)} = H^{-1}(E) \cap {p_u}^{-1}(k) \cap {p_\theta}^{-1}(l)$$

on U is given in local coordinates by

$$\frac{l^2}{r^2} - \left(1 - \frac{2M}{r}\right)^{-1} k^2 \leq 2E.$$

(f) Conclude that if $r_0 > 6M$ then the circular orbit of radius r_0 is stable.

(6) Recall that the Lagrange top is the mechanical system determined by the Lagrangian $L : TSO(3) \to \mathbb{R}$ given in local coordinates by

$$L = \frac{I_1}{2}\left(\dot{\theta}^2 + \dot{\varphi}^2 \sin^2\theta\right) + \frac{I_3}{2}\left(\dot{\psi} + \dot{\varphi}\cos\theta\right)^2 - Mgl\cos\theta,$$

where (θ, φ, ψ) are the Euler angles, M is the top's mass and l is the distance from the fixed point to the center of mass.

(a) Compute the Legendre transformation, show that L is hyper-regular and write an expression in local coordinates for the Hamiltonian $H : T^*SO(3) \to \mathbb{R}$.
(b) Prove that H is completely integrable.
(c) Show that the solutions found in Exercise 3.20.13 traverse degenerate 2-dimensional tori. Use this fact to argue that these solutions are stable.

(7) Consider the sequence formed by the first digit of the decimal expansion of each of the integers 2^n for $n \in \mathbb{N}_0$:

$$1, 2, 4, 8, 1, 3, 6, 1, 2, 5, 1, 2, 4, 8, 1, 3, 6, 1, 2, 5, \ldots$$

The purpose of this exercise is to answer the following question: is there a 7 in this sequence?

(a) Show that if $\nu \in \mathbb{R} \setminus \mathbb{Z}$ then

$$\lim_{n \to +\infty} \frac{1}{n+1} \sum_{k=0}^{n} e^{2\pi i \nu k} = 0.$$

(b) Prove the following discrete version of Birkhoff's theorem: if a differentiable function $f : \mathbb{R} \to \mathbb{R}$ is periodic with period 1 and $\nu \in \mathbb{R} \setminus \mathbb{Q}$ then for all $x \in \mathbb{R}$

$$\lim_{n \to +\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(x + \nu k) = \int_0^1 f(x) dx.$$

(c) Show that $\log 2$ is an irrational multiple of $\log 10$.
(d) Is there a 7 in the sequence above?

8. Notes on Chapter 5

8.1. Bibliographical notes. The material in this chapter follows [**Oli02**] and [**Arn97**] closely. There are of course many other excellent books on Mechanics, both traditional [**GPS02**] and geometric [**AM78, MR99**]. Nonholonomic systems (including control theory) are treated in greater detail in [**Blo03, BL05**]. For more information on completely integrable systems see [**CB97, Aud96**].

CHAPTER 6

Relativity

In this chapter we study one of the most important applications of Riemannian geometry, namely **General Relativity**.

In Section 1 we discuss **Galileo spacetime**, the geometric structure underlying Newtonian mechanics, which hinges on the existence of arbitrarily fast motions; if, however, a maximum speed is assumed to exist, then it must be replaced by **Minkowski spacetime**, whose geometry is studied in **Special Relativity** (Section 2).

Section 3 shows how to include Newtonian gravity in Galileo spacetime by introducing the symmetric **Cartan connection**. By trying to generalize this procedure we are led to consider general **Lorentzian manifolds** satisfying the **Einstein field equation**, of which Minkowski spacetime is the simplest example (Section 4).

Other simple solutions are analyzed in the subsequent sections: the **Schwarzschild solution**, modeling the gravitational field outside spherically symmetric bodies or **black holes** (Section 5), and the **Friedmann-Robertson-Walker** models of **cosmology**, describing the behavior of the Universe as a whole (Section 6).

This chapter concludes with a discussion of the **causal structure** of a Lorentz manifold (Section 7), in preparation for the proof of one of the **Hawking-Penrose singularity theorems** (Section 8).

1. Galileo Spacetime

The set of all physical occurrences can be modeled as a connected 4-dimensional manifold M, which we call **spacetime**, and whose points we refer to as **events**. We assume that M is diffeomorphic to $\mathbb{R}^4$, and that there exists a special class of diffeomorphisms $x : M \to \mathbb{R}^4$, called **inertial frames**. An inertial frame yields global coordinates $(x^0, x^1, x^2, x^3) = (t, x, y, z)$. We call the coordinate $t : M \to \mathbb{R}$ the **time function** associated to a given inertial frame. Two events $p, q \in M$ are said to be **simultaneous** on that frame if $t(p) = t(q)$. The level functions of the time function are therefore called **simultaneity hypersurfaces**. The **distance** between two simultaneous events $p, q \in M$ is given by

$$d(p, q) = \sqrt{\sum_{i=1}^{3} (x^i(p) - x^i(q))^2}.$$

The motion of a particle is modeled by a smooth curve $c : I \to M$ such that $dt(\dot{c}) \neq 0$. A special class of motion is the motions of s **free particle**, i.e., a particle which is not acted upon by any external force. The special property that inertial frames have to satisfy is that the motions of a free particle is always represented by a straight line. In other words, free particles move with constant velocity relative to inertial frames (**Newton's law of inertia**). In particular, motions of particles at rest in an inertial frame are motions of free particles.

Inertial frames are not unique: if $x : M \to \mathbb{R}^4$ is an inertial frame and $T : \mathbb{R}^4 \to \mathbb{R}^4$ is an invertible affine transformation then $T \circ x$ is another inertial frame. In fact, any two inertial frames must be related by such an affine transformation (cf. Exercise 1.1.2).

The **Galileo spacetime**, which underlies Newtonian mechanics, is obtained by further requiring that inertial frames should:

(1) agree on the time interval between any two events (and hence on whether two given events are simultaneous);
(2) agree on the distance between simultaneous events.

Therefore, up to translations and reflections, all coordinate transformations between inertial frames belong to the **Galileo group** $Gal(4)$, the group of linear orientation-preserving maps which preserve time functions and the Euclidean structures of the simultaneity hypersurfaces.

When analyzing problems in which only one space dimension is important, we can use a simpler 2-dimensional Galileo spacetime. If (t, x) are the spacetime coordinates associated to an inertial frame and $T \in Gal(2)$ is a Galileo change of basis to a new inertial frame with global coordinates (t', x'), then

$$\begin{aligned} \frac{\partial}{\partial t'} &:= T\left(\frac{\partial}{\partial t}\right) = \frac{\partial}{\partial t} + v\frac{\partial}{\partial x} \\ \frac{\partial}{\partial x'} &:= T\left(\frac{\partial}{\partial x}\right) = \frac{\partial}{\partial x} \end{aligned}$$

with $v \in \mathbb{R}$, since we must have

$$dt\left(\frac{\partial}{\partial t'}\right) = dt'\left(\frac{\partial}{\partial t'}\right) = 1,$$

and we want the orientation-preserving map T to be an isometry of $\{t = 0\} \equiv \{t' = 0\}$. The change of basis matrix is then

$$S = \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix},$$

with inverse

$$S^{-1} = \begin{pmatrix} 1 & 0 \\ -v & 1 \end{pmatrix}.$$

Therefore the corresponding coordinate transformation is

$$\begin{cases} t' = t \\ x' = x - vt \end{cases} \qquad (v \in \mathbb{R})$$

(**Galileo transformation**), and hence the new frame is moving with velocity v with respect to the old one (as the curve $x' = 0$ is the curve $x = vt$). Notice that S^{-1} is obtained from S simply by reversing the sign of v, as one would expect, as the old frame must be moving relative to the new one with velocity $-v$. We shall call this observation the **Relativity Principle**.

EXERCISES 1.1.

(1) (*Lucas Problem*) By the late 19th century there existed a regular transatlantic service between Le Havre and New York. Every day at noon (GMT) a transatlantic ship would depart Le Havre and another one would depart New York. The journey took exactly seven days, so that arrival would also take place at noon (GMT). Therefore, a transatlantic ship traveling from Le Havre to New York would meet a transatlantic ship just arriving from New York at departure, and another one just leaving New York on arrival. Besides these, how many other ships would it meet? At what times? What was the total number of ships needed for this service? (**Hint:** Represent the ships' motions as curves in a 2-dimensional Galileo spacetime).

(2) Let $f : \mathbb{R}^n \to \mathbb{R}^n$ ($n \geq 2$) be a bijection that takes straight lines to straight lines. Show that f must be an affine function, i.e., that

$$f(x) = Ax + b$$

for all $x \in \mathbb{R}^n$, where $A \in GL(n, \mathbb{R})$ and $b \in \mathbb{R}^n$.

(3) Prove that the Galileo group $Gal(4)$ is the subset of $GL(4, \mathbb{R})$ formed by matrices of the form

$$\begin{pmatrix} 1 & 0 \\ v & R \end{pmatrix}$$

where $v \in \mathbb{R}^3$ and $R \in SO(3)$. Conclude that $Gal(4)$ is isomorphic to the group of orientation-preserving isometries of the Euclidean 3-space $\mathbb{R}^3$.

(4) Show that $Gal(2)$ is a subgroup of $Gal(4)$.

2. Special Relativity

The Galileo spacetime assumption that all inertial observers should agree on the time interval between two events is intimately connected with the possibility of synchronizing clocks in different frames using signals of arbitrarily high speeds. Experience reveals that this is actually impossible. Instead, there appears to be a maximum propagation speed, the speed of light, which is the same at all events and in all directions, and that we can therefore take to be 1 by choosing suitable units (for instance, measuring time in years and distance in light-years). Therefore a more accurate requirement is that any two inertial frames should

(1') agree on whether a given particle is moving at the speed of light.

Notice that we no longer require that different inertial frames should agree on the time interval between two events, or even if two given events are simultaneous. However we still require that any two inertial frames should

(2') agree on the distance between events which are simultaneous in **both** frames.

Fix a particular inertial frame with coordinates (x^0, x^1, x^2, x^3). A free particle moving at the speed of light will be a straight line whose tangent vector

$$v = v^0 \frac{\partial}{\partial x^0} + v^1 \frac{\partial}{\partial x^1} + v^2 \frac{\partial}{\partial x^2} + v^3 \frac{\partial}{\partial x^3}$$

must satisfy

$$(v^0)^2 = (v^1)^2 + (v^2)^2 + (v^3)^2$$

(since the distance travelled must equal the elapsed time). In other words, v must satisfy $\langle v, v \rangle = 0$, where

$$\langle v, w \rangle = -v^0 w^0 + v^1 w^1 + v^2 w^2 + v^3 w^3 = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} v^\mu w^\nu,$$

with $(\eta_{\mu\nu}) = \operatorname{diag}(-1, 1, 1, 1)$. Notice that $\langle \cdot, \cdot \rangle$ is a symmetric non-degenerate tensor which is not positive definite; we call it the **Minkowski (pseudo) inner product**. The coordinate basis

$$\left\{ \frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right\}$$

is an orthonormal basis for this inner product (cf. Exercise 2.2.1), as

$$\left\langle \frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu} \right\rangle = \eta_{\mu\nu}$$

$(\mu, \nu = 0, 1, 2, 3)$.

By assumption (1'), given a motion of a free particle at the speed of light, all inertial observers must agree that the particle is moving at this (maximum) speed. Therefore, if $(x^{0\prime}, x^{1\prime}, x^{2\prime}, x^{3\prime})$ are the coordinates associated to another inertial frame, the vectors

$$\frac{\partial}{\partial x^{0\prime}} \pm \frac{\partial}{\partial x^{i\prime}}$$

$(i = 1, 2, 3)$ must be tangent to a motion at the speed of light, i.e.,

$$\left\langle \frac{\partial}{\partial x^{0\prime}} \pm \frac{\partial}{\partial x^{i\prime}}, \frac{\partial}{\partial x^{0\prime}} \pm \frac{\partial}{\partial x^{i\prime}} \right\rangle = 0.$$

This implies that

$$\left\langle \frac{\partial}{\partial x^{0\prime}}, \frac{\partial}{\partial x^{0\prime}} \right\rangle = - \left\langle \frac{\partial}{\partial x^{i\prime}}, \frac{\partial}{\partial x^{i\prime}} \right\rangle ;$$
$$\left\langle \frac{\partial}{\partial x^{0\prime}}, \frac{\partial}{\partial x^{i\prime}} \right\rangle = 0.$$

Similarly, we must have

$$\left\langle \sqrt{2}\frac{\partial}{\partial x^{0\prime}} + \frac{\partial}{\partial x^{i\prime}} + \frac{\partial}{\partial x^{j\prime}}, \sqrt{2}\frac{\partial}{\partial x^{0\prime}} + \frac{\partial}{\partial x^{i\prime}} + \frac{\partial}{\partial x^{j\prime}} \right\rangle = 0$$

$(i \neq j)$, and hence

$$\left\langle \frac{\partial}{\partial x^{i\prime}}, \frac{\partial}{\partial x^{j\prime}} \right\rangle = 0.$$

Since $\langle \cdot, \cdot \rangle$ is non-degenerate, we conclude that there must exist $k \neq 0$ such that

$$\left\langle \frac{\partial}{\partial x^{\mu\prime}}, \frac{\partial}{\partial x^{\nu\prime}} \right\rangle = k\eta_{\mu\nu}$$

$(\mu, \nu = 0, 1, 2, 3)$.

The simultaneity hypersurfaces $\{x^0 = \text{const.}\}$ and $\{x^{0\prime} = \text{const.}\}$ are 3-planes in $\mathbb{R}^4$. If they are parallel, they coincide; otherwise, they must intersect along 2-planes of events which are simultaneous in both frames. Let $v \neq 0$ be a vector tangent to one of these 2-planes. Then $dx^0(v) = dx^{0\prime}(v) = 0$, and hence

$$v = \sum_{i=1}^{3} v^i \frac{\partial}{\partial x^i} = \sum_{i=1}^{3} v^{i\prime} \frac{\partial}{\partial x^{i\prime}}.$$

By assumption (2'), we must have

$$\sum_{i=1}^{3} \left(v^i\right)^2 = \sum_{i=1}^{3} \left(v^{i\prime}\right)^2.$$

Consequently, from

$$\sum_{i=1}^{3} \left(v^i\right)^2 = \langle v, v \rangle = \left\langle \sum_{i=1}^{3} v^{i\prime} \frac{\partial}{\partial x^{i\prime}}, \sum_{i=1}^{3} v^{i\prime} \frac{\partial}{\partial x^{i\prime}} \right\rangle = k \sum_{i=1}^{3} \left(v^{i\prime}\right)^2$$

we conclude that we must have $k = 1$. Therefore the coordinate basis

$$\left\{ \frac{\partial}{\partial x^{0\prime}}, \frac{\partial}{\partial x^{1\prime}}, \frac{\partial}{\partial x^{2\prime}}, \frac{\partial}{\partial x^{3\prime}} \right\}$$

must also be an orthonormal basis. In particular, this means that the Minkowski inner product $\langle \cdot, \cdot \rangle$ is well defined (i.e., is independent of the inertial frame we choose to define it), and that we can identify inertial frames with orthonormal bases of $(\mathbb{R}^4, \langle \cdot, \cdot \rangle)$.

Definition 2.1. $(\mathbb{R}^4, \langle\cdot,\cdot\rangle)$ *is said to be the* **Minkowski spacetime**. *The* **length** *of a vector* $v \in \mathbb{R}^4$ *is* $|v| = |\langle v, v\rangle|^{\frac{1}{2}}$.

The study of the geometry of Minkowski spacetime is usually called **Special Relativity**. A vector $v \in \mathbb{R}^4$ is said to be:

(1) **timelike** if $\langle v, v\rangle < 0$; in this case, there exists an inertial frame $(x^{0\prime}, x^{1\prime}, x^{2\prime}, x^{3\prime})$ such that
$$v = |v|\frac{\partial}{\partial x^{0\prime}}$$
(cf. Exercise 2.2.1), and consequently any two events p and $p+v$ occur on the same location in this frame, separated by a time interval $|v|$;

(2) **spacelike** if $\langle v, v\rangle > 0$; in this case, there exists an inertial frame $(x^{0\prime}, x^{1\prime}, x^{2\prime}, x^{3\prime})$ such that
$$v = |v|\frac{\partial}{\partial x^{1\prime}}$$
(cf. Exercise 2.2.1), and consequently any two events p and $p+v$ occur simultaneously in this frame, a distance $|v|$ apart;

(3) **lightlike**, or **null**, if $\langle v, v\rangle = 0$; in this case any two events p and $p+v$ are connected by a motion at the speed of light in **any** inertial frame.

The set of all null vectors is called the **light cone**, and in a way is the structure that replaces the absolute simultaneity hypersurfaces of Galileo spacetime. It is the boundary of the set of all timelike vectors, which has two connected components; we represent by $C(v)$ the connected component of a given timelike vector v. A **time orientation** for Minkowski spacetime is a choice of one of these components, whose elements are said to be **future-pointing**; this is easily extended to nonzero null vectors.

An inertial frame (x^0, x^1, x^2, x^3) determines a time orientation, namely that for which the future-pointing timelike vectors are the elements of $C\left(\frac{\partial}{\partial x^0}\right)$. Up to translations and reflections, all coordinate transformations between inertial frames belong to the **(proper) Lorentz group** $SO_0(3,1)$, the group of linear maps which preserve orientation, time orientation and the Minkowski inner product (hence the light cone).

A curve $c : I \subset \mathbb{R} \to \mathbb{R}^4$ is said to be **timelike** if $\langle \dot{c}, \dot{c}\rangle < 0$. Timelike curves represent motions of particles with nonzero mass, since only for these curves is it possible to find an inertial frame in which the particle is instantaneously at rest. In other words, massive particles must always move at less than the speed of light (cf. Exercise 2.2.13). The **proper time** measured by the particle between events $c(a)$ and $c(b)$ is
$$\tau(c) := \int_a^b |\dot{c}(s)|ds.$$

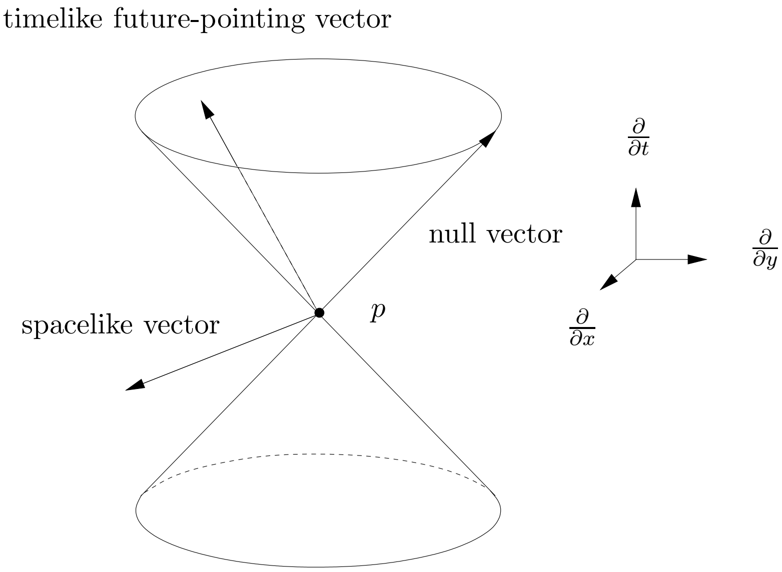

FIGURE 1. Minkowski geometry (it is traditional represented with the t-axis pointing upwards).

When analyzing problems in which only one space dimension is important, we can use a simpler 2-dimensional Minkowski spacetime. If (t, x) are the spacetime coordinates associated to an inertial frame and $T \in SO_0(1,1)$ is a Lorentzian change of basis to a new inertial frame with global coordinates (t', x'), we must have

$$\frac{\partial}{\partial t'} := T\left(\frac{\partial}{\partial t}\right) = \cosh u \frac{\partial}{\partial t} + \sinh u \frac{\partial}{\partial x}$$
$$\frac{\partial}{\partial x'} := T\left(\frac{\partial}{\partial x}\right) = \sinh u \frac{\partial}{\partial t} + \cosh u \frac{\partial}{\partial x}$$

with $u \in \mathbb{R}$ (cf. Exercise 2.2.3). The change of basis matrix is

$$S = \begin{pmatrix} \cosh u & \sinh u \\ \sinh u & \cosh u \end{pmatrix},$$

with inverse

$$S^{-1} = \begin{pmatrix} \cosh u & -\sinh u \\ -\sinh u & \cosh u \end{pmatrix}.$$

Therefore the corresponding coordinate transformation is

$$\begin{cases} t' = t \cosh u - x \sinh u \\ x' = x \cosh u - t \sinh u \end{cases}$$

(**Lorentz transformation**), and hence the new frame is moving with velocity $v = \tanh u$ with respect to the old one (as the curve $x' = 0$ is the curve $x = vt$; notice that $|v| < 1$). The matrix S^{-1} is obtained from S simply by reversing the sign of u, or, equivalently, of v; therefore, the Relativity Principle still holds for Lorentz transformations.

Moreover, since

$$\cosh u = \left(1 - v^2\right)^{-\frac{1}{2}};$$
$$\sinh u = v\left(1 - v^2\right)^{-\frac{1}{2}},$$

one can also write the Lorentz transformation as

$$\begin{cases} t' = \left(1 - v^2\right)^{-\frac{1}{2}} t - v\left(1 - v^2\right)^{-\frac{1}{2}} x \\ x' = \left(1 - v^2\right)^{-\frac{1}{2}} x - v\left(1 - v^2\right)^{-\frac{1}{2}} t \end{cases}.$$

In everyday life situations, we deal with frames whose relative speed is much smaller that the speed of light, $|v| \ll 1$, and with events for which $|x| \ll |t|$ (distances traveled by particles in one second are much smaller that 300,000 kilometers). An approximate expression for the Lorentz transformations in these situations is then

$$\begin{cases} t' = t \\ x' = x - vt \end{cases}$$

which is just a Galileo transformation. In other words, the Galileo group is a convenient low-speed approximation of the Lorentz group.

Suppose that two distinct events p and q occur in the same spatial location in the inertial frame (t', x'),

$$q - p = \Delta t' \frac{\partial}{\partial t'} = \Delta t' \cosh u \frac{\partial}{\partial t} + \Delta t' \sinh u \frac{\partial}{\partial x} = \Delta t \frac{\partial}{\partial t} + \Delta x \frac{\partial}{\partial x}.$$

We see that the time separation between the two events in a different inertial frame (t, x) is **bigger**,

$$\Delta t = \Delta t' \cosh u > \Delta t'.$$

Loosely speaking, moving clocks run slower when compared to stationary ones (**time dilation**).

If, on the other hand, two distinct events p and q occur simultaneously in the inertial frame (t', x'),

$$q - p = \Delta x' \frac{\partial}{\partial x'} = \Delta x' \sinh u \frac{\partial}{\partial t} + \Delta x' \cosh u \frac{\partial}{\partial x} = \Delta t \frac{\partial}{\partial t} + \Delta x \frac{\partial}{\partial x},$$

then they will **not** be simultaneous in the inertial frame (t, x), where the time difference between them is

$$\Delta t = \Delta x' \sinh u \neq 0$$

(**relativity of simultaneity**).

Finally, consider two particles at rest in the inertial frame (t', x'). Their motions are the lines $x' = x'_0$ and $x' = x'_0 + l'$. In the inertial frame (t, x), these lines have equations

$$x = \frac{x'_0}{\cosh u} + vt \quad \text{ and } \quad x = \frac{x'_0 + l'}{\cosh u} + vt,$$

which describe motions of particles moving with velocity v and separated by a distance

$$l = \frac{l'}{\cosh u} < l'.$$

Loosely speaking, moving objects shrink in the direction of their motion (**length contraction**).

EXERCISES 2.2.

(1) Let $\langle\cdot,\cdot\rangle$ be a nondegenerate symmetric 2-tensor on an n-dimensional vector space V. Show that there always exists an **orthonormal basis** $\{v_1,\ldots,v_n\}$, i.e. a basis such that $\langle v_i, v_j\rangle = \varepsilon_{ij}$, where $\varepsilon_{ii} = \pm 1$ and $\varepsilon_{ij} = 0$ for $i \neq j$. Moreover, show that $s = \sum_{i=1}^{n} \varepsilon_{ii}$ (known as the **signature** of $\langle\cdot,\cdot\rangle$) does not depend on the choice of orthonormal basis.

(2) Consider the Minkowski inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^4$ with a given time orientation.
   (a) Let $v \in \mathbb{R}^4$ be timelike and future-pointing. Show that:
      (i) if $w \in \mathbb{R}^4$ is timelike or null and future-pointing then $\langle v, w\rangle < 0$;
      (ii) if $w \in \mathbb{R}^4$ is timelike or null and future-pointing then $v + w$ is timelike and future-pointing;
      (iii) $\{v\}^\perp = \{w \in \mathbb{R}^4 \mid \langle v, w\rangle = 0\}$ is a hyperplane containing only spacelike vectors (and the zero vector).
   (b) Let $v \in \mathbb{R}^4$ be null and future-pointing. Show that:
      (i) if $w \in \mathbb{R}^4$ is timelike or null and future-pointing then $\langle v, w\rangle \leq 0$, with equality iff $w = \lambda v$ for some $\lambda > 0$;
      (ii) if $w \in \mathbb{R}^4$ is timelike or null and future-pointing then $v + w$ is timelike or null and future-pointing, being null iff $w = \lambda v$ for some $\lambda > 0$;
      (iii) $\{v\}^\perp$ is a hyperplane containing only spacelike and null vectors, all of which are multiples of v.
   (c) Let $v \in \mathbb{R}^4$ be spacelike. Show that $\{v\}^\perp$ is a hyperplane containing timelike, null and spacelike vectors.

(3) Show that if (t, x) are the spacetime coordinates associated to an inertial frame and $T \in SO_0(1,1)$ is a Lorentzian change of basis to a new inertial frame with global coordinates (t', x'), we must have

$$\frac{\partial}{\partial t'} = T\left(\frac{\partial}{\partial t}\right) = \cosh u \frac{\partial}{\partial t} + \sinh u \frac{\partial}{\partial x}$$
$$\frac{\partial}{\partial x'} = T\left(\frac{\partial}{\partial x}\right) = \sinh u \frac{\partial}{\partial t} + \cosh u \frac{\partial}{\partial x}$$

for some $u \in \mathbb{R}$.

(4) (*Twin Paradox*) Twins Alice and Bob separate on their 20^{th} anniversary: while Alice stays on Earth (which is approximately an inertial frame), Bob leaves at 80% of the speed of light towards a

planet 8 light-years away from Earth, which he therefore reaches 10 years later (as measured in Earth's frame). After a short stay, Bob returns to Earth, again at 80% of the speed of light. Consequently, Alice is 40 years old when they meet again.

   (a) How old is Bob at this meeting?
   (b) How do you explain the asymmetry in the twin's ages? Notice that, from Bob's point of view, he is the one who is stationary, while the Earth moves away and back again.
   (c) Imagine that each twin has a very powerful telescope. What does each of them **see**? In particular, how much time elapses for each of them as they see their twin experiencing one year? (**Hint:** Notice that light rays are represented by null lines, i.e. lines whose tangent vector is null; therefore, if event p in Alice's history is seen by Bob at event q then there must exist a future-directed null line connecting p to q).

(5) (*Car and Garage Paradox*) A 5-meter long car moves at 80% of light speeed towards a 4-meter long garage with doors at both ends.
   (a) Compute the length of the car in the garage's frame, and show that if the garage doors are closed at the right time the car will be completely inside the garage for a few moments.
   (b) Compute the garage's length in the car's frame, and show that in this frame the car is never completely inside the garage. How do you explain this apparent contradiction?

(6) Let (t', x') be an inertial frame moving with velocity v with respect to the inertial frame (t, x). Prove the **velocity addition formula**: if a particle moves with velocity w' in the frame (t', x'), the particle's velocity in the frame (t, x) is

$$w = \frac{w' + v}{1 + w'v}.$$

What happens when $w' = \pm 1$?

(7) (*Hyperbolic angle*)
   (a) Show that
      (i) $\mathfrak{so}(1,1) = \left\{ \begin{pmatrix} 0 & u \\ u & 0 \end{pmatrix} \mid u \in \mathbb{R} \right\}$;
      (ii) $\exp \begin{pmatrix} 0 & u \\ u & 0 \end{pmatrix} = \begin{pmatrix} \cosh u & \sinh u \\ \sinh u & \cosh u \end{pmatrix} = S(u)$;
      (iii) $S(u)S(u') = S(u + u')$.
   (b) Consider the Minkowski inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^2$ with a given time orientation. If $v, w \in \mathbb{R}^2$ are unit timelike future-pointing vectors then there exists a unique $u \in \mathbb{R}$ such that $w = S(u)v$ (which we call the **hyperbolic angle** between v and w). Show that:
      (i) $|u|$ is the length of the curve formed by all unit timelike vectors between v and w;

(ii) $\frac{1}{2}|u|$ is the area of the region swept by the position vector of the curve above;
(iii) hyperbolic angles are additive;
(iv) the velocity addition formula of Exercise 6 is simply the formula for the hyperbolic tangent of a sum.

(8) (*Generalized Twin Paradox*) Let $p, q \in \mathbb{R}^4$ be two events connected by a timelike straight line l. Show that the proper time between p and q measured along l is bigger than the proper time between p and q measured along any other timelike curve connecting these two events. In other words, if an inertial observer and a (necessarily) accelerated observer separate at a given event and are rejoined at a later event, then the inertial observer always measures a bigger (proper) time interval between the two events. In particular, prove the **reversed triangle inequality**: if $v, w \in \mathbb{R}^4$ are timelike vectors with $w \in C(v)$ then $|v + w| \geq |v| + |w|$.

(9) (*Doppler effect*) Use the spacetime diagram in Figure 2 to show that an observer moving with velocity v away from a source of light of period T measures the period to be

$$T' = T\sqrt{\frac{1+v}{1-v}}$$

(**Remark:** This effect allows astronomers to measure the radial velocity of stars and galaxies relative to the Earth).

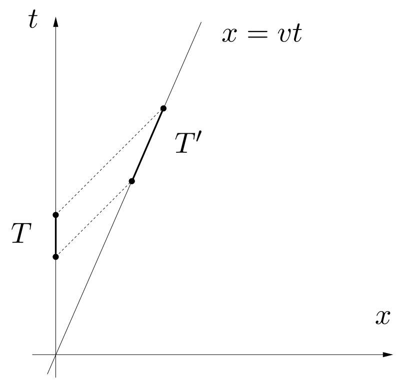

FIGURE 2. Doppler effect.

(10) (*Aberration*) Suppose that the position in the sky of the star Sirius makes an angle θ with the x-axis of a given inertial observer. Show that the angle θ' measured by a second inertial observer moving with velocity $v = \tanh u$ along the x-axis of the first observer satisfies

$$\tan\theta' = \frac{\sin\theta}{\cosh u\cos\theta + \sinh u}.$$

(11) Minkowski geometry can be used in many contexts. For instance, let $l = \mathbb{R}\frac{\partial}{\partial t}$ represent the motion of an observer at rest in the atmosphere and choose units such that the speed of sound is 1.

(a) Let $\tau : \mathbb{R}^4 \to \mathbb{R}$ the map such that $\tau(p)$ is the t coordinate of the event in which the observer hears the sound generated at p. Show that the level surfaces of τ are the conical surfaces

$$\tau^{-1}(t_0) = \left\{ p \in \mathbb{R}^4 \mid t_0 \frac{\partial}{\partial t} - p \text{ is null and future-pointing} \right\}.$$

(b) Show that $c : I \to \mathbb{R}^4$ represents the motion of a supersonic particle iff

$$\left\langle \dot{c}, \frac{\partial}{\partial t} \right\rangle < 0 \quad \text{ and } \quad \langle \dot{c}, \dot{c} \rangle > 0.$$

(c) Argue that the observer hears a sonic boom whenever c is tangent to a surface $\tau =$ constant. Assuming that c is a straight line, what does the observer hear before and after the boom?

(12) Let $c : \mathbb{R} \to \mathbb{R}^4$ be the motion of a particle in Minkowski spacetime parametrized by the proper time τ.

(a) Show that

$$\langle \dot{c}, \dot{c} \rangle = -1$$

and

$$\langle \dot{c}, \ddot{c} \rangle = 0.$$

Conclude that $\ddot{c}$ is the particle's acceleration as measured in the particle's **instantaneous rest frame**, i.e., in the inertial frame (t, x, y, z) for which $\dot{c} = \frac{\partial}{\partial t}$. For this reason, $\ddot{c}$ is called the particle's **proper acceleration**, and $|\ddot{c}|$ is interpreted as the acceleration measured by the particle.

(b) Compute the particles's motion assuming that it is moving along the x-axis with constant proper acceleration $|\ddot{c}| = a$.

(c) Consider a spaceship launched from Earth towards the center of the Galaxy (at a distance of 30,000 light-years) with $a = g$, where g represents the gravitational acceleration at the surface of the Earth. Using the fact that $g \simeq 1$ year^{-1} in units such that $c = 1$, compute the proper time measured aboard the spaceship for this journey. How long would the journey take as measured from Earth?

(13) (*The faster-than-light missile*) While conducting a surveillance mission on the home planet of the wicked Klingons, the *Enterprise* uncovers their evil plan to build a faster-than-light missile and attack Earth, 12 light-years away. Captain Kirk immediately orders the *Enterprise* back to Earth at its top speed ($\frac{12}{13}$ of the speed of light), and at the same time sends out a radio warning. Unfortunately, it is too late: eleven years later (as measured by them), the Klingons

launch their missile, moving at 12 times the speed of light. Therefore the radio warning, traveling at the speed of light, reaches Earth at the same time as the missile, twelve years after its emission, and the *Enterprise* arrives on the ruins of Earth one year later.

(a) How long does the *Enterprise* trip take according to its crew?

(b) On Earth's frame, let $(0,0)$ be the (t,x) coordinates of the event in which the *Enterprise* discovers the Klingon plan, $(11,0)$ the coordinates of the missile's launch, $(12,12)$ the coordinates of Earth's destruction and $(13,12)$ the coordinates of the *Enterprise*'s arrival on Earth's ruins. Compute the (t',x') coordinates of the same events on the *Enterprise*'s frame.

(c) Plot the motions of the *Enterprise*, the Klingon planet, Earth, the radio signal and the missile on *Enterprise*'s frame. Does the missile motion according to the *Enterprise* crew make sense?

3. The Cartan Connection

Let $(x^0,x^1,x^2,x^3)=(t,x,y,z)$ be an inertial frame on Galileo spacetime, which we can therefore identify with $\mathbb{R}^4$. Recall that Newtonian gravity is described by a **gravitational potential** $\Phi:\mathbb{R}^4\to\mathbb{R}$. This potential determines the motions of free-falling particles through

$$\frac{d^2x^i}{dt^2}=-\frac{\partial\Phi}{\partial x^i}$$

$(i=1,2,3)$, and is in turn determined by the **matter density function** $\rho:\mathbb{R}^4\to\mathbb{R}$ through the **Poisson equation**

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}=4\pi\rho$$

(we are using units in which Newton's universal gravitation constant G is set equal to 1). The vacuum Poisson equation (corresponding to the case in which all matter is concentrated on singularities of the field) is the well known **Laplace equation**

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}=0.$$

Notice that the equation of motion is the same for all particles, irrespective of their mass. This observation, dating back to Galileo, was made into the so-called **Equivalence Principle** by Einstein. Thus a gravitational field determines special curves on Galileo spacetime, namely the motions of free-falling particles. These curves are the geodesics of a symmetric connection, known as the **Cartan connection**, defined through the nonvanishing Christoffel symbols

$$\Gamma^i_{00}=\frac{\partial\Phi}{\partial x^i}$$

(cf. Exercise 3.1.1), corresponding to the nonvanishing connection forms

$$\omega_0^i = \frac{\partial \Phi}{\partial x^i} dt.$$

Cartan's structure equations

$$\Omega_\nu^\mu = d\omega_\nu^\mu + \sum_{\alpha=0}^{3} \omega_\alpha^\mu \wedge \omega_\nu^\alpha$$

still hold for this connection (cf. Exercise 2.8.2 in Chapter 4), and hence we have the nonvanishing curvature forms

$$\Omega_0^i = \sum_{j=1}^{3} \frac{\partial^2 \Phi}{\partial x^j \partial x^i} dx^j \wedge dt.$$

The Ricci curvature tensor of this connection is

$$Ric = \left(\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} \right) dt \otimes dt$$

(cf. Exercise 3.1.2), and hence the Poisson equation can be written as

$$Ric = 4\pi\rho \; dt \otimes dt.$$

In particular, the Laplace equation can be written as

$$Ric = 0.$$

EXERCISES 3.1.

(1) Check that the motions of free-falling particles are indeed geodesics of the Cartan connection. What other geodesics are there? How would you interpret them?
(2) Check the formula for the Ricci curvature tensor of the Cartan connection.
(3) Show that the Cartan connection ∇ is compatible with Galileo structure, i.e., show that
   (a) $\nabla_X dt = 0$ for all $X \in \mathfrak{X}(\mathbb{R}^4)$ (cf. Exercise 2.6.3 in Chapter 3).
   (b) If $E, F \in \mathfrak{X}(\mathbb{R}^4)$ are tangent to the simultaneity hypersurfaces and parallel along some curve $c : \mathbb{R} \to \mathbb{R}^4$, then $\langle E, F \rangle$ is constant.
(4) Show that the Cartan connection is not the Levi-Civita connection of any pseudo-Riemannian metric on $\mathbb{R}^4$ (cf. Section 4).

4. General Relativity

Gravity can be introduced in Newtonian mechanics through the symmetric Cartan connection, which preserves Galileo spacetime structure. A natural idea for introducing gravity in Special Relativity is then to searching for symmetric connections preserving the Minkowski inner product. To formalize this, we introduce the following

Definition 4.1. *A* **pseudo-Riemannian manifold** *is a pair* (M, g)*, where* M *is a connected* n*-dimensional differentiable manifold and* g *is a symmetric nondegenerate differentiable* 2*-tensor field (*g *is said to be a* **pseudo-Riemannian metric** *in* M*). The* **signature** *of a pseudo-Riemannian manifold is just the signature of* g *at any tangent space. A* **Lorentzian manifold** *is a pseudo-Riemannian manifold with signature* $n-2$*.*

The Minkowski spacetime $(\mathbb{R}^4, \langle\cdot,\cdot\rangle)$ is obviously a Lorentzian manifold. It is easily seen that the Levi-Civita Theorem still holds for pseudo-Riemannian manifolds: given a pseudo-Riemannian manifold (M, g) there exists a unique symmetric connection ∇ which is compatible with g (given by the Koszul formula). Therefore there exists just one symmetric connection preserving the Minkowski metric: the trivial connection (obtained in Cartesian coordinates by taking all Christoffel symbols equal to zero), whose geodesics are straight lines.

To introduce gravity through a symmetric connection we must therefore consider more general 4-dimensional Lorentzian manifolds, which we will still call **spacetimes**. These are no longer required to be diffeomorphic to $\mathbb{R}^4$, or to have inertial charts. The study of the geometry of these spacetimes is usually called **General Relativity**.

Each spacetime comes equipped with its unique Levi-Civita connection, and hence with its geodesics. If $c : I \subset \mathbb{R} \to M$ is a geodesic, then $\langle\dot{c}, \dot{c}\rangle$ is constant, as

$$\frac{d}{ds}\langle\dot{c}(s), \dot{c}(s)\rangle = 2\left\langle\frac{D\dot{c}}{ds}(s), \dot{c}(s)\right\rangle = 0.$$

A geodesic is called **timelike**, **null**, or **spacelike** according to whether $\langle\dot{c}, \dot{c}\rangle < 0$, $\langle\dot{c}, \dot{c}\rangle = 0$ or $\langle\dot{c}, \dot{c}\rangle > 0$ (i.e. according to whether its tangent vector is timelike, spacelike or null). By analogy with the Cartan connection, we will take timelike geodesics to represent the free-falling motions of massive particles. This ensures that the Equivalence Principle holds. Null geodesics will be taken to represent the motions of light rays.

In general, any curve $c : I \subset \mathbb{R} \to M$ is said to be **timelike** if $\langle\dot{c}, \dot{c}\rangle < 0$. In this case, c represents the motion of a particle with nonzero mass (which is accelerating unless c is a geodesic). The **proper time** measured by the particle between events $c(a)$ and $c(b)$ is

$$\tau(c) = \int_a^b |\dot{c}(s)| ds.$$

To select physically relevant spacetimes we must impose some sort of constraint. By analogy with the formulation of the Laplace equation in terms of the Cartan connection, we make the following

Definition 4.2. *We say that the Lorentzian manifold* (M, g) *is a* **vacuum solution** *of the* **Einstein field equation** *if its Levi-Civita connection satisfies* $Ric = 0$*.*

The general Einstein field equation is

$$Ric = 8\pi T,$$

where T is the so-called **reduced energy-momentum tensor** of the matter content of the spacetime. The simplest model of such a matter content is that of a pressureless perfect fluid, which is described by a **rest density function** $\rho \in C^\infty(M)$ and a unit **velocity vector field** $U \in \mathfrak{X}(M)$ (whose integral lines are the motions of the fluid particles). The reduced energy-momentum tensor for this matter model turns out to be

$$T = \rho\left(\nu \otimes \nu + \frac{1}{2}g\right),$$

where $\nu \in \Omega^1(M)$ is the 1-form associated to U by the metric g. Consequently, the Einstein field for this matter model is

$$Ric = 4\pi\rho(2\nu \otimes \nu + g)$$

(compare this to Poisson's equation in terms of the Cartan connection).

It turns out that spacetimes satisfying the Einstein field equation model astronomical phenomena with great accuracy.

Exercises 4.3.

(1) Show that the signature of a pseudo-Riemannian manifold (M, g) is well defined, i.e., show that the signature of $g_p \in \mathcal{T}^2(T_pM)$ does not depend on $p \in M$.
(2) Let (M, g) be a pseudo-Riemannian manifold and $f : N \to M$ an immersion. Show that f^*g is not necessarily a pseudo-Riemannian metric on N.
(3) Let (M, g) be the $(n+1)$-dimensional Minkowski spacetime, i.e., $M = \mathbb{R}^{n+1}$ and

$$g = -dx^0 \otimes dx^0 + dx^1 \otimes dx^1 + \ldots + dx^n \otimes dx^n.$$

Let

$$N = \{v \in M : \langle v, v\rangle = -1 \text{ and } v^0 > 0\},$$

and $i : N \to M$ the inclusion map. Show that (N, i^*g) is the n-dimensional hyperbolic space H^n.
(4) Let $c : I \subset \mathbb{R} \to \mathbb{R}^4$ be a timelike curve in Minkowski space parametrized by the proper time, $U = \dot{c}$ the tangent unit vector and $A = \ddot{c}$ the proper acceleration. A vector field $V : I \to \mathbb{R}^4$ is said to be **Fermi-Walker transported** along c if

$$\frac{DV}{d\tau} = \langle V, A\rangle U - \langle V, U\rangle A.$$

(a) Show that U is Fermi-Walker transported along c.
(b) Show that if V and W are Fermi-Walker transported along c then $\langle V, W\rangle$ is constant.

(c) If $\langle V, U\rangle = 0$ then V is tangent at U to the submanifold

$$N = \{v \in \mathbb{R}^4 : \langle v, v\rangle = -1 \text{ and } v^0 > 0\},$$

which is isometric to the hyperbolic 3-space. Show that in this case V is Fermi-Walker transported **iff** it is parallel transported along $U : I \to N$.

(d) Assume that c describes a circular motion with constant speed v and $\langle V, U\rangle = 0$. Compute the angle by which V varies (or **precesses**) after one revolution. (**Remark:** It is possible to prove that the angular momentum vector of a spinning particle is Fermi-Walker transported along its motion and orthogonal to it; the above precession, which has been observed for spinning particles such as electrons, is called the **Thomas precession**).

(5) (*Twin Paradox on a Cylinder)* Consider the vacuum solution of the Einstein field equation obtained by quotienting Minkowski spacetime by the discrete isometry group generated by the translation $\xi : \mathbb{R}^4 \to \mathbb{R}^4$ defined by $\xi(t, x, y, z) = (t, x + 8, y, z)$. Assume that Earth's motion is represented by the line $x = y = z = 0$, and that once again as Bob turns 20 he leaves his twin sister Alice on Earth and departs at 80% of the speed of light along the x-axis. Because of the topology of space, the two twins meet again after 10 years (as measured on Earth), *without Bob ever having accelerated.*

(a) Compute the age of each twin in their meeting.

(b) From Bob's viewpoint, it is the Earth which moves away from him. How do you explain the asymmetry in the twins' ages?

(6) (*Rotating frame*)

(a) Show that the metric of Minkowski spacetime can be written as

$$g = -dt \otimes dt + dr \otimes dr + r^2 d\theta \otimes d\theta + dz \otimes dz$$

by using cylindrical coordinates (r, θ, z) in $\mathbb{R}^3$.

(b) Let $\omega > 0$ and consider the coordinate change given by $\theta = \theta' + \omega t$. Show that in these coordinates the metric is written as

$$\begin{aligned} g = &- (1 - \omega^2 r^2) dt \otimes dt + \omega r^2 dt \otimes d\theta' + \omega r^2 d\theta' \otimes dt \\ &+ dr \otimes dr + r^2 d\theta' \otimes d\theta' + dz \otimes dz. \end{aligned}$$

(c) Show that in the region $U = \{r < \frac{1}{\omega}\}$ the coordinate curves of constant (r, θ', z) are timelike curves corresponding to (accelerated) observers rotating rigidly with respect to the inertial observers of constant (r, θ, z).

(d) The set of the rotating observers is a 3-dimensional smooth manifold Σ with local coordinates (r, θ', z), and there exists a natural projection $\pi : U \to \Sigma$. We introduce a Riemannian metric h on Σ as follows: if $v \in T_{\pi(p)}\Sigma$ then

$$h(v,v) = g\left(v^\dagger, v^\dagger\right),$$

where $v^\dagger \in T_pU$ satisfies

$$(d\pi)_p\, v^\dagger = v \quad \text{ and } \quad g\left(v^\dagger, \left(\frac{\partial}{\partial t}\right)_p\right) = 0.$$

Show that h is well defined and

$$h = dr \otimes dr + \frac{r^2}{1-\omega^2 r^2}\, d\theta' \otimes d\theta' + dz \otimes dz.$$

(**Remark:** This is the metric resulting from local distance measurements between the rotating observers; Einstein used the fact that this metric has curvature to argue for the need to use non-Euclidean geometry in the relativistic description of gravity).

(e) The image of a curve $c : \mathbb{R} \to U$ consists of simultaneous events from the point of view of the rotating observers if $\dot{c}$ is orthogonal to $\frac{\partial}{\partial t}$ at each point. Show that this is equivalent to requiring that $\alpha(\dot{c}) = 0$, where

$$\alpha = dt - \frac{\omega r^2}{1-\omega^2 r^2} d\theta'.$$

In particular, show that in general synchronization of the rotating observers' clocks around closed paths leads to inconsistencies. (**Remark:** This is the so-called **Sagnac effect**; it must be taken into account when synchronizing the very precise atomic clocks on the GPS system ground stations).

(7) Let (Σ, h) be a 3-dimensional Riemannian manifold and consider the 4-dimensional Lorentzian manifold (M, g) determined by $M = \mathbb{R} \times \Sigma$ and

$$g = -e^{2\,\Phi\circ\pi} dt \otimes dt + \pi^* h,$$

where t is the usual coordinate in $\mathbb{R}$, $\pi : M \to \Sigma$ is the natural projection and $\Phi : \Sigma \to \mathbb{R}$ is a smooth function.

(a) Let $c : I \subset \mathbb{R} \to M$ be a timelike geodesic, and $\gamma = \pi \circ c$. Show that

$$\frac{D\dot\gamma}{d\tau} = (1 + h(\dot\gamma, \dot\gamma))\, G,$$

where $G = -\operatorname{grad}(\Phi)$ is the vector field associated to $-d\Phi$ by h and can be thought of as the gravitational field. Show that this equation implies that the quantity

$$E = (1 + h(\dot\gamma, \dot\gamma))^{\frac{1}{2}} e^{\Phi}$$

is a constant of motion.

(b) Let $c : I \subset \mathbb{R} \to M$ be a lightlike geodesic, $\tilde{c}$ its reparametrization by the coordinate time t, and $\tilde{\gamma} = \pi \circ \tilde{c}$. Show that $\tilde{\gamma}$ is a geodesic of the **Fermat metric**

$$l = e^{-2\Phi} h.$$

(c) Show that the vacuum Einstein field equation for g is equivalent to

$$\begin{aligned} &\operatorname{div} G = h(G, G); \\ &Ric + \nabla d\Phi = d\Phi \otimes d\Phi, \end{aligned}$$

where Ric and ∇ are the Ricci curvature and the Levi-Civita connection of h; $\nabla d\Phi$ is the tensor defined by $\nabla d\Phi(X, Y) = (\nabla_X d\Phi)(Y)$ for all $X, Y \in \mathfrak{X}(\Sigma)$ (cf. Exercise 2.6.3 in Chapter 3).

5. The Schwarzschild Solution

The vacuum Einstein field equation is nonlinear, and hence much harder to solve that the Laplace equation. One of the first solutions to be discovered was the so-called **Schwarzschild solution**, which can be obtained from the simplifying hypotheses of time independence and spherical symmetry, i.e. looking for solutions of the form

$$g = -A^2(r) dt \otimes dt + B^2(r) dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2 \theta d\varphi \otimes d\varphi$$

for unknown positive smooth functions $A, B : \mathbb{R} \to \mathbb{R}$. Notice that this expression reduces to the Minkowski metric in spherical coordinates for $A \equiv B \equiv 1$).

It is easily seen that Cartan's structure equations still hold for pseudo-Riemannian manifolds. We have

$$g = -\omega^0 \otimes \omega^0 + \omega^r \otimes \omega^r + \omega^\theta \otimes \omega^\theta + \omega^\varphi \otimes \omega^\varphi$$

with

$$\begin{aligned} \omega^0 &= A(r) dt; \\ \omega^r &= B(r) dr; \\ \omega^\theta &= r d\theta; \\ \omega^\varphi &= r \sin\theta d\varphi, \end{aligned}$$

and hence $\{\omega^0, \omega^r, \omega^\theta, \omega^\varphi\}$ is an orthonormal coframe. The first structure equations,

$$\begin{aligned} d\omega^\mu &= \sum_{\nu=0}^{3} \omega^\nu \wedge \omega^\mu_\nu; \\ dg_{\mu\nu} &= \sum_{\alpha=0}^{3} g_{\mu\alpha} \omega^\alpha_\nu + g_{\nu\alpha} \omega^\alpha_\mu, \end{aligned}$$

together with

$$\begin{aligned}
d\omega^0 &= \frac{A'}{B}\omega^r \wedge dt;\\
d\omega^r &= 0;\\
d\omega^\theta &= \frac{1}{B}\omega^r \wedge d\theta;\\
d\omega^\varphi &= \frac{\sin\theta}{B}\omega^r \wedge d\varphi + \cos\theta\omega^\theta \wedge d\varphi,
\end{aligned}$$

yield

$$\begin{aligned}
\omega^0_r &= \omega^r_0 = \frac{A'}{B}dt;\\
\omega^\theta_r &= -\omega^r_\theta = \frac{1}{B}d\theta;\\
\omega^\varphi_r &= -\omega^r_\varphi = \frac{\sin\theta}{B}d\varphi;\\
\omega^\varphi_\theta &= -\omega^\theta_\varphi = \cos\theta d\varphi.
\end{aligned}$$

The curvature forms can be computed from the second structure equations

$$\Omega^\mu_\nu = d\omega^\mu_\nu + \sum_{\alpha=0}^{3} \omega^\mu_\alpha \wedge \omega^\alpha_\nu,$$

and are found to be

$$\begin{aligned}
\Omega^0_r &= \Omega^r_0 = \frac{A''B - A'B'}{AB^3}\omega^r \wedge \omega^0;\\
\Omega^0_\theta &= \Omega^\theta_0 = \frac{A'}{rAB^2}\omega^\theta \wedge \omega^0;\\
\Omega^0_\varphi &= \Omega^\varphi_0 = \frac{A'}{rAB^2}\omega^\varphi \wedge \omega^0;\\
\Omega^\theta_r &= -\Omega^r_\theta = \frac{B'}{rB^3}\omega^\theta \wedge \omega^r;\\
\Omega^\varphi_r &= -\Omega^r_\varphi = \frac{B'}{rB^3}\omega^\varphi \wedge \omega^r;\\
\Omega^\varphi_\theta &= -\Omega^\theta_\varphi = \frac{B^2-1}{r^2B^2}\omega^\varphi \wedge \omega^\theta.
\end{aligned}$$

Thus the components of the curvature tensor on the orthonormal frame can be read off from the curvature forms using

$$\Omega^\mu_\nu = \sum_{\alpha<\beta} R_{\alpha\beta\nu}{}^{\mu}\omega^\alpha \wedge \omega^\beta,$$

and in turn be used to compute the components of the Ricci curvature tensor *Ric* on the same frame. The nonvanishing components of *Ric* on this frame turn out to be

$$R_{00} = \frac{A''B - A'B'}{AB^3} + \frac{2A'}{rAB^2};$$
$$R_{rr} = -\frac{A''B - A'B'}{AB^3} + \frac{2B'}{rB^3};$$
$$R_{\theta\theta} = R_{\varphi\varphi} = -\frac{A'}{rAB^2} + \frac{B'}{rB^3} + \frac{B^2-1}{r^2B^2}.$$

Thus the vacuum Einstein field equation $Ric = 0$ is equivalent to the ODE system

$$\begin{cases} \frac{A''}{A} - \frac{A'B'}{AB} + \frac{2A'}{rA} = 0 \\ \\ \frac{A''}{A} - \frac{A'B'}{AB} - \frac{2B'}{rB} = 0 \\ \\ \frac{A'}{A} - \frac{B'}{B} - \frac{B^2-1}{r} = 0 \end{cases} \Leftrightarrow \begin{cases} \frac{A'}{A} + \frac{B'}{B} = 0 \\ \\ \left(\frac{A'}{A}\right)' + 2\left(\frac{A'}{A}\right)^2 + \frac{2A'}{rA} = 0 \\ \\ \frac{2B'}{B} + \frac{B^2-1}{r} = 0 \end{cases}$$

The last equation can be immediately solved to yield

$$B = \left(1 - \frac{2m}{r}\right)^{-\frac{1}{2}},$$

where $m \in \mathbb{R}$ is an integration constant. The first equation implies that $A = \frac{\alpha}{B}$ for some constant $\alpha > 0$. By rescaling the time coordinate t we can assume that $\alpha = 1$. Finally, it is easily checked that the second ODE is identically satisfied. Therefore there exists a one-parameter family of solutions of the vacuum Einstein field equation of the form we seeked, given by

$$g = -\left(1 - \frac{2m}{r}\right) dt \otimes dt + \left(1 - \frac{2m}{r}\right)^{-1} dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi.$$

To interpret this family of solutions, we compute the proper acceleration (cf. Exercise 2.2.12) of the **stationary observers**, whose motions are the integral curves of $\frac{\partial}{\partial t}$. If $\{E_0, E_r, E_\theta, E_\varphi\}$ is the orthonormal frame obtained by normalizing $\left\{\frac{\partial}{\partial t}, \frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \varphi}\right\}$ (hence dual to $\{\omega^0, \omega^r, \omega^\theta, \omega^\varphi\}$), we have

$$\nabla_{E_0} E_0 = \sum_{\mu=0}^{3} \omega_0^\mu(E_0) E_\mu = \omega_0^r(E_0) E_r = \frac{A'}{AB}\omega^0(E_0)E_r = \frac{m}{r^2}\left(1 - \frac{2m}{r}\right)^{-\frac{1}{2}} E_r.$$

Therefore, each stationary observer is accelerating with a proper acceleration $\frac{m}{r^2}\left(1 - \frac{2m}{r}\right)^{-\frac{1}{2}}$ away from the origin, to prevent falling towards it. In other words, they are experiencing a gravitational field of intensity $\frac{m}{r^2}\left(1 - \frac{2m}{r}\right)^{-\frac{1}{2}}$, directed towards the origin. Since for large values of r this approaches the familiar acceleration $\frac{m}{r^2}$ of the Newtonian gravitational field generated by a point particle of mass m, we interpret the Schwarzschild solution as the general relativistic field of a point particle of mass m. Accordingly, we will assume that $m > 0$ (notice that $m = 0$ corresponds to Minkowski spacetime).

When obtaining the Schwarzschild solution we assumed $A(r) > 0$, and hence $r > 2m$. However, it is easy to check that it is also a solution of Einstein's vacuum field equation for $r < 2m$. Notice that the coordinate system (t, r, θ, φ) is singular at $r = 2m$, and hence covers only the two disconnected open sets $\{r > 2m\}$ and $\{r < 2m\}$. Both these sets are geodesically incomplete, as for instance radial timelike or null geodesics cannot be continued past $r = 0$ or $r = 2m$. While this is to be expected for $r = 0$, as the curvature blows up along geodesics approaching this limit, this is not the case for $r = 2m$. It turns out that it is possible to fit these two open sets together to obtain a solution of Einstein's vacuum field equation regular at $r = 2m$. To do so, we introduce the so-called **Painlevé time coordinate**

$$t' = t + \int \sqrt{\frac{2m}{r}} \left(1 - \frac{2m}{r}\right)^{-1} dr.$$

In the coordinate system (t', r, θ, φ), the Schwarzschild metric is written

$$g = -dt' \otimes dt' + \left(dr + \sqrt{\frac{2m}{r}}\, dt'\right) \otimes \left(dr + \sqrt{\frac{2m}{r}}\, dt'\right) + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta\, d\varphi \otimes d\varphi.$$

This expression is nonsingular at $r = 2m$, and is a solution of Einstein's vacuum field equation for $\{r > 2m\}$ and $\{r < 2m\}$. By continuity, it must be a solution also at $r = 2m$.

The submanifold $r = 2m$ is called the **event horizon**, and is ruled by null geodesics. This is easily seen from the fact that $\frac{\partial}{\partial t'} = \frac{\partial}{\partial t}$ becomes null at $r = 2m$, and hence its integral curves are (reparametrizations of) null geodesics.

The causal properties of the Schwarzschild spacetime are best understood by studying the **light cones**, i.e. the set of tangent null vectors at each point. For instance, radial null vectors $v = v^0 \frac{\partial}{\partial t'} + v^r \frac{\partial}{\partial r}$ satisfy

$$-\left(v^0\right)^2 + \left(v^r + \sqrt{\frac{2m}{r}}\, v^0\right)^2 = 0 \Leftrightarrow v^r = \left(\pm 1 - \sqrt{\frac{2m}{r}}\right) v^0.$$

For $r \gg 2m$ we obtain approximately the usual light cones of Minkowski spacetime. as r approaches $2m$, however, the light cones "tip over" towards the origin, becoming tangent to the event horizon at $r = 2m$ (cf. Figure 3). Since the tangent vector to a timelike curve must be inside the light cone, we see that no particle which crosses the event horizon can ever leave the region $r = 2m$ (which for this reason is called a **black hole**). Once inside the black hole, the light cones tip over even more, forcing the particle into the singularity $r = 0$.

Notice that the Schwarzschild solution in Painlevé coordinates is still not geodesically complete at the event horizon, as outgoing radial timelike and

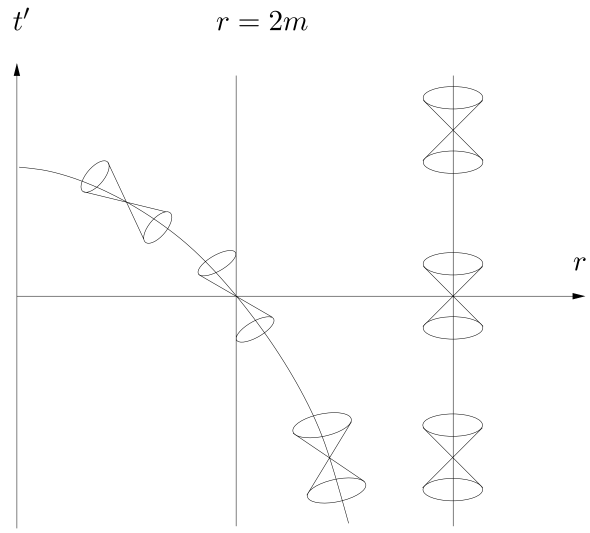

FIGURE 3. Light cones in Painlevé coordinates.

null geodesics cannot be continued to the past through $r = 2m$. Physically, this is not important: black holes are thought to form through the collapse of (approximately) spherical stars, whose surface follows a radial timelike curve in the spacetime diagram of Figure 3. Since only outside the star is there vacuum, the Schwarzschild solution in expected to hold only above this curve, thereby removing the region of $r = 2m$ leading to incompleteness. Nevertheless, it is possible to glue two copies of the Schwarzschild spacetime in Painlevé coordinates to obtain a solution of the vacuum Einstein field equation which is geodesically incomplete only at the two copies of $r = 0$. This solution, known as the **Kruskal extension**, contains a black hole and its time-reversed version, known as a **white hole**.

For some time it was thought that the curvature singularity at $r = 0$ was an artifact of the high symmetry of Schwarzschild spacetime, and that more realistic models of collapsing stars would be singularity-free. Penrose and Hawking (see [**Pen65, HP70**]) proved that this was is the case: once the collapse has begun, no matter how asymmetric, nothing can prevent a singularity from forming (cf. Section 8).

EXERCISES 5.1.

(1) Show that Cartan's structure equations still hold for pseudo-Riemannian manifolds
(2) Let (M, g) be a 2-dimensional Lorentzian manifold.
   (a) Consider an orthonormal frame $\{E_0, E_1\}$ on an open set $U \subset M$, with associated coframe $\{\omega^0, \omega^1\}$. Show that Cartan's

structure equations are

$$\begin{aligned}
&\omega_1^0 = \omega_0^1;\\
&d\omega^0 = \omega^1 \wedge \omega_1^0;\\
&d\omega^1 = \omega^0 \wedge \omega_1^0;\\
&\Omega_1^0 = d\omega_1^0.
\end{aligned}$$

(b) Let $\{F_0, F_1\}$ be another orthonormal frame such that $F_0 \in C(E_0)$, with associated coframe $\{\overline{\omega}^0, \overline{\omega}^1\}$ and connection form $\overline{\omega}_1^0$. Show that $\sigma = \overline{\omega}_1^0 - \omega_1^0$ is given locally by $\sigma = du$, where u is the hyperbolic angle between F_0 and E_0 (cf. Exercise 2.2.7).

(c) Consider a triangle $\Delta \subset U$ whose sides are timelike geodesics, and let α, β and γ be the hyperbolic angles between them (cf. Figure 4). Show that

$$\gamma = \alpha + \beta + \int_\Delta \Omega_1^0,$$

where, following the usual convention for spacetime diagrams, we orient U so that $\{E_0, E_1\}$ is **negative**.

(d) Provide a physical interpretation for the formula above in the case in which (M, g) is a totally geodesic submanifold of the Schwarzschild spacetime obtained by fixing (θ, φ) (cf. Exercise 5.7.3 in Chapter 4).

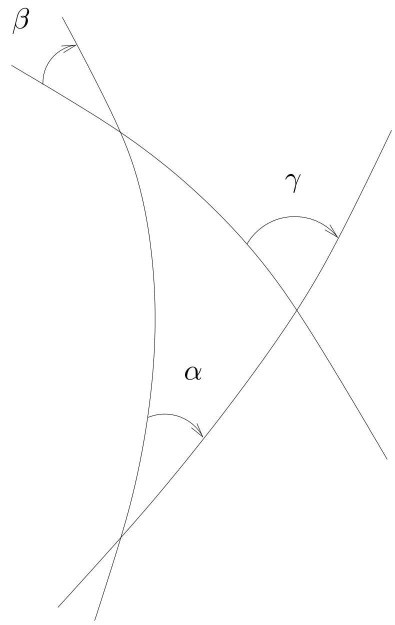

FIGURE 4. Timelike geodesic triangle.

(3) Consider the Schwarzschild spacetime with local coordinates (t, r, θ, φ). An **equatorial circular curve** is a curve given in these coordinates by $(t(\tau), r(\tau), \theta(\tau), \varphi(\tau))$ with $\dot{r}(\tau) \equiv 0$ and $\theta(\tau) \equiv \frac{\pi}{2}$.

(a) Show that the conditions for such a curve to be a timelike geodesic parametrized by its proper time are

$$\begin{cases} \ddot{t} = 0 \\ \ddot{\varphi} = 0 \\ r\dot{\varphi}^2 = \frac{m}{r^2}\dot{t}^2 \\ \left(1 - \frac{3m}{r}\right)\dot{t}^2 = 1 \end{cases}$$

Conclude that massive particles can orbit the central mass in circular orbits for all $r > 3m$.

(b) Show that there exists an equatorial circular null geodesic for $r = 3m$. What does a stationary observer placed at $r = 3m$, $\theta = \frac{\pi}{2}$ see as he looks along the direction of this lightlike geodesic?

(c) The angular momentum vector of a free-falling spinning particle is parallel-transported along its motion, and orthogonal to it (cf. Exercise 4.3.4). Consider a spinning particle on a circular orbit around a pointlike mass m. Show that the axis precesses by an angle

$$\delta = 2\pi\left(1 - \left(1 - \frac{3m}{r}\right)^{\frac{1}{2}}\right),$$

after one revolution, if initially aligned with the radial direction. (**Remark:** The above precession, which has been observed for spinning quartz spheres in orbit around the Earth during the Gravity Probe B experiment, is called the **geodesic precession**).

(4) We consider again the Schwarzschild spacetime with local coordinates (t, r, θ, φ).

(a) Show that the proper time interval $\Delta\tau$ measured by a stationary observer between two events on his history is

$$\Delta\tau = \left(1 - \frac{2m}{r}\right)^{\frac{1}{2}} \Delta t,$$

where Δt is the difference between the time coordinates of the two events (loosely speaking, clocks closer to the central mass run slower).

(b) Show that if $(t(\tau), r(\tau), \theta(\tau), \varphi(\tau))$ is a geodesic then so is $(t(\tau) + \Delta t, r(\tau), \theta(\tau), \varphi(\tau))$ for any $\Delta t \in \mathbb{R}$. Conclude that the time coordinate t can be thought of as the time between events at a fixed location as seen by stationary observers at infinity.

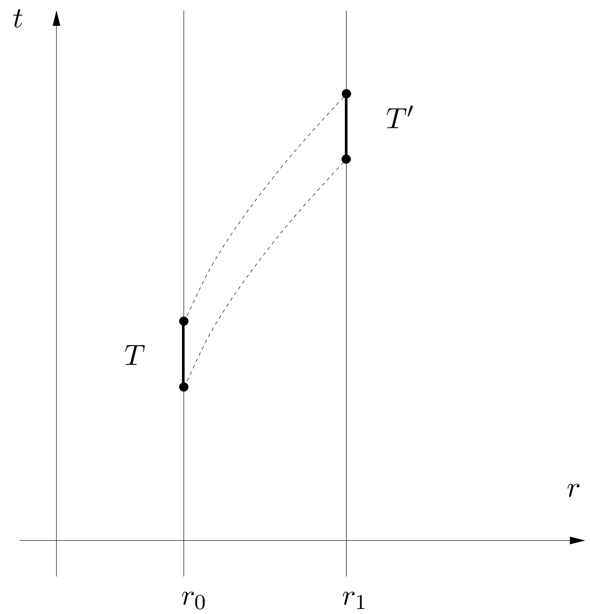

FIGURE 5. Gravitational redshift.

(c) (*Gravitational redshift*) Use the spacetime diagram in Figure 5 to show that if a stationary observer at $r = r_0$ measures a light signal to have period T, a stationary observer at $r = r_1$ measures a period

$$T' = T\sqrt{\frac{1 - \frac{2m}{r_1}}{1 - \frac{2m}{r_0}}}$$

for the same signal.

(d) Show that the proper time interval $\Delta\tau$ measured by an observer moving on a circular orbit between two events on his history is

$$\Delta\tau = \left(1 - \frac{3m}{r}\right)^{\frac{1}{2}} \Delta t,$$

where Δt is the difference between the time coordinates of the two events. (**Remark:** Notice that in particular the period of a circular orbit as measured by a free-falling orbiting observer is **smaller** than the period of the same orbit as measured by an accelerating stationary observer; thus a circular orbit over a full period is a non-maximizing geodesic – cf. Exercise 8.12.9).

(e) By setting $c = G = 1$, one can measure both time intervals and masses in meters. In these units, Earth's mass is approximately 0.0044 meters. Assume the atomic clock at a GPS ground station in the equator (whose radius is approximately

6, 400 kilometers) and the atomic clock on a GPS satellite moving on a circular orbit at an altitude of 20, 200 kilometers are initially synchronized. By how much will the two clocks be offset after one day? (**Remark:** This has important consequences for the GPS navigational system, which uses very accurate time measurements to compute the receiver's coordinates: if it were not taken into account, the error in the calculated position would be of the order of the time offset you just computed).

(5) Let (M, g) be the region $r > 2m$ of the Schwarzschild solution with the Schwarzschild metric. The set of all stationary observers in M is a 3-dimensional smooth manifold Σ with local coordinates (r, θ, φ), and there exists a natural projection $\pi : M \to \Sigma$. We introduce a Riemannian metric h on Σ as follows: if $v \in T_{\pi(p)}\Sigma$ then

$$h(v, v) = g\left(v^\dagger, v^\dagger\right),$$

where $v^\dagger \in T_pM$ satisfies

$$(d\pi)_p\, v^\dagger = v \quad \text{and} \quad g\left(v^\dagger, \left(\frac{\partial}{\partial t}\right)_p\right) = 0$$

(cf. Exercise 4.3.6).

(a) Show that h is well defined and

$$h = \left(1 - \frac{2m}{r}\right)^{-1} dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi.$$

(b) Show that h is not flat, but has zero scalar curvature.

(c) Show that the equatorial plane $\theta = \frac{\pi}{2}$ is isometric to the revolution surface generated by the curve $z(r) = \sqrt{8m(r - 2m)}$ when rotated around the z-axis (cf. Figure 6).

(**Remark:** This is the metric resulting from local distance measurements between the stationary observers; loosely speaking, gravity deforms space).

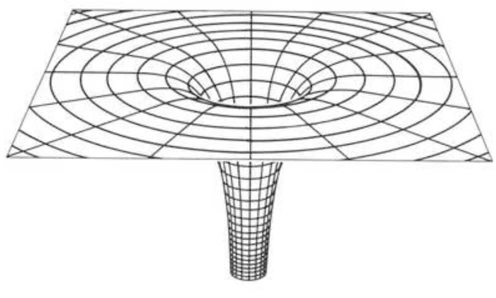

FIGURE 6. Surface of revolution isometric to the equatorial plane.

(6) In this exercise we study in detail the timelike and null geodesics of the Schwarzschild spacetime. We start by observing that the submanifold $\theta = \frac{\pi}{2}$ is totally geodesic (cf. Exercise 5.7.3 in Chapter 4). By adequately choosing the angular coordinates (θ, φ), one can always assume that the initial condition of the geodesic is tangent to this submanifold; hence it suffices to study the timelike and null geodesics of the 3-dimensional Lorentzian manifold (M, g), where

$$g = -\left(1 - \frac{2m}{r}\right) dt \otimes dt + \left(1 - \frac{2m}{r}\right)^{-1} dr \otimes dr + r^2 d\varphi \otimes d\varphi.$$

(a) Show that $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial \varphi}$ are Killing fields (cf. 3.3.8 in Chapter 3).

(b) Conclude that the equations for a curve $c : \mathbb{R} \to M$ to be a future-directed geodesic (parametrized by proper time if timelike) can be written as

$$\begin{cases} g(\dot{c}, \dot{c}) = -\sigma \\ g\left(\frac{\partial}{\partial t}, \dot{c}\right) = E \\ g\left(\frac{\partial}{\partial \varphi}, \dot{c}\right) = L \end{cases} \Leftrightarrow \begin{cases} \dot{r}^2 = E^2 - \left(\sigma + \frac{L^2}{r^2}\right)\left(1 - \frac{2m}{r}\right) \\ \left(1 - \frac{2m}{r}\right)\dot{t} = E \\ r^2 \dot{\varphi} = L \end{cases}$$

where $E > 0$ and L are integration constants, $\sigma = 1$ for timelike geodesics and $\sigma = 0$ for null geodesics.

(c) Show that if $L \neq 0$ then $u = \frac{1}{r}$ satisfies

$$\frac{d^2u}{d\varphi^2} + u = \frac{m\sigma}{L^2} + 3mu^2.$$

(d) For situations where relativistic corrections are small one has $mu \ll 1$, and hence the approximate equation

$$\frac{d^2u}{d\varphi^2} + u = \frac{m}{L^2}$$

holds for timelike geodesics. Show that the solution to this equation is the equation for a conic section in polar coordinates,

$$u = \frac{m}{L^2}(1 + \varepsilon \cos(\varphi - \varphi_0)),$$

where the integration constants $\varepsilon \geq 0$ and φ_0 are the eccentricity and the argument of the pericenter.

(e) Show that for $\varepsilon \ll 1$ this approximate solution satisfies

$$u^2 = \frac{2m}{L^2}u - \frac{m^2}{L^4}.$$

Argue that timelike geodesics close to circular orbits where relativistic corrections are small yield approximate solutions of the equation

$$\frac{d^2u}{d\varphi^2} + \left(1 - \frac{6m^2}{L^2}\right) u = \frac{m}{L^2}\left(1 - \frac{3m^2}{L^2}\right),$$

and hence the pericenter advances by approximately

$$\frac{6\pi m}{r}$$

radians per revolution. (**Remark:** The first success of General Relativity was due to this effect, which explained the anomalous **precession of Mercury's perihelion** – 43 arcseconds per century.).

(f) Show that if one neglects relativistic corrections then null geodesics satisfy

$$\frac{d^2u}{d\varphi^2} + u = 0.$$

Show that the solution to this equation is the equation for a straight line in polar coordinates,

$$u = \frac{1}{b}\sin(\varphi - \varphi_0)),$$

where the integration constants $b > 0$ and φ_0 are the **impact parameter** (distance of closest approach to the center) and the angle between the line and the x-axis.

(g) Assume that $mu \ll 1$. Let us include relativistic corrections by looking for approximate solutions of the form

$$u = \frac{1}{b}\left(\sin\varphi + \frac{m}{b}v\right)$$

(where we take $\varphi_0 = 0$ for simplicity). Show that v is an approximate solution of the equation

$$\frac{d^2v}{d\varphi^2} + v = 3\sin^2\varphi,$$

and hence u is approximately given by

$$u = \frac{1}{b}\left(\sin\varphi + \frac{m}{b}\left(\frac{3}{2} + \frac{1}{2}\cos(2\varphi) + \alpha\cos\varphi + \beta\sin\varphi\right)\right),$$

where α and β are integration constants.

(h) Show that for the incoming part of the null geodesic ($\varphi \simeq 0$) one has approximately

$$u = 0 \Leftrightarrow \varphi = -\frac{m}{b}\left(2 + \alpha\right).$$

Similarly, show that for the outgoing part of the null geodesic ($\varphi \simeq \pi$) one has approximately

$$u = 0 \Leftrightarrow \varphi = \pi + \frac{m}{b}\left(2 - \alpha\right).$$

Conclude that φ varies by approximately

$$\Delta\varphi = \pi + \frac{4m}{b}$$

radians along its path, and hence the null geodesic is deflected towards the center by approximately

$$\frac{4m}{b}$$

radians. (**Remark:** The measurement of this **deflection of light** by the Sun – 1.75 arcseconds – was the first experimental confirmation of General Relativity, and made Einstein a world celebrity overnight).

(7) (*Birkhoff Theorem*) Prove that the only Ricci-flat Lorentzian metric given in local coordinates (t, r, θ, φ) by

$$g = A^2(t,r)dt \otimes dt + B^2(t,r)dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi$$

is the Schwarzschild metric. Loosely speaking, spherically symmetric mass configurations do not radiate.

(8) Show that observers satisfying

$$\frac{dr}{dt'} = -\sqrt{\frac{2m}{r}}$$

in Painlevé's coordinates are free-falling, and that t' is their proper time.

(9) What does a stationary observer at infinity see as a particle falls into a black hole?

(10) Show that an observer who crosses the horizon will hit the singularity in proper time at most πm.

6. Cosmology

The the purpose of cosmology is the study of the behavior of the Universe as a whole. Experimental observations (chiefly that of the cosmic background radiation) suggest that space is isotropic at Earth's location. Assuming the **Copernican Principle** that Earth's location in the Universe is not in any way special, we take an isotropic (hence constant curvature) 3-dimensional Riemannian manifold (Σ, h) as our model of space. We can always find local coordinates (r, θ, φ) on Σ such that

$$h = a^2 \left(\frac{1}{1-kr^2} dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi \right),$$

where $a > 0$ is the "radius" of space and $k = -1, 0, 1$ according to whether the curvature is negative, zero or positive (cf. Exercise 6.1.1). Allowing for the possibility that the "radius" of space may be varying in time, we take our model of the Universe to be (M, g), where $M = \mathbb{R} \times \Sigma$ and

$$g = -dt \otimes dt + a^2(t) \left(\frac{1}{1-kr^2} dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi \right).$$

These are the so-called **Friedmann-Robertson-Walker** models of cosmology.

One can easily compute the Ricci curvature for the metric g: we have

$$g = -\omega^0 \otimes \omega^0 + \omega^r \otimes \omega^r + \omega^\theta \otimes \omega^\theta + \omega^\varphi \otimes \omega^\varphi$$

with

$$\begin{aligned}
&\omega^0 = dt;\\
&\omega^r = a(t)\left(1 - kr^2\right)^{-\frac{1}{2}} dr;\\
&\omega^\theta = rd\theta;\\
&\omega^\varphi = r\sin\theta d\varphi,
\end{aligned}$$

and hence $\{\omega^0, \omega^r, \omega^\theta, \omega^\varphi\}$ is an orthonormal coframe. The first structure equations yield

$$\begin{aligned}
&\omega^0_r = \omega^r_0 = \dot{a}\left(1 - kr^2\right)^{-\frac{1}{2}} dr;\\
&\omega^0_\theta = \omega^\theta_0 = \dot{a}rd\theta;\\
&\omega^0_\varphi = \omega^\varphi_0 = \dot{a}r\sin\theta d\varphi;\\
&\omega^\theta_r = -\omega^r_\theta = \left(1 - kr^2\right)^{\frac{1}{2}} d\theta;\\
&\omega^\varphi_r = -\omega^r_\varphi = \left(1 - kr^2\right)^{\frac{1}{2}} \sin\theta d\varphi;\\
&\omega^\varphi_\theta = -\omega^\theta_\varphi = \cos\theta d\varphi.
\end{aligned}$$

The curvature forms can be computed from the second structure equations, and are found to be

$$\begin{aligned}
&\Omega^0_r = \Omega^r_0 = \frac{\ddot{a}}{a}\omega^0 \wedge \omega^r;\\
&\Omega^0_\theta = \Omega^\theta_0 = \frac{\ddot{a}}{a}\omega^0 \wedge \omega^\theta;\\
&\Omega^0_\varphi = \Omega^\varphi_0 = \frac{\ddot{a}}{a}\omega^0 \wedge \omega^\varphi;\\
&\Omega^\theta_r = -\Omega^r_\theta = \left(\frac{k}{a^2} + \frac{\dot{a}^2}{a^2}\right)\omega^\theta \wedge \omega^r;\\
&\Omega^\varphi_r = -\Omega^r_\varphi = \left(\frac{k}{a^2} + \frac{\dot{a}^2}{a^2}\right)\omega^\varphi \wedge \omega^r;\\
&\Omega^\varphi_\theta = -\Omega^\theta_\varphi = \left(\frac{k}{a^2} + \frac{\dot{a}^2}{a^2}\right)\omega^\varphi \wedge \omega^\theta.
\end{aligned}$$

The components of the curvature tensor on the orthonormal frame can be read off from the curvature forms, and can in turn be used to compute the components of the Ricci curvature tensor Ric on the same frame. The nonvanishing components of Ric on this frame turn out to be

$$R_{00} = -\frac{3\ddot{a}}{a};$$
$$R_{rr} = R_{\theta\theta} = R_{\varphi\varphi} = \frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2} + \frac{2k}{a^2}.$$

At very large scales, galaxies and clusters of galaxies are expected to behave as particles of a pressureless fluid, which we take to be our matter model. Therefore the Einstein field equation is

$$Ric = 4\pi\rho(2dt \otimes dt + g),$$

and is equivalent to the ODE system

$$\begin{cases} -\dfrac{3\ddot{a}}{a} = 4\pi\rho \\ \\ \dfrac{\ddot{a}}{a} + \dfrac{2\dot{a}^2}{a^2} + \dfrac{2k}{a^2} = 4\pi\rho \end{cases} \Leftrightarrow \begin{cases} \ddot{a} + \dfrac{\dot{a}^2}{2a} + \dfrac{k}{2a} = 0 \\ \\ \rho = -\dfrac{3\ddot{a}}{4\pi a} \end{cases}$$

The first equation allows us to determine the function $a(t)$, and the second yields ρ (which in particular must be a function of the t coordinate only; this is to be taken to mean that the average density of matter at cosmological scales is spatially constant). It is easy to check that the first equation implies

$$\ddot{a} = -\frac{\alpha}{a^2}$$

for some integration constant α (we take $\alpha > 0$ so that $\rho > 0$). Substituting in the first equation we get the first order ODE

$$\frac{\dot{a}^2}{2} - \frac{\alpha}{a} = -\frac{k}{2}.$$

This is formally identical to the energy conservation equation for a particle falling on a Keplerian potential $V(a) = -\frac{\alpha}{a}$ with total energy $-\frac{k}{2}$. Thus we see that $a(t)$ will be bounded if and only if $k = 1$. Notice that in all cases $a(t)$ explodes for some value of t, conventionally taken to be $t = 0$ (**Big Bang**). Again it was thought that this could be due to the high symmetry of the Friedmann-Robertson-Walker models. Hawking and Penrose (see [**Haw67, HP70**]) showed that actually the big bang is a generic feature of cosmological models (cf. Section 8).

The function

$$H(t) = \frac{\dot{a}}{a}$$

is (somewhat confusingly) called **Hubble's constant**. It is easy to see from the above equations that

$$H^2 + \frac{k}{a^2} = \frac{8\pi}{3}\rho.$$

Therefore, in these models one has $k = -1$, $k = 0$ or $k = 1$ according to whether the average density ρ of the Universe is smaller than, equal to or bigger than the so-called **critical density**

$$\rho_c = \frac{3H^2}{8\pi}.$$

These models were the standard models for cosmology for a long time. Currently, however, things are thought to be slightly more complicated (cf. Exercise 6.1.7).

EXERCISES 6.1.

(1) Show that the Riemannian metric h given in local coordinates (r, θ, φ) by

$$h = a^2\left(\frac{1}{1-kr^2}dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi\right)$$

has constant curvature $K = \frac{k}{a^2}$.

(2) The motions of galaxies and groups of galaxies in the Friedmann-Robertson-Walker models are the integral curves of $\frac{\partial}{\partial t}$. Show that these are timelike geodesics, and that the time coordinate t is the proper time of such observers.

(3) (a) Show that the differential equation for $a(t)$ implies that this function explodes in finite time (usually the singularity is taken to be at $t = 0$).
(b) Show that if $k = -1$ or $k = 0$ then the solution can be extended to all values of $t > 0$.
(c) Show that if $k = 1$ then the solution cannot be extended past some positive value $t = T > 0$ (**Big Crunch**).
(d) Show that if the spatial sections are 3-spheres (hence $k = 1$) then the light which leaves some galaxy at the Big Bang travels once around the 3-sphere and is just reaching it at the Big Crunch. Conclude that no observer can circumnavigate the Universe, no matter how fast he moves.

(4) Show that the solutions to the Einstein equation for the Friedmann-Robertson-Walker models can be given parametrically by:
(a) $k = 1$:

$$\begin{cases} a = \alpha(1 - \cos u) \\ t = \alpha(u - \sin u) \end{cases}$$

(b) $k = 0$:

$$\begin{cases} a = \frac{\alpha}{2}u^2 \\ t = \frac{\alpha}{6}u^3 \end{cases}$$

(c) $k = -1$:

$$\begin{cases} a = \alpha(\cosh u - 1) \\ t = \alpha(\sinh u - u) \end{cases}$$

(5) Show that the Friedmann-Robertson-Walker model with $k = 1$ is isometric to the hypersurface with equation

$$\sqrt{x^2 + y^2 + z^2 + w^2} = 2\alpha - \frac{t^2}{8\alpha}$$

in the 5-dimensional Minkowski spacetime $(\mathbb{R}^5, g)$ with metric

$$g = -dt \otimes dt + dx \otimes dx + dy \otimes dy + dz \otimes dz + dw \otimes dw.$$

(6) (*A model of collapse*) Show that the radius of a free-falling spherical shell $r = r_0$ in a Friedmann-Robertson-Walker model changes with proper time in exactly the same fashion as the radius of a free-falling spherical shell in a Schwarzschild spacetime of mass parameter m moving with energy parameter E (cf. Exercise 5.1.6), provided that

$$\begin{cases} M = \alpha r_0^3 \\ E^2 - 1 = -k r_0^3 \end{cases}$$

Therefore these two spacetimes ca be matched along the 3-dimensional hypersurface determined by the spherical shell's history to yield a model of collapsing matter. Can you physically interpret the three cases $k = 1$, $k = 0$ and $k = -1$?

(7) Show that if we allow for a **cosmological constant** $\Lambda \in \mathbb{R}$, i.e. for an Einstein equation of the form

$$Ric = 4\pi\rho(2\nu \otimes \nu + g) + \Lambda g$$

then the equations for the Friedmann-Robertson-Walker models become

$$\begin{cases} \dfrac{\dot{a}^2}{2} - \dfrac{\alpha}{a} - \dfrac{\Lambda}{6}a^2 = -\dfrac{k}{2} \\ \\ \dfrac{4\pi}{3}a^3\rho = \alpha \end{cases}$$

Analyze the possible behaviors of the function $a(t)$. (**Remark:** It is currently thought that there exists indeed a positive cosmological constant, also known as **dark energy**. The model favored by experimental observations seems to be $k = 0$, $\Lambda > 0$).

(8) Consider the 5-dimensional Minkowski spacetime $(\mathbb{R}^5, g)$ with metric

$$g = -dt \otimes dt + dx \otimes dx + dy \otimes dy + dz \otimes dz + dw \otimes dw.$$

Show that the induced metric on each of the following hypersurfaces determines generalized Friedmann-Robertson-Walker models with the indicated parameters:

(a) (*Einstein universe*) The "cylinder" of equation

$$x^2 + y^2 + z^2 + w^2 = \frac{1}{\Lambda},$$

satisfies $k = 1$, $\Lambda > 0$ and $\rho = \frac{\Lambda}{4\pi}$.

(b) (*de Sitter universe*) The "sphere" of equation

$$-t^2 + x^2 + y^2 + z^2 + w^2 = \frac{3}{\Lambda}$$

satisfies $k = 1$, $\Lambda > 0$ and $\rho = 0$.

7. Causality

In this section we will study the causal features of spacetimes. This is a subject which has no parallel in Riemannian geometry, where the metric is positive definite. Although we will focus on 4-dimensional Lorentzian manifolds, the discussion can be easily generalized to any number $n \geq 2$ of dimensions.

A spacetime (M, g) is said to be **time-orientable** if there exists a vector field $T \in \mathfrak{X}(M)$ such that $\langle T, T \rangle < 0$. In this case, we can define a time orientation on each tangent space T_pM (which is, of course, isometric to Minkowski spacetime) by choosing $C(T_p)$ to be the future-pointing timelike vectors.

Assume that (M, g) is **time-oriented** (i.e. time-orientable with a definite choice of time orientation). A timelike curve $c : I \subset \mathbb{R} \to M$ is said to be **future-directed** if $\dot{c}$ is future-pointing. The **chronological future** of $p \in M$ is the set $I^+(p)$ of all points to which p can be connected by a future-directed timelike curve. A **future-directed causal curve** is a curve $c : I \subset \mathbb{R} \to M$ such that $\dot{c}$ is non-spacelike and future-pointing (if nonzero). The **causal future** of $p \in M$ is the set $J^+(p)$ of all points to which p can be connected by a future-directed causal curve. Notice that $I^+(p)$ is simply the set of all events which are accessible to a particle with nonzero mass at p, whereas $J^+(p)$ is the set of events which can be causally influenced by p (as this causal influence cannot propagate faster than the speed of light). Analogously, the **chronological past** of $p \in M$ is the set $I^-(p)$ of all points which can be connected to p by a future-directed timelike curve, and the **causal past** of $p \in M$ is the set $J^-(p)$ of all points which can be connected to p by a future-directed causal curve.

In general, the chronological and causal pasts and futures can be quite complicated sets, because of global features of the spacetime. Locally, however, causal properties are similar to those of Minkowski spacetime. More precisely, we have the following statement:

PROPOSITION 7.1. *Let (M, g) be a time-oriented spacetime. Then each point $p_0 \in M$ has an open neighborhood $V \subset M$ such that the spacetime (V, g) obtained by restricting g to V satisfies:*

(1) If $p, q \in V$ then there exists a unique geodesic (up to reparametrization) joining p to q (i.e. V is **geodesically convex***);*
(2) $q \in I^+(p)$ iff there exists a future-directed timelike geodesic connecting p to q;
(3) $J^+(p) = \overline{I^+(p)}$;

(4) *$q \in J^+(p)$ iff there exists a future-directed timelike or null geodesic connecting p to q.*

PROOF. Let U be a normal neighborhood of p_0 and choose **normal coordinates** (x^0, x^1, x^2, x^3) on U, given by the parametrization

$$\varphi(x^0, x^1, x^2, x^3) = \exp_{p_0}(x^0 v_0 + x^1 v_1 + x^2 v_2 + x^3 v_3),$$

where $\{v_0, v_1, v_2, v_3\}$ is a basis of $T_{p_0}(M)$ (cf. Exercise 4.8.2 in Chapter 3). Let $D : U \to \mathbb{R}$ be the differentiable function

$$D(p) := \sum_{\alpha=0}^{3} (x^\alpha(p))^2,$$

and let us define for each $\varepsilon > 0$ the set

$$B_\varepsilon = \{p \in U \mid D(p) < \varepsilon\},$$

which for sufficiently small ε is diffeomorphic to an open ball in $T_{p_0}M$. Assume, for simplicity, that U is one such set.

Let us show that there exists $k > 0$ such that if $c : I \subset \mathbb{R} \to B_k$ is a geodesic then all critical points of $D(t) := D(c(t))$ are strict local minima. In fact, setting $x^\mu(t) := x^\mu(c(t))$, we have

$$\begin{aligned}
\dot{D}(t) &= 2\sum_{\alpha=0}^{3} x^\alpha(t)\dot{x}^\alpha(t);\\
\ddot{D}(t) &= 2\sum_{\alpha=0}^{3} (\dot{x}^\alpha(t))^2 + 2\sum_{\alpha=0}^{4} x^\alpha(t)\ddot{x}^\alpha(t)\\
&= 2\sum_{\mu,\nu=0}^{3}\left(\delta_{\mu\nu} - \sum_{\alpha=0}^{3}\Gamma^\alpha_{\mu\nu}(c(t))x^\alpha(t)\right)\dot{x}^\mu(t)\dot{x}^\nu(t),
\end{aligned}$$

and for k sufficiently small the matrix

$$\delta_{\mu\nu} - \sum_{\alpha=0}^{3}\Gamma^\alpha_{\mu\nu}x^\alpha$$

is positive definite on B_k.

Consider the map $F : W \subset TM \to M \times M$, defined on some open neighborhood W of $0 \in T_{p_0}M$ by

$$F(v) = (\pi(v), \exp(v)).$$

As was established in the Riemannian case (cf. Chapter 3, Section 4), this map is a local diffeomorphism at $0 \in T_{p_0}M$. Choosing $\delta > 0$ sufficiently small and reducing W, we can assume that F maps W diffeomorphically to $B_\delta \times B_\delta$, and that $\exp(tv) \in B_k$ for all $t \in [0,1]$ and $v \in W$.

Finally, set $V = B_\delta$. If $p, q \in V$ and $v = F^{-1}(p,q)$, then $c(t) = \exp_p(tv)$ is a geodesic connecting p to q whose image is contained in B_k. If it image were not contained in V, there would necessarily be a point of local maximum of $D(t)$, which cannot occur. Therefore, there exists a geodesic in V

connecting p to q. Since $\exp_p$ is a diffeomorphism onto V, this geodesic is unique (up to reparametrization). This proves (1).

To prove assertion (2), we start by noticing that if there exists a future-directed timelike geodesic connecting p to q then it is obvious that $q \in I^+(p)$. Suppose now that $q \in I^+(p)$; then there exists a future-directed timelike curve $c : [0,1] \to V$ such that $c(0) = p$ and $c(1) = q$. Choose normal coordinates (x^0, x^1, x^2, x^3) given by the parametrization

$$\varphi(x^0, x^1, x^2, x^3) = \exp_p(x^0 E_0 + x^1 E_1 + x^2 E_2 + x^3 E_3),$$

where $\{E_0, E_1, E_2, E_3\}$ is an orthonormal basis of T_pM (with E_0 timelike and future-pointing). These are global coordinates in V, since $F : W \to V \times V$ is a diffeomorphism. Defining

$$\begin{aligned} W_p(q) &:= -\left(x^0(q)\right)^2 + \left(x^1(q)\right)^2 + \left(x^2(q)\right)^2 + \left(x^3(q)\right)^2 \\ &= \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} x^\mu(q) x^\nu(q), \end{aligned}$$

we have to show that $W_p(q) < 0$. Let $W_p(t) := W_p(c(t))$. Since $x^\mu(p) = 0$ ($\mu = 0, 1, 2, 3$), we have $W_p(0) = 0$. Setting $x^\mu(t) = x^\mu(c(t))$, we have

$$\begin{aligned} \dot{W}_p(t) &= 2 \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} x^\mu(t) \dot{x}^\nu(t); \\ \ddot{W}_p(t) &= 2 \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} x^\mu(t) \ddot{x}^\nu(t) + 2 \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} \dot{x}^\mu(t) \dot{x}^\nu(t), \end{aligned}$$

and consequently (recalling that $\left(d \exp_p\right)_p = \mathrm{id}$)

$$\begin{aligned} \dot{W}_p(0) &= 0; \\ \ddot{W}_p(0) &= 2\langle \dot{c}(0), \dot{c}(0) \rangle < 0. \end{aligned}$$

Therefore there exists $\varepsilon > 0$ such that $W_p(t) < 0$ for $t \in (0, \varepsilon)$.

Using the same ideas as in the Riemannian case (cf. Chapter 3, Section 4), it is easy to prove that the level surfaces of W_p are orthogonal to the geodesics through p. Therefore, if $c_v(t) = \exp_p(tv)$ is the geodesic with initial condition $v \in T_pM$, we have

$$(\operatorname{grad} W_p)_{c_v(1)} = a(v) \dot{c}_v(1),$$

where the gradient of a function is defined as in the Riemannian case (notice however that in the Lorentzian case a smooth function f **decreases** along the direction of $\operatorname{grad} f$ if $\operatorname{grad} f$ is timelike). Now

$$\begin{aligned} \left\langle (\operatorname{grad} W_p)_{c_v(t)}, \dot{c}_v(t) \right\rangle &= \frac{d}{dt} W_p(c_v(t)) = \frac{d}{dt} W_p(c_{tv}(1)) \\ &= \frac{d}{dt} \left(t^2 W_p(c_v(1)) \right) = 2t W_p(c_v(1)), \end{aligned}$$

and hence

$$\left\langle (\operatorname{grad} W_p)_{c_v(1)}, \dot{c}_v(1) \right\rangle = 2W_p(c_v(1)).$$

On the other hand,

$$\begin{aligned}\left\langle (\operatorname{grad} W_p)_{c_v(1)}, \dot{c}_v(1) \right\rangle &= \langle a(v)\dot{c}_v(1), \dot{c}_v(1) \rangle \\ &= a(v)\langle v, v \rangle = a(v)W_p(c_v(1)).\end{aligned}$$

We conclude that $a(v) = 2$, and therefore

$$(\operatorname{grad} W_p)_{c_v(1)} = 2\dot{c}_v(1).$$

Consequently $\operatorname{grad} W_p$ is tangent to geodesics through p, being future-pointing on future-directed geodesics.

Suppose that $W_p(t) < 0$. Then

$$\dot{W}(t) = \left\langle (\operatorname{grad} W_p)_{c(t)}, \dot{c}(t) \right\rangle < 0$$

as both $(\operatorname{grad} W_p)_{c(t)}$ and $\dot{c}(t)$ are timelike future-pointing (cf. Exercise 2.2.2). We conclude that we must have $W_p(t) < 0$ for all $t \in [0, 1]$. In particular, $W_p(q) = W_p(1) < 0$, and hence there exists a future-directed timelike geodesic connecting p to q.

Assertion (3) can be proved by using the global normal coordinates (x^0, x^1, x^2, x^3) of V to approximate causal curves by timelike curves. We leave the details of this as an exercise. Once this is done, (4) is obvious from the fact that $\exp_p$ is a diffeomorphism onto V. □

The generalized twin paradox (cf. Exercise 2.2.8) also holds locally for general spacetimes. More precisely, we have the following statement:

PROPOSITION 7.2. *Let (M, g) be a time-oriented spacetime and $p_0 \in M$. Then there exists a geodesically convex open neighborhood $V \subset M$ of p_0 such that the spacetime (V, g) obtained by restricting g to V satisfies the following property: if $q \in I^+(p)$, c is the timelike geodesic connecting p to q and γ is any timelike curve connecting p to q, then $\tau(\gamma) \leq \tau(c)$, with equality iff γ is a is a reparametrization of c.*

PROOF. Choose V as in the proof of Proposition 7.1. Any timelike curve $\gamma : [0, 1] \to V$ satisfying $\gamma(0) = p$, $\gamma(1) = q$ can be written as

$$\gamma(t) = \exp_p(r(t)n(t)),$$

for $t \in [0, 1]$, where $r(t) \geq 0$ and $\langle n(t), n(t) \rangle = -1$. We have

$$\dot{\gamma}(t) = (\exp_p)_* \left(\dot{r}(t)n(t) + r(t)\dot{n}(t) \right).$$

Since $\langle n(t), n(t) \rangle = -1$, we have $\langle \dot{n}(t), n(t) \rangle = 0$, and consequently $\dot{n}(t)$ is tangent to the level surfaces of the function $v \mapsto \langle v, v \rangle$. We conclude that

$$\dot{\gamma}(t) = \dot{r}(t)X_{\gamma(t)} + Y(t),$$

where X is the unit tangent vector field to timelike geodesics through p and $Y(t) = r(t)(\exp_p)_* \dot{n}(t)$ is tangent to the level surfaces of W_p – hence orthogonal to $X_{\gamma(t)}$. Consequently,

$$\begin{aligned}\tau(\gamma) &= \int_0^1 \left|\left\langle \dot{r}(t)X_{\gamma(t)} + Y(t), \dot{r}(t)X_{\gamma(t)} + Y(t)\right\rangle\right|^{\frac{1}{2}} dt \\ &= \int_0^1 \left(\dot{r}(t)^2 - |Y(t)|^2\right)^{\frac{1}{2}} dt \\ &\leq \int_0^1 \dot{r}(t)dt = r(1) = \tau(c),\end{aligned}$$

(where we've used the facts that $\dot{r}(t) > 0$ for all $t \in [0,1]$, as $\dot{c}$ is future-pointing, and $\tau(c) = r(1)$, as $q = \exp_p(r(1)n(1))$. It should be clear that $\tau(\gamma) = \tau(c)$ if and only if $|Y(t)| \equiv 0 \Leftrightarrow Y(t) \equiv 0$ ($Y(t)$ is spacelike) for all $t \in [0,1]$, implying that n is constant. In this case, $\gamma(t) = \exp_p(r(t)n)$ is, up to reparametrization, the geodesic through p with initial condition $n \in T_pM$. □

There is also a local property characterizing null geodesics:

PROPOSITION 7.3. *Let (M, g) be a time-oriented spacetime and $p_0 \in M$. Then there exists a geodesically convex open neighborhood $V \subset M$ of p_0 such that the spacetime (V, g) obtained by restricting g to V satisfies the following property: if there exists a future-directed null geodesic c connecting p to q and γ is a causal curve connecting p to q then γ is a reparametrization of c.*

PROOF. Again choose V as in the proof of Proposition 7.1. Since p and q are connected by a null geodesic, we conclude from Proposition 7.1 that $q \in J^+(p) \setminus I^+(p)$. Let $\gamma : [0,1] \to V$ be a causal curve connecting p to q. Then we must have $\gamma(t) \in J^+(p) \setminus I^+(p)$ for all $t \in [0,1]$, since $\gamma(t_0) \in I^+(p)$ implies $\gamma(t) \in I^+(p)$ for all $t > t_0$ (again by Proposition 7.1). Consequently, we have

$$\left\langle (\operatorname{grad} W_p)_{\gamma(t)}, \dot{\gamma}(t) \right\rangle = 0.$$

The formula $(\operatorname{grad} W_p)_{c_v(1)} = 2\dot{c}_v(1)$, which was proved for timelike geodesics c_v with initial condition $v \in T_pM$, must also hold for null geodesics (by continuity). Hence $\operatorname{grad} W_p$ is tangent to the null geodesics ruling $J^+(p) \setminus I^+(p)$ and future-pointing. Since $\dot{\gamma}(t)$ is also future-pointing, we conclude that $\dot{\gamma}$ is proportional to $\operatorname{grad} W_p$ (cf. Exercise 2.2.8), and therefore γ must be a reparametrization of a null geodesic (which must be c). □

It is not difficult to show that if $r \in I^+(p)$ and $q \in J^+(r)$ (or $r \in J^+(p)$ and $q \in I^+(r)$) then $q \in I^+(p)$ (cf. Exercise 7.8.3). Therefore, we see that if p and q are connected by a future-directed causal curve which is not a null geodesic then $q \in I^+(p)$ (cf. Exercise 7.8.4).

For physical applications, it is important to require that the spacetime satisfies reasonable causality conditions. The simplest of these conditions excludes time travel, i.e. the possibility of a particle returning to an event in its past history.

DEFINITION 7.4. *A spacetime (M, g) is said to satisfy the* **chronology condition** *if it does not contain closed timelike curves.*

This condition is violated by compact spacetimes:

PROPOSITION 7.5. *Any compact spacetime (M, g) contains closed timelike curves.*

PROOF. Taking if necessary the time-orientable double covering (cf. Exercise 7.8.1), we can assume that (M, g) is time-oriented. Since $I^+(p)$ is an open set for any $p \in M$ (cf. Exercise 7.8.3), it is clear that $\{I^+(p)\}_{p\in M}$ is an open cover of M. If M is compact, we can obtain a finite subcover $\{I^+(p_1), \ldots, I^+(p_N)\}$. Now if $p_1 \in I^+(p_i)$ for $i \neq 1$ then $I^+(p_1) \subset I^+(p_i)$, and we can exclude $I^+(p_1)$ from the subcover. Therefore, we can assume without loss of generality that $p_1 \in I^+(p_1)$, and hence there exists a closed timelike curve starting and ending at p_1. □

A stronger restriction on the causal behavior of the spacetime is the following:

DEFINITION 7.6. *A spacetime (M, g) is said to be* **stably causal** *if there exists a* **global time function**, *i.e. a smooth function $t : M \to \mathbb{R}$ such that* $\operatorname{grad}(t)$ *is timelike.*

In particular, a stably causal spacetime is time-orientable. We choose the time orientation defined by $-\operatorname{grad}(t)$, so that t increases along future-directed timelike curves. Notice that this implies that no closed timelike curves can exist, i.e. any stably causal spacetime satisfies the chronology condition. In fact, any small perturbation of a causally stable spacetime still satisfies the chronology condition (cf. Exercise 7.8.5).

Let (M, g) be a time-oriented spacetime. A smooth future-directed causal curve $c : (a, b) \to M$ (with possibly $a = -\infty$ or $b = +\infty$) is said to be **future-inextendible** if $\lim_{t\to b} c(t)$ does not exist. The definition of a **past-inextendible** causal curve is analogous. The **future domain of dependence** of $S \subset M$ is the set $D^+(S)$ of all events $p \in M$ such that any past-inextendible causal curve starting at p intersects S. Therefore any causal influence on an event $p \in D^+(S)$ had to register somewhere in S, and one can expect that what happens at p can be predicted from data on S. Similarly, the **past domain of dependence** of S is the set $D^-(S)$ of all events $p \in M$ such that any future-inextendible causal curve starting at p intersects S. Therefore any causal influence of an event $p \in D^+(S)$ will register somewhere in S, and one can expect that what happened at p can be retrodicted from data on S. The **domain of dependence** of S is simply the set $D(S) = D^+(S) \cup D^-(S)$.

Let (M, g) be a stably causal spacetime with time function $t : M \to \mathbb{R}$. The level sets $S_a = t^{-1}(a)$ are said to be **Cauchy hypersurfaces** if

$D(S_a) = M$. Spacetimes for which this happens have particularly good causal properties.

DEFINITION 7.7. *A stably causal spacetime possessing a time function whose level sets are Cauchy hypersurfaces is said to be* **globally hyperbolic**.

Notice that the future and past domains of dependence of the Cauchy surfaces S_a are $D^+(S_a) = t^{-1}([a, +\infty))$ and $D^-(S_a) = t^{-1}((-\infty, a])$.

EXERCISES 7.8.

(1) (*Time-orientable double covering*) Using ideas similar to those of Exercise 8.6.9 in Chapter 1, show that if (M, g) is a non-time-orientable Lorentzian manifold then there exists a **time-orientable double covering**, i.e. a time-orientable Lorentzian manifold $(\overline{M}, \overline{g})$ and a local isometry $\pi : \overline{M} \to M$ such that every point in M has two preimages by π. Use this to conclude that the only compact surfaces which admit a Lorentzian metric are the torus T^2 and the Klein bottle K^2.
(2) Complete the proof of Proposition 7.1.
(3) Let (M, g) be a time oriented spacetime and $p \in M$. Show that:
 (a) $I^+(p)$ is open;
 (b) $J^+(p)$ is not necessarily closed;
 (c) $J^+(p) \subset \overline{I^+(p)}$;
 (d) if $r \in I^+(p)$ and $q \in J^+(r)$ then $q \in I^+(p)$;
 (e) if $r \in J^+(p)$ and $q \in I^+(r)$ then $q \in I^+(p)$;
 (f) it may happen that $I^+(p) = M$;
 (g) if U is an open set such that $H = \partial I^+(p) \cap U$ is a hypersurface, then the normal vector to H is null;
 (h) H is ruled by null geodesics.
(4) Consider the 3-dimensional Minkowski spacetime $(\mathbb{R}^3, g)$, where
$$g = -dt \otimes dt + dx \otimes dx + dy \otimes dy.$$
Let $c : \mathbb{R} \to \mathbb{R}^3$ be the curve $c(t) = (t, \cos t, \sin t)$. Show that although $\dot{c}(t)$ is null for all $t \in \mathbb{R}$ we have $c(t) \in I^+(c(0))$ for all $t > 0$. What kind of motion does this curve represent?
(5) Let (M, g) be a causally stable spacetime and h an arbitrary $(2, 0)$-tensor field with compact support. Show that for sufficiently small $\varepsilon > 0$ the tensor field $g_\varepsilon = g + \varepsilon h$ is still a Lorentzian metric on M, and (M, g_ε) satisfies the chronology condition.
(6) Let (M, g) be the quotient of Minkowski 2-dimensional spacetime by the discrete group of isometries generated by the map $f(t, x) = (t + 1, x + 1))$. Show that (M, g) satisfies the chronology condition, but there exist arbitrarily small perturbations of (M, g) (in the sense of Exercise 7.8.5) which do not.
(7) Let (M, g) be a time oriented spacetime and $S \subset M$. Show that:

(a) $S \subset D^+(S)$;
(b) $D^+(S)$ is not necessarily open;
(c) $D^+(S)$ is not necessarily closed;
(d) if U is an open set such that $H = \partial D^+(S) \cap U$ is a hypersurface, then the normal vector to H is null;
(e) H is ruled by null geodesics.

(8) Show that the following spacetimes are globally hyperbolic:
(a) Minkowski spacetime;
(b) Friedmann-Robertson-Walker spacetimes;
(c) The region $\{r > 2m\}$ of Schwarzschild spacetime;
(d) The region $\{r < 2m\}$ of Schwarzschild spacetime.

(9) Let (M, g) be the 2-dimensional spacetime obtained by removing the positive x-semi-axis of Minkowski 2-dimensional spacetime (cf. Figure 7). Show that:
(a) (M, g) is stably causal but not globally hyperbolic.
(b) There exist points $p, q \in M$ such that $J^+(p) \cap J^-(q)$ is not compact.
(c) There exist points $p, q \in M$ with $q \in I^+(p)$ such that the supremum of the lengths of timelike curves connecting p to q is not attained by any timelike curve.

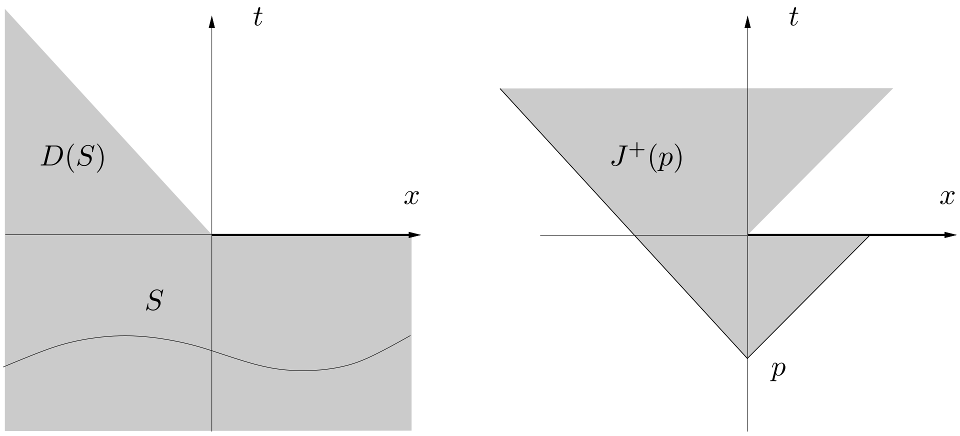

FIGURE 7. Stably causal but not globally hyperbolic spacetime.

(10) Let (Σ, h) be a 3-dimensional Riemannian manifold. Show that the spacetime $(M, g) = (\mathbb{R} \times \Sigma, -dt \otimes dt + h)$ is globally hyperbolic *iff* (Σ, h) is complete.

(11) Let (M, g) be a global hyperbolic spacetime with Cauchy surface S. Show that M is diffeomorphic to $\mathbb{R} \times S$.

8. Singularity Theorem

As we have seen in Sections 5 and 6, both the Schwarzschild solution and the Friedmann-Robertson-Walker cosmological models display singularities, beyond which timelike geodesics cannot be continued.

DEFINITION 8.1. *A spacetime (M,g) is* **singular** *if it is not geodesically complete.*

It was once thought that the examples above were singular due to their high degree of symmetry, and that more realistic spacetimes would be non-singular. Following Hawking and Penrose (cf. [**Pen65, Haw67, HP70**]), we will show that this is not the case: any sufficiently small perturbation of these solutions will still be singular.

The question of whether a given Riemannian manifold is geodesically complete is settled by the Hopf-Rinow Theorem. Unfortunately, this theorem does not hold on Lorentzian geometry (essentially because one cannot use the metric to define a distance function). For instance, compact manifolds are not necessarily geodesically complete (cf. Exercise 8.12.1), and the exponential map is not necessarily surjective in geodesically complete manifolds (cf. Exercise 8.12.2).

Let (M,g) be a globally hyperbolic spacetime and S a Cauchy hypersurface with future-pointing normal vector field n. Let c_p be the timelike geodesic with initial condition n_p for each point $p \in S$. We define a smooth map $\exp : U \to M$ on an open set $U \subset \mathbb{R} \times S$ containing $\{0\} \times S$ as $\exp(t,p) = c_p(t)$.

DEFINITION 8.2. *The critical values of* $\exp$ *are said to be* **conjugate points** *to S.*

Loosely speaking, conjugate points are points where geodesics starting orthogonally at nearby points of S intersect.

Let $q = \exp(t_0, p)$ be a point not conjugate to S, and let (x^1, x^2, x^3) be local coordinates on S around p . Then (t, x^1, x^2, x^3) are local coordinates on some open set $V \ni q$. Since $\frac{\partial}{\partial t}$ is the unit tangent field to the geodesics orthogonal to S, we have $g_{00} = \left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial t} \right\rangle = -1$. On the other hand,

$$\begin{aligned}\frac{\partial g_{0i}}{\partial t} &= \frac{\partial}{\partial t}\left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial x^i}\right\rangle = \left\langle \frac{\partial}{\partial t}, \nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial x^i}\right\rangle \\ &= \left\langle \frac{\partial}{\partial t}, \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial t}\right\rangle = \frac{1}{2}\frac{\partial}{\partial x^i}\left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right\rangle = 0,\end{aligned}$$

and since $g_{0i} = 0$ on S we have $g_{0i} = 0$ on V. Therefore the surfaces of constant t are orthogonal to the geodesics tangent to $\frac{\partial}{\partial t}$; for this reason, (t, x^1, x^2, x^3) is said to be a **synchronized** coordinate system. On this coordinate system we have

$$g = -dt \otimes dt + \sum_{i,j=1}^{3} \gamma_{ij}(t) dx^i \otimes dx^j,$$

where the functions γ_{ij} define a positive definite matrix. This matrix is well defined along c_p, even at points where the synchronized coordinate system breaks down. These are the points along c_p which are conjugate to S, and are also those where $\gamma(t) = \det(\gamma_{ij}(t))$ vanishes, since only then will $\left\{\frac{\partial}{\partial t}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right\}$ fail to be linearly independent.

It is easy to see that

$$\Gamma^0_{00} = \Gamma^i_{00} = 0 \quad \text{ and } \quad \Gamma^i_{0j} = \sum_{k=1}^{3} \gamma^{ik}\beta_{kj},$$

where $(\gamma^{ij}) = (\gamma_{ij})^{-1}$ and $\beta_{ij} = \frac{1}{2}\frac{\partial \gamma_{ij}}{\partial t}$. Consequently,

$$\begin{aligned} R_{00} &= \sum_{i=1}^{3} R_{i00}{}^{i} = \sum_{i=1}^{3} \left(\frac{\partial \Gamma^i_{00}}{\partial x^i} - \frac{\partial \Gamma^i_{i0}}{\partial t} + \sum_{j=1}^{3} \Gamma^j_{00}\Gamma^i_{ij} - \sum_{j=1}^{3} \Gamma^j_{i0}\Gamma^i_{0j} \right) \\ &= -\frac{\partial}{\partial t} \sum_{i,j=1}^{3} \gamma^{ij}\beta_{ij} - \sum_{i,j,k,l=1}^{3} \gamma^{ik}\gamma^{jl}\beta_{ij}\beta_{kl}. \end{aligned}$$

(cf. Chapter 4, Section 1). The quantity

$$\theta = \sum_{i,j=1}^{3} \gamma^{ij}\beta_{ij}$$

appearing in this expression is called the **expansion** of the synchronized observers, and has an important geometric meaning:

$$\theta = \frac{1}{2}\operatorname{tr}\left((\gamma_{ij})^{-1} \frac{\partial}{\partial t}(\gamma_{ij}) \right) = \frac{1}{2}\frac{\partial}{\partial t}\log\gamma = \frac{\partial}{\partial t}\log\gamma^{\frac{1}{2}}$$

where we have used the formula

$$(\log\det A)' = \operatorname{tr}\left(A^{-1}A'\right)$$

for any smooth matrix function $A : \mathbb{R} \to GL(n)$ (cf. Example 7.1.4 in Chapter 1). Therefore the expansion measures the variation of the spatial volume spanned by neighboring synchronized observers. More importantly for our purposes, we see that a singularity of the expansion indicates a zero of γ, i.e. a conjugate point to S.

Definition 8.3. *A spacetime (M, g) is said to satisfy the* **strong energy condition** *if $Ric(V, V) \geq 0$ for any timelike vector field $V \in \mathfrak{X}(M)$.*

By the Einstein equation, this is equivalent to requiring that the reduced energy-momentum tensor T satisfies $T(V, V) \geq 0$ for any timelike vector field $V \in \mathfrak{X}(M)$. In the case of a pressureless fluid with rest density function $\rho \in C^\infty(M)$ and unit velocity vector field $U \in \mathfrak{X}(M)$, this requirement becomes

$$\rho\left(\langle U, V\rangle^2 + \frac{1}{2}\langle V, V\rangle \right) \geq 0,$$

or, since the term in brackets is always positive (cf. Exercise 8.12.3), simply $\rho \geq 0$. For more complicated matter models, the strong energy condition produces equally reasonable restrictions.

PROPOSITION 8.4. *Let (M, g) be a globally hyperbolic spacetime satisfying the strong energy condition, $S \subset M$ a Cauchy hypersurface and $p \in S$ be a point where $\theta = \theta_0 < 0$. Then the geodesic c_p contains at least a point conjugate to S, at a distance of at most $-\frac{3}{\theta_0}$ to the future of S.*

PROOF. Since (M, g) satisfies the strong energy condition, we have $R_{00} = Ric\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) \geq 0$ on any synchronized frame. Consequently,

$$\frac{\partial \theta}{\partial t} + \sum_{i,j,k,l=1}^{3} \gamma^{ik}\gamma^{jl}\beta_{ij}\beta_{kl} \leq 0$$

on such a frame. Using the identity

$$(\operatorname{tr} A)^2 \leq n \operatorname{tr}(A^t A),$$

which holds for square $n \times n$ matrices (as a simple consequence of the Cauchy-Schwarz inequality), it is easy to show that

$$\sum_{i,j,k,l=1}^{3} \gamma^{ik}\gamma^{jl}\beta_{ij}\beta_{kl} \geq \frac{1}{3}\theta^2.$$

Consequently θ must satisfy

$$\frac{\partial \theta}{\partial t} + \frac{1}{3}\theta^2 \leq 0.$$

Integrating this inequality yields

$$\frac{1}{\theta} \geq \frac{1}{\theta_0} + \frac{t}{3},$$

and hence θ must blow up at a value of t no greater than $-\frac{3}{\theta_0}$. $\square$

PROPOSITION 8.5. *Let (M, g) be a globally hyperbolic spacetime, S a Cauchy hypersurface, $p \in M$ and c a timelike geodesic through p orthogonal to S. If there exists a conjugate point between S and p then c does not maximize length (among the timelike curves connecting S to p).*

PROOF. We will offer only a sketch of the proof. Let q be the first conjugate point along c between S and p. Then we can use a synchronized coordinate system around the portion of c between S and q. Since q is conjugate to S, there exists another geodesic $\tilde{c}$, orthogonal to S, with the same (approximate) length $t(q)$, which (approximately) intersects c at q. Let V be a geodesically convex neighborhood of q, $r \in V$ a point along $\tilde{c}$ between S and q, and $s \in V$ a point along c between q and p (cf. Figure 8). Then the piecewise smooth timelike curve obtained by following $\tilde{c}$ between S and r, the unique geodesic in V between r and s, and c between s and p connects S to p and has strictly bigger length than c (by the generalized twin paradox). This curve can be easily smoothed while retaining bigger length than c. $\square$

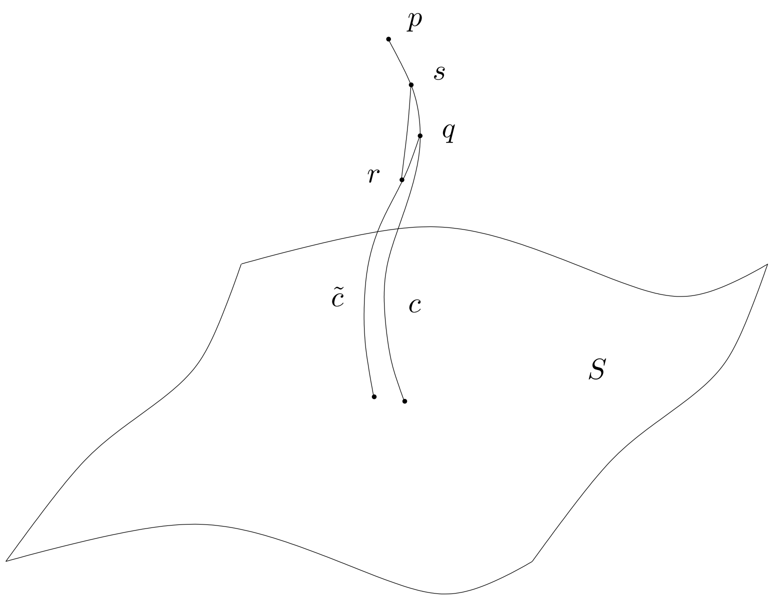

FIGURE 8. Proof of Proposition 8.5.

PROPOSITION 8.6. *Let (M, g) be a globally hyperbolic spacetime, S a Cauchy hypersurface and $p \in D^+(S)$. Then $D^+(S) \cap J^-(p)$ is compact.*

PROOF. Let us define a **simple neighborhood** $U \subset M$ to be a geodesically convex open set diffeomorphic to an open ball whose boundary is a compact submanifold of a geodesically convex open set (therefore ∂U is diffeomorphic to S^3 and $\overline{U}$ is compact). It is clear that simple neighborhoods form a basis for the topology of M. Also, it is easy to show that any open cover $\{V_\alpha\}_{\alpha \in A}$ has a countable, locally finite refinement $\{U_n\}_{n \in \mathbb{N}}$ by simple neighborhoods (cf. Exercise 8.12.5).

If $A = D^+(S) \cap J^-(p)$ were not compact, there would exist a countable, locally finite open cover $\{U_n\}_{n \in \mathbb{N}}$ of A by simple neighborhoods not admitting any finite subcover. Take $q_n \in A \cap U_n$ such that $q_m \neq q_n$ for $m \neq n$. The sequence $\{q_n\}_{n \in \mathbb{N}}$ cannot have accumulation points, for any point in M has a neighborhood intersecting only finite simple neighborhoods U_n. Consequently, each simple neighborhood U_n contains only finite points in the sequence (as $\overline{U_n}$ is compact).

Set $p_1 = p$. Since $p_1 \in A$, we have $p_1 \in U_{n_1}$ for some $n_1 \in \mathbb{N}$. Let $q_n \notin U_{n_1}$. Since $q_n \in J^-(p_1)$, there exists a future-directed causal curve c_n connecting q_n to p_1. This curve will necessarily intersect ∂U_{n_1}. Let $r_{1,n}$ be an intersection point. Since U_{n_1} contains only finite points in the sequence $\{q_n\}_{n\in\mathbb{N}}$, there will exist infinite intersection points $r_{1,n}$. As ∂U_{n_1} is compact, these will accumulate to some point $p_2 \in \partial U_{n_1}$.

Because $\overline{U_{n_1}}$ is contained in a geodesically convex open set, $p_2 \in J^-(p_1)$: if $\gamma_{1,n}$ is the unique causal geodesic connecting p_1 to $r_{1,n}$, parametrized by the global time function $t : M \to \mathbb{R}$, then the subsequence of $\{\gamma_{1,n}\}$ corresponding to a convergent subsequence of $\{r_{1,n}\}$ will converge to a causal geodesic γ_1 connecting p_1 to p_2. Since $t(r_{1,n}) \geq 0$, we have $t(p_2) \geq 0$, and therefore $p_2 \in A$. Since $p_2 \notin U_{n_1}$, there must exist $n_2 \in \mathbb{N}$ such that $p_2 \in U_{n_2}$.

Since U_{n_2} contains only finite points in the sequence $\{q_n\}_{n\in\mathbb{N}}$, infinite curves c_n must intersect ∂U_{n_2} to the past of $r_{1,n}$. Let $r_{2,n}$ be the intersection points. As ∂U_{n_2} is compact, $\{r_{2,n}\}$ must accumulate to some point $p_3 \in \partial U_{n_2}$. Because $\overline{U_{n_2}}$ is contained in a geodesically convex open set, $p_3 \in J^-(p_2)$: if $\gamma_{2,n}$ is the unique causal geodesic connecting $r_{1,n}$ to $r_{2,n}$, parametrized by the global time function, then the subsequence of $\{\gamma_{2,n}\}$ corresponding to convergent subsequences of both $\{r_{1,n}\}$ and $\{r_{2,n}\}$ will converge to a causal geodesic connecting p_2 to p_3. Since $J^-(p_2) \subset J^-(p_1)$ and $t(r_{2,n}) \geq 0 \Rightarrow t(p_3) \geq 0$, we have $p_3 \in A$.

Iterating the procedure above, we can construct a sequence $\{p_i\}_{i\in\mathbb{N}}$ of points in A satisfying $p_i \in U_{n_i}$ with $n_i \neq n_j$ if $i \neq j$, such that p_i is connected p_{i+1} by a causal geodesic γ_i. It is clear that γ_i cannot intersect S, for $t(p_{i+1}) > t(p_{i+2}) \geq 0$. On the other hand, the piecewise smooth causal curve obtained by joining the curves γ_i can easily be smoothed into a past-directed causal curve starting at p_1 which does not intersect S. Finally, such curve is inextendible: it cannot converge to any point, as $\{p_i\}_{i\in\mathbb{N}}$ cannot accumulate. But since $p_1 \in D^+(S)$, this curve would have to intersect S. Therefore A must be compact. $\square$

Corollary 8.7. *Let (M, g) be a globally hyperbolic spacetime and $p, q \in M$. Then:*

(i) $J^+(p)$ is closed;
(ii) $J^+(p) \cap J^-(q)$ is compact.

We leave the proof of this corollary as an easy exercise. Proposition 8.6 is a key ingredient in establishing the following fundamental result:

Theorem 8.8. *Let (M, g) be a globally hyperbolic spacetime with Cauchy hypersurface S, and $p \in D^+(S)$. Then among all timelike curves connecting p to M there exists a timelike curve with maximal length. This curve is a timelike geodesic, orthogonal to S.*

Proof. Consider the set $T(S, p)$ of all timelike curves connecting S to p. Since we can always use the global time function $t : M \to \mathbb{R}$ as a parameter, these curves are determined by their images, which are compact subsets of the compact set $A = D^+(S) \cap J^-(p)$. As is well known (cf. [**Mun00**]), the set $C(A)$ of all compact subsets of A is a compact metric space for the

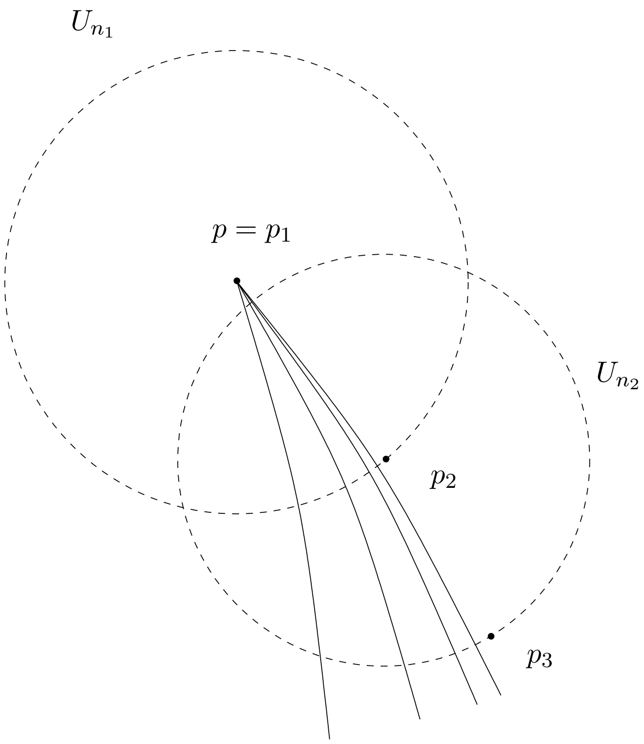

FIGURE 9. Proof of Proposition 8.6.

Hausdorff metric d_H, defined as follows: if $d : M \times M \to \mathbb{R}$ is a metric yielding the topology of M,

$$d_H(K, L) = \inf\{\varepsilon > 0 \mid K \subset U_\varepsilon(L) \text{ and } L \subset U_\varepsilon(K)\},$$

where $U_\varepsilon(K)$ is a ε-neighborhood of K for the metric d. Therefore, the closure $C(S, p) = \overline{T(S, p)}$ is a compact subset of $C(A)$. It is not difficult to show that $C(S, p)$ can be identified with the set of **continuous causal curves** connecting S to p (a continuous curve $c : [0, t(p)] \to M$ is said to be **causal** if $c(t_2) \in J^+(c(t_1))$ whenever $t_2 > t_1$).

The length function $\tau : T(S, p) \to \mathbb{R}$ is defined by

$$\tau(c) = \int_0^{t(p)} |\dot{c}(t)| dt.$$

This function is **upper semicontinuous**, i.e. continuous for the topology

$$\mathcal{O} = \{(-\infty, a) \mid -\infty \leq a \leq +\infty\}$$

in $\mathbb{R}$. Indeed, let $c \in T(S, p)$ be parametrized by its arclength $\mathcal{T}$. For sufficiently small $\varepsilon > 0$, the function $\mathcal{T}$ can be extended to the ε-neighborhood $U_\varepsilon(c)$ in such a way that its level hypersurfaces are spacelike and orthogonal to c (i.e. $-\operatorname{grad} \mathcal{T}$ is timelike and coincides with $\dot{c}$ on c), and $S \cap U_\varepsilon(c)$ is one of these surfaces. If $\gamma \in T(S,p)$ is in the open ball $B_\varepsilon(c) \subset C(A)$ then we can use $\mathcal{T}$ as a parameter, thus obtaining

$$\dot{\gamma} \cdot \mathcal{T} = 1 \Leftrightarrow \langle \dot{\gamma}, \operatorname{grad} \mathcal{T} \rangle = 1.$$

Therefore $\dot{\gamma}$ can be decomposed as

$$\dot{\gamma} = \frac{1}{\langle \operatorname{grad} \mathcal{T}, \operatorname{grad} \mathcal{T} \rangle} \operatorname{grad} \mathcal{T} + X$$

where X is spacelike and orthogonal to $\operatorname{grad} \mathcal{T}$. Consequently,

$$\tau(\gamma) = \int_0^{\tau(c)} |\dot{\gamma}| \, d\mathcal{T} = \int_0^{\tau(c)} \left| \frac{1}{\langle \operatorname{grad} \mathcal{T}, \operatorname{grad} \mathcal{T} \rangle} + \langle X, X \rangle \right|^{\frac{1}{2}} d\mathcal{T}.$$

Given $\delta > 0$, we can choose $\varepsilon > 0$ sufficiently small so that

$$-\frac{1}{\langle \operatorname{grad} \mathcal{T}, \operatorname{grad} \mathcal{T} \rangle} < \left(1 + \frac{\delta}{\tau(c)} \right)^2$$

on the ε-neighborhood $U_\varepsilon(c)$ (as $\langle \operatorname{grad} \mathcal{T}, \operatorname{grad} \mathcal{T} \rangle = -1$ on c). Consequently,

$$\begin{aligned} \tau(\gamma) &= \int_0^{\tau(c)} \left| -\frac{1}{\langle \operatorname{grad} \mathcal{T}, \operatorname{grad} \mathcal{T} \rangle} - \langle X, X \rangle \right|^{\frac{1}{2}} d\mathcal{T} \\ &< \int_0^{\tau(c)} \left(1 + \frac{\delta}{\tau(c)} \right) d\mathcal{T} = \tau(c) + \delta, \end{aligned}$$

proving upper semicontinuity. As a consequence, the length function and can be extended to $C(S,p)$ through

$$\tau(c) = \lim_{\varepsilon \to 0} \sup \{ \tau(\gamma) \mid \gamma \in B_\varepsilon(c) \cap T(S,p) \}$$

(as for $\varepsilon > 0$ sufficiently small the supremum will be finite). Also, it is clear that if $c \in T(S,p)$ then the upper semicontinuity of the length forces the two definitions of $\tau(c)$ to coincide. The extension of the length function to $C(S,p)$ is trivially upper semicontinuous: given $c \in C(S,p)$ and $\delta > 0$, let $\varepsilon > 0$ be such that $\tau(\gamma) < \tau(c) + \frac{\delta}{2}$ for any $\gamma \in B_{2\varepsilon}(c) \cap T(S,p)$. Then it is clear that $\tau(c') < \tau(c) + \delta$ for any $c' \in B_\varepsilon(c)$.

Finally, we notice that the compact sets of $\mathbb{R}$ for the topology $\mathcal{O}$ are sets with maximum. Therefore, the length function attains a maximum at some point $c \in C(S,p)$. All that remains to be seen is that the maximum is also attained at a smooth timelike curve γ. To do so, cover c with finitely many geodesically convex neighborhoods and choose points $p_1, \ldots, p_k$ in c such that $p_1 \in S$, $p_k = p$ and the portion of c between p_{i-1} and p_i is contained in a geodesically convex neighborhood for all $i = 2, \ldots, k$. It is clear that there exists a sequence $c_n \in T(S,p)$ such that $c_n \to c$ and $\tau(c_n) \to \tau(c)$. Let $t_i = t(p_i)$ and $p_{i,n}$ be the intersection of c_n with $t^{-1}(t_i)$. Replace c_n by the sectionally geodesic curve γ_n obtained by joining $p_{i-1,n}$ to $p_{i,n}$ in the corresponding geodesically convex neighborhood. Then $\tau(\gamma_n) \geq \tau(c_n)$, and therefore $\tau(\gamma_n) \to \tau(c)$. Since each sequence $p_{i,n}$ converges to p_i, γ_n converges to the sectionally geodesic curve γ obtained by joining p_{i-1} to p_i

$(i = 2, \dots, k)$, and it is clear that $\tau(\gamma_n) \to \tau(\gamma) = \tau(c)$. Therefore γ is a point of maximum for the length. Finally, we notice that γ must be smooth at the points p_i, for otherwise we could increase its length by using the generalized twin paradox. Therefore γ must be a timelike geodesic. Using a synchronized coordinate system around $\gamma(0)$, it is clear that γ must be orthogonal to S, for otherwise it would be possible to increase its length. $\square$

We have now all the necessary ingredients to prove the singularity theorem:

THEOREM 8.9. *Let (M, g) be a globally hyperbolic spacetime satisfying the strong energy condition, and suppose that the expansion satisfies $\theta \leq \theta_0 < 0$ on a Cauchy hypersurface S. Then (M, g) is singular.*

PROOF. We will show that no future-directed timelike geodesic orthogonal to S can be extended to proper time greater than $\tau_0 = -\frac{3}{\theta_0}$ to the future of S. Suppose that this was not so. Then there would exist a future-directed timelike geodesic c orthogonal to S defined in an interval $[0, \tau_0 + 2\varepsilon)$ for some $\varepsilon > 0$. Let $p = c(\tau_0 + \varepsilon)$. According to Theorem 8.8, there would exist a timelike geodesic γ with maximal length connecting S to p, orthogonal to S. Because $\tau(c) = \tau_0 + \varepsilon$, we would necessarily have $\tau(\gamma) \geq \tau_0 + \varepsilon$. Proposition 8.4 guarantees that γ would develop a conjugate point at a distance of at most τ_0 to the future of S, and Proposition 8.5 states that γ would cease to be maximizing beyond this point. Therefore we arrive at a contradiction. $\square$

REMARK 8.10. It should be clear that (M, g) is singular if the condition $\theta \leq \theta_0 < 0$ on a Cauchy hypersurface S is replaced by the condition $\theta \geq \theta_0 > 0$ on S. In this case, no **past-directed** timelike geodesic orthogonal to S can be extended to proper time greater than $\tau_0 = \frac{3}{\theta_0}$ to the **past** of S.

EXAMPLE 8.11.

(1) The Friedmann-Robertson-Walker models are globally hyperbolic (cf. Exercise 7.8.8), and satisfy the strong energy condition (as $\rho > 0$). Furthermore,
$$\beta_{ij} = \frac{\dot{a}}{a}\gamma_{ij} \Rightarrow \theta = \frac{3\dot{a}}{a}.$$
Assume that the model is expanding at time t_0. Then $\theta = \theta_0 = \frac{3\dot{a}(t_0)}{a(t_0)} > 0$ on the Cauchy hypersurface $S = \{t = t_0\}$, and hence Theorem 8.9 guarantees that this model is singular to the past of S (i.e. there exists a Big Bang). Furthermore, Theorem 8.9 implies that this singularity is generic: any sufficiently small perturbation of an expanding Friedmann-Robertson-Walker model satisfying the strong energy condition will also be singular. Loosely speaking, any expanding universe must have begun at a Big Bang.

(2) The region $\{r < 2m\}$ of the Schwarzschild solution is globally hyperbolic (cf. Exercise 7.8.8), and satisfies the strong energy condition (as $Ric = 0$). The metric can be written is this region as

$$g = -d\tau \otimes d\tau + \left(\frac{2m}{r} - 1\right) dt \otimes dt + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi,$$

where

$$\tau = \int_r^{2m} \left(\frac{2m}{u} - 1\right)^{-\frac{1}{2}} du.$$

Therefore the inside of the black hole can be pictured as a cylinder $\mathbb{R} \times S^2$ whose shape is evolving in time. As $r \to 0$, the S^2 contracts to a singularity, with the t-direction expanding. Since

$$\sum_{i,j=1}^{3} \beta_{ij} dx^i \otimes dx^j = \frac{dr}{d\tau}\left(-\frac{m}{r^2} dt \otimes dt + r d\theta \otimes d\theta + r \sin^2\theta d\varphi \otimes d\varphi\right),$$

we have

$$\theta = \left(\frac{2m}{r} - 1\right)^{-\frac{1}{2}} \left(\frac{2}{r} - \frac{3m}{r^2}\right).$$

Therefore $\theta = \theta_0 < 0$ on any Cauchy hypersurface $S = \{r = r_0\}$ with $r_0 < \frac{3m}{2}$, and hence Theorem 8.9 guarantees that the Schwarzschild solution is singular to the future of S. Furthermore, Theorem 8.9 implies that this singularity is generic: any sufficiently small perturbation of the Schwarzschild solution satisfying the strong energy condition will also be singular. Loosely speaking, once the collapse has advanced long enough, nothing can prevent the formation of a singularity.

Exercises 8.12.

(1) (*Clifton-Pohl torus*) Consider the Lorentzian metric

$$\overline{g} = \frac{1}{u^2 + v^2}(du \otimes dv + dv \otimes du)$$

on $\overline{M} = \mathbb{R}^2 \setminus \{0\}$. The Lie group $\mathbb{Z}$ acts freely and properly on $\overline{M}$ by isometries through

$$n \cdot (u, v) = (2^n u, 2^n v),$$

and this determines a Lorentzian metric g on $M = \overline{M}/\mathbb{Z} \cong T^2$. Show that (M, g) is not geodesically complete (although M is compact). (**Hint:** Look for null geodesics with $v \equiv 0$).

(2) (2-*dimensional Anti-de Sitter spacetime*) Consider $\mathbb{R}^3$ with the pseudo-Riemannian metric

$$-du \otimes du - dv \otimes dv + dw \otimes dw,$$

and let (M, g) be the universal covering of the submanifold

$$\{(u, v, w) \in \mathbb{R}^3 \mid u^2 + v^2 - w^2 = 1)\}$$

with the induced metric. Show that:

(a) A model for (M, g) is $M = \mathbb{R} \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and
$$g = \frac{1}{\cos^2 x}(-dt \otimes dt + dx \otimes dx)$$
(hence (M, g) is not globally hyperbolic).

(b) (M, g) is geodesically complete, but $\exp_p$ is not surjective for any $p \in M$. (**Hint:** Notice each isometry of $\mathbb{R}^3$ with the given pseudo-Riemannian metric determines an isometry of (M, g)).

(c) There exist points $p, q \in M$ connected by arbitrarily long timelike curves (cf. Exercise 9).

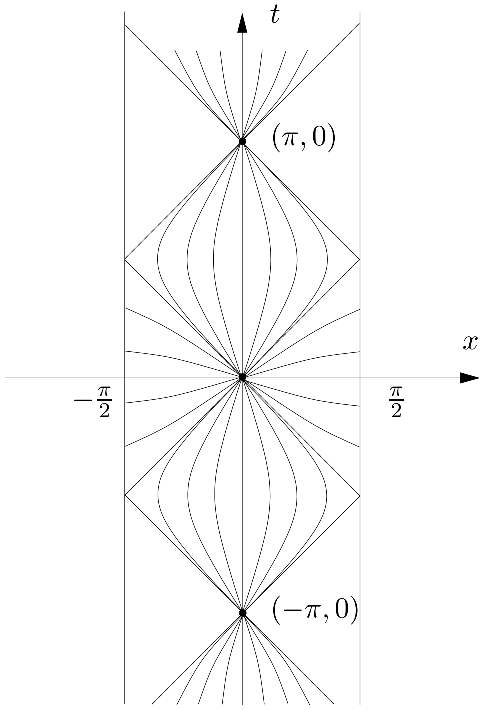

FIGURE 10. The exponential map is not surjective in 2-dimensional Anti-de Sitter space.

(3) Show that if U is a unit timelike vector field and V is any timelike vector field then $\langle U, V\rangle^2 + \frac{1}{2}\langle V, V\rangle$ is a positive function.

(4) Show that a spacetime (M, g) whose matter content is a pressureless fluid with rest density function $\rho \in C^\infty(M)$ and a cosmological constant $\Lambda \in \mathbb{R}$ (cf. Exercise 6.1.7) satisfies the strong energy condition *iff* $\rho \geq \frac{\Lambda}{4\pi}$.

(5) Let (M, g) be a spacetime. Show that any open cover $\{V_\alpha\}_{\alpha \in A}$ has a countable, locally finite refinement $\{U_n\}_{n \in \mathbb{N}}$ by simple neighborhoods (i.e., $\cup_{n \in \mathbb{N}} U_n = \cup_{\alpha \in A} V_\alpha$, for each $n \in \mathbb{N}$ there exists $\alpha \in A$

such that $U_n \subset V_\alpha$, and each point $p \in M$ has a neighborhood which intersects only finite simple neighborhoods U_n).

(6) Prove Corollary 8.7.

(7) Let (M, g) be a globally hyperbolic spacetime, $t : M \to \mathbb{R}$ a global time function, $S = t^{-1}(0)$ a Cauchy hypersurface, $p \in D^+(S)$ and $A = D^+(S) \cap J^-(p)$. Show that the closure $C(S,p) = \overline{T(S,p)}$ in the space $C(A)$ of all compact subsets of A with the Hausdorff metric can be identified with the set of continuous causal curves connecting S to p (parametrized by t).

(8) Show that if (M, g) is a globally hyperbolic spacetime and S is a Cauchy surface then $\exp : U \subset \mathbb{R} \times S \to M$ is surjective.

(9) Let (M, g) be a globally hyperbolic spacetime and $p, q \in M$ with $q \in I^+(p)$. Show that among all timelike curves connecting p to q there exists a timelike curve with maximal length, which is a timelike geodesic.

(10) Use ideas similar to those leading to the proof of Theorem 8.9 to prove the following theorem of Riemannian geometry: if (M, g) is a complete Riemannian manifold whose Ricci curvature satisfies $Ric(X, X) \geq \varepsilon\langle X, X\rangle$ for some $\varepsilon > 0$ then M is compact. Is it possible to prove a singularity theorem in Riemannian geometry?

(11) Explain why each of the following spacetimes does not have to be singular.
   (a) Minkowski spacetime.
   (b) Einstein universe (cf. Exercise 6.1.8).
   (c) de Sitter universe (cf. Exercise 6.1.8).
   (d) 2-dimensional Anti-de Sitter spacetime (cf. Exercise 2).

(12) Prove that any sufficiently small perturbation of the model of collapse in Exercise 6.1.6 is also singular.

9. Notes on Chapter 6

9.1. Bibliographical notes. There are many excellent texts on General Relativity, usually containing also the relevant differential and Lorentzian geometry. These range from introductory [**Sch02**] to more advanced [**Wal84**] to encyclopedic [**MTW73**]. A more mathematically oriented treatment can be found in [**BEE96, O'N83**] ([**GHL04**] also contains a brief glance at pseudo-Riemannian geometry). For more information on Special Relativity and the Lorentz group see [**Nab92, Oli02**]. Causality and the singularity theorems are treated in greater detail in [**Pen87, HE95, Nab88**].

Bibliography

[AM78] R. Abraham and J. Marsden, *Foundations of mechanics*, Addison Wesley, 1978.

[Arn92] V. I. Arnold, *Ordinary differential equations*, Springer, 1992.

[Arn97] V.I. Arnold, *Mathematical methods of classical mechanics*, Springer, 1997.

[Aud96] M. Audin, *Spinning tops*, Cambridge University Press, 1996.

[BEE96] J. Beem, P. Ehrlich, and K. Easley, *Global lorentzian geometry*, Marcel Dekker, 1996.

[BL05] F. Bullo and A. Lewis, *Geometric control of mechanical systems*, Springer, 2005.

[Blo96] E. Bloch, *A first course in geometric topology and differential geometry*, Birkäuser, 1996.

[Blo03] A. Bloch, *Nonholonomic mechanics and control*, Springer, 2003.

[Boo03] W. Boothby, *An introduction to differentiable manifolds and riemannian geometry*, Academic Press, 2003.

[BT82] R. Bott and L. Tu, *Differential forms in algebraic topology*, Springer, 1982.

[BtD03] T. Bröcker and T. tom Diek, *Representations of compact lie groups*, Springer, 2003.

[CB97] R. Cushmann and L. Bates, *Global aspects of classical integrable systems*, Birkhäuser, 1997.

[dC76] M. do Carmo, *Differential geometry of curves and surfaces*, Prentice-Hall, 1976.

[dC93] ______, *Riemannian geometry*, Birkhäuser, 1993.

[dC94] ______, *Differential forms and applications*, Springer, 1994.

[DK99] J. Duistermaat and J. Kolk, *Lie groups*, Springer, 1999.

[Fre82] M. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), no. 3, 357–453.

[GHL04] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, Springer, 2004.

[Gom83] R. Gompf, *Three exotic $\mathbb{R}^4$'s and other anomalies*, J. Differential Geom. **18** (1983), no. 2, 317–328.

[GP73] V. Gillemin and A. Pollack, *Differential topology*, Prentice-Hall, 1973.

[GPS02] H. Goldstein, C. Poole, and J. Safko, *Classical mechanics*, Addison Wesley, 2002.

[Gro70] M. Gromov, *Isometric embeddings and immersions in riemannian geometry*, Usp. Mat. Nauk **25** (1970), 3–62.

[Haw67] S. Hawking, *The occurrence of singularities in cosmology. iii. causality and singularities*, Proc. Roy. Soc. Lon. A **300** (1967), 187–201.

[HE95] S. Hawking and G. Ellis, *The large scale structure of space-time*, Cambridge University Press, 1995.

[Hel01] S. Helgasson, *Differential geometry, lie groups and symmetric spaces*, American Mathematical Society, 2001.

[HP70] S. Hawking and R. Penrose, *The singularities of gravitational collapse and cosmology*, Proc. Roy. Soc. Lon. A **314** (1970), 529–548.

[Jos02] J. Jost, *Riemannian geometry and geometric analysis*, Springer, 2002.

[Ker60] M. Kervaire, *A manifold wich does not admit any differentiable structure*, Comment. Math. Helv. **34** (1960), 257–270.

[KM63] M. Kervaire and J. Milnor, *Groups of homotopy spheres. i.*, Ann. of Math. (2) **77** (1963), 504–537.

[KN96] S. Kobayashi and K. Nomizu, *Foundations of differential geometry*, vol. I and II, Wiley, 1996.

[Mil56a] J. Milnor, *On manifolds homeomorphic to the 7-sphere*, Ann. of Math. **64** (1956), no. (2), 399–405.

[Mil56b] ______, *On the relationship between differentiable manifolds and combinatorial manifolds*, Collected Papers III: Differential Topology, 1956, (in preparation).

[Mil59] ______, *Differentiable structures on spheres*, Amer. J. Math. **81** (1959), 962–972.

[Mil97] ______, *Topology from the differentiable viewpoint*, Princeton University Press, 1997.

[Mor98] F. Morgan, *Riemannian geometry*, A K Peters, 1998.

[MR99] J. Marsden and T. Ratiu, *Introduction to mechanics and symmetry*, Springer, 1999.

[MTW73] C. Misner, K. Thorne, and J. A. Wheeler, *Gravitation*, Freeman, 1973.

[Mun00] J. Munkres, *Topology*, Prentice-Hall, 2000.

[Nab88] G. Naber, *Spacetime and singularities – an introduction*, Cambridge University Press, 1988.

[Nab92] ______, *The geometry of minkowski spacetime*, Springer, 1992.

[Nas56] J. Nash, *The imbedding problem for riemannian manifolds*, Annals of Mathematics **63** (1956), 20–63.

[Nov65] S. P. Novikov, *Topological invariance of rational pontrjagin classes*, Soviet Math. Dokl. **6** (1965), 921–923.

[Oli02] W. Oliva, *Geometric mechanics*, Springer, 2002.

[O'N83] B. O'Neill, *Semi-riemannian geometry*, Academic Press, 1983.

[Pen65] R. Penrose, *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. **14** (1965), 57–59.

[Pen87] ______, *Techniques of differential topology in relativity*, Society for Industrial and Applied Mathematics, 1987.

[Sch02] B. Schutz, *A first course in general relativity*, Cambridge University Press, 2002.

[Sma60] S. Smale, *The generalized poincaré conjecture in higher dimensions*, Bull. AMS **66** (1960), 373–375.

[Wal84] R. Wald, *General relativity*, University of Chicago Press, 1984.

[War83] F. Warner, *Foundations of differentiable manifolds and lie groups*, Springer, 1983.

[Whi44a] H. Whitney, *The selfintersections of a smooth n-manifold in $(2n-1)$-space*, Annals of Mathematics **45** (1944), 247–293.

[Whi44b] ______, *The selfintersections of a smooth n-manifold in $2n$-space*, Annals of Mathematics **45** (1944), 220–246.

[Wol78] J. A. Wolf, *Spaces of constant curvature*, Publish or Perish, 1978.

Index