Handbook of geometry and all that...
(A work in progress)

Edoardo Provenzi

Contents

Preface

This handbook is intended to be a self-contained reference for the most fundamental methods of differential and hyperbolic geometry, together with some applications to physics.

It must not be intended as an exhaustive treatise but as a (hopefully) clear exposition of these topics. In particular, we have tried to reduce to the minimum one of the major problems reported by students when learning for the first time differential geometry: notation! Formulae in differential geometry can easily become notationally unbearable if a bad choice of notation is performed. This implies that some reasonable shortcut must be implicitly assumed to avoid this problem and keep equations as readable and meaningful as possible.

Coherently with our main concern, the major sources of inspiration for our handbook (among others, that are duly quoted) are listed below.

- The extremely clear videos about differential geometry by Francesco Bottacin, professor at the university of Padova, Italy. They are available online (in Italian) at the following url: https://www.math.unipd.it/~bottacin/geomdiff.htm. A great deal of this handbook can be thought as a free translation of his notes and videos. Professor Bottacin is warmly acknowledged.

- J. Lee's treatise: 'Introduction to smooth manifolds' [10], one of the clearest, most complete, introductory books about differential geometry.

- C. Isham's splendid big little book [6], for once, a book about mathematical concepts written for physicists that does not treat them as 'dummies'.

- J.G. Ratcliffe's book: 'Foundations of hyperbolic manifolds' [18], to our knowledge, the treatise on hyperbolic geometry that fits best with the spirit of this handbook.

Of course, every mistake in this document must be referred to me and not to the books and material quoted above.

The author.

Part I:

Introduction to differential geometry

In many cases, proofs based on coordinate free local representations in charts are clearer than proofs which are repleate with the claws of a rather unpleasant prying insect such as Γ^i_{jkl}.
S. Lang, 'Differential and Riemannian Manifolds', 1995

Chapter 1

Differential manifolds: definitions and basic properties

> Determinations of measure require magnitude to be independent of location, a state of things which can occur in more than one way.
>
> B. Riemann, 1854

In this first chapter we introduce the basic definitions and properties of differential manifolds. The reader not used to Einstein's convention for sum over repeated indices and differential calculus in $\mathbb{R}^n$ is referred to the appendices.

1.1 Differential manifolds

The first mathematician to conceive the idea of what we call today a differential manifold was Bernhard **Riemann** $(1826 - 1866)$ who, in his groundbreaking 1854 habilitation defense [20], introduced the concept of an abstract manifold not necessarily embedded in a Euclidean space, as, instead, it was thought by his PhD advisor, the prince of mathematicians C.F. **Gauss** $(1777 - 1855)$.

Riemann's ideas have been further refined until the modern definition of differential manifold that we report in this document, first introduced in the literature by Charles **Ehresmann** $(1905 - 1979)$ [4] in 1943. In this definition a (finite dimensional) differential manifold is seen as a topological space (with some suitable requests to make calculus easier) with the fundamental requirement to be locally identifiable with a model space, which is a topological vector space.

The reason for considering topological vector spaces as local models lies in the fact that one of the fundamental elements of calculus, the derivative, represents a *local linearization* of a function, which explains the need of a linear structure on the model space that makes it a vector space. Moreover, the computation of derivatives requires the concept of limit, which implies that a topology coherent with the linear structure should be present. Finally, the fact that derivatives are defined in a local neighborhoods of points will allow us **transporting the differential structure of topological vector spaces to more general topological spaces that 'resemble' to them only locally**.

This local resemblance is provided by means of **homeomorphisms**, i.e. **bicontinuous** maps between topological spaces (continuous bijective functions with a continuous inverse). Depending on the particular choice of topological vector space that is considered as local model, different differential manifolds can be defined. Classically, the local model is chosen to be $\mathbb{R}^n$, $n < +\infty$, but of course it can be $\mathbb{C}^n$ or an infinite-dimensional Frechet, Banach or Hilbert space and so on. Here, the local model will always be $\mathbb{R}^n$.

Before going through the details of differential manifolds, let us spend just a few words on topological manifolds.

Def. 1.1.1 (Topological manifold) *The couple given by a connected topological space M and a set of couples $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ (where A is an index set, U_α are open subsets of M), satisfying:*

- $M = \bigcup_{\alpha \in A} U_\alpha$, *i.e. the union of the sets U_α covers M*
- $\varphi_\alpha : U_\alpha \to \mathbb{R}^n$ *are homeomorphisms*[1]*,*

is said to be a topological manifold of dimension n.

The definition of the dimension is well posed, in fact either there is a single homeomorphism that covers M, and so n is univocally defined, or at least the domain of two homeomorphisms has a non empty intersection. Suppose that these homeomorphisms are $\varphi_\alpha : U_\alpha \to \mathbb{R}^n$ and $\varphi_\beta : U_\beta \to \mathbb{R}^m$, with $U_\alpha \cap U_\beta = U_{\alpha\beta} \neq \varnothing$. Then $\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_{\alpha\beta}) \subseteq \mathbb{R}^n \to \varphi_\beta(U_{\alpha\beta}) \subset \mathbb{R}^m$ is a homeomorphism (as composition of homeomorphisms), this implies that $n = m$ because it cannot exist a homeomorphism between $\mathbb{R}^n$ and $\mathbb{R}^m$ if $n \neq m$, see e.g. [10]. Thus, n is an invariant in the definition of a topological manifold.

The fact that M is locally homeomorphic to an open set of $\mathbb{R}^n$ guarantees that, **locally, a topological manifold M defined as before has all the properties of $\mathbb{R}^n$**, e.g. M **is locally connected** (and locally connected by paths) and M **is locally compact**, i.e. every point $p \in M$ has a compact neighborhood, i.e., there exists an open set $U \subset M$ and a compact set $K \subset M$, such that $x \in U \subseteq K$. Other properties, e.g. the Hausdorff and second countable property, must be separately required.

Let us now move a step forward towards the concept of differential manifold.

Def. 1.1.2 *A topological space M is a **locally Euclidean space of dimension n** $\in \mathbb{N}$, $n < +\infty$, if:*

1. *it is a **Hausdorff space**[2]: for every couple of elements $p, q \in M$, there exist two open neighborhoods U_p and U_q such that $U_p \cap U_q = \varnothing$;*
2. *it is **second countable**[3]: there exists a countable collection $\mathcal{U} = \{U_i\}_{i=1}^{\infty}$ of open subsets of M such that any open subset of M can be written as a union of elements of some subfamily of $\mathcal{U}$;*

[1] i.e. bicontinuous functions: continuous invertible functions with continuous inverse. Homeomorphisms are open and closed maps, i.e. they send open sets in open sets and closed sets in closed sets. This implies that $\varphi_\alpha(U_\alpha)$ **is open in** $\mathbb{R}^n$, to see this explicitly, notice that $\varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha) \to U_\alpha$, so $\varphi_\alpha(U_\alpha) = (\varphi_\alpha^{-1})^{-1}(U_\alpha)$, the anti-image of the continuous function φ_α^{-1} of the open set U_α.

[2] The Hausdorff property serves to assure that convergent sequences in M have a **unique limit**.

[3] The second countability is needed to assure the existence of a **partition of unity**, an essential tool to extend local objects to global ones.

3. *it is **locally homeomorphic to** $\mathbb{R}^n$: for every point $p \in M$ it exists an open neighborhood $U \subseteq M$ containing p and a homeomorphism:*

$$\begin{array}{rrcl} \varphi: & U \subseteq M & \xrightarrow{\sim} & \varphi(U) \equiv V \subseteq \mathbb{R}^n \\ & p & \longmapsto & \varphi(p) = x = (x^1, \dots, x^n). \end{array}$$

*The couple (U, φ) is called a **local chart in p**, it is said to be **centered in p** if $\varphi(p) = 0 \in \mathbb{R}^n$. U is called **chart domain** and φ **chart map**.*

1.1.1 Local coordinates of a point

We are going to show that it is always possible to represent the position of any point p in a manifold M of dimension n with the coordinates of the local model $\mathbb{R}^n$ as long as we remain inside the chart domain U of a local chart (U, φ) in p.

The first step consists of course in applying the chart function φ to p to obtain the vector $x = \varphi(p)$ which lives in an open subset of $\mathbb{R}^n$ and the second step consists simply in extracting its components by using the functionals ε^j of the dual canonical basis of $\mathbb{R}^n$. The composition of these two steps gives rise to the following real-valued functions:

$$\begin{array}{rrcl} x^j: & U \subseteq M & \longrightarrow & \mathbb{R} \\ & p & \longmapsto & x^j(p) = (\varepsilon^j \circ \varphi)(p). \end{array}$$

The x^j's are nothing but *the components functions of φ interpreted as a vector-valued function*[4], thus we can write:

$$\varphi \equiv (x^1, \dots, x^n), \quad \text{or} \quad \varphi \equiv (x^j)_{j=1}^n.$$

Def. 1.1.3 (Local coordinates) *The locally-defined real-valued functions*

$$x^j \equiv \varepsilon^j \circ \varphi : U \to \mathbb{R}$$

*are called **local coordinate functions** and the couple $(U, (x^j))$ is said to be a **local coordinate system** in p, $j = 1, \dots, n$.*

A typical **abuse of notation** consists in writing with x^j both the components of the image of $p \in M$ via the local chart φ w.r.t. the canonical basis of $\mathbb{R}^n$, which are *real numbers*, *and* the *real-valued functions* $\varepsilon^j \circ \varphi : U \to \mathbb{R}$.

On one side, this abuse of notation implies the weird formula $x^j(p) = x^j$, however, on the other side, in general it is clear when x^j refers to a function or a to real number and this notational simplification improves enormously the readability of expressions involving coordinates.

Following the idea of transporting the differential structure of $\mathbb{R}^n$ to a locally Euclidean space M, we must assure two things: the first is that **all the points of M are covered by a local chart**, the second is that **two intersecting charts** are compatible in the sense that the differential structure that they induce on M is **not in conflict**. The formalization of these ideas is given in the following definition.

Def. 1.1.4 (Atlas) *Given a locally Euclidean space M of dimension n, an **atlas** for M is a collection of charts $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$, satisfying:*

[4] in fact some author denote them more correctly as φ^j instead of x^j.

1. ***Covering****:* $\{(U_\alpha, \varphi_\alpha)\}_{\alpha\in A}$ *covers* M*, i.e.*
$$M = \bigcup_{\alpha\in A} U_\alpha$$
2. ***Compatibility****: whenever* $U_{\alpha\beta} \equiv U_\alpha \cap U_\beta \neq \varnothing$*, the function:*
$$\begin{array}{rccl} \eta_{\beta\alpha} := \varphi_\beta \circ \varphi_\alpha^{-1}: & \varphi_\alpha(U_{\alpha\beta}) \subseteq \mathbb{R}^n & \longrightarrow & \varphi_\beta(U_{\alpha\beta}) \subseteq \mathbb{R}^n \\ & x & \longmapsto & \tilde{x} := \eta_{\beta\alpha}(x) = \varphi_\beta(\varphi_\alpha^{-1}(x)), \end{array}$$
is ***smooth****, i.e. it belongs to* $\mathscr{C}^\infty(\varphi_\alpha(U_{\alpha\beta}))$*.*

The function $\eta_{\beta\alpha}$ is called **transition function** from the local representation $(U_\alpha, \varphi_\alpha)$ to (U_β, φ_β). It is invertible, being a composition of invertible functions, and its inverse is
$$\eta_{\beta\alpha}^{-1} = \eta_{\alpha\beta} = \varphi_\alpha \circ \varphi_\beta^{-1}.$$

In general, showing that the chart domains of an atlas cover M and the smoothness of the chart maps is not a difficult task. **What requires much work is to verify** the compatibility, i.e. **that the transition functions are smooth**.

If the transition function $\eta_{\beta\alpha}$ is of class $\mathscr{C}^r$, then the compatibility will be called of class $\mathscr{C}^r$, but here we will always consider the smooth compatibility, unless otherwise stated.

By composing the transition functions with the elements of the canonical dual basis of $\mathbb{R}^n$ we obtain the functions that allow us transforming the local coordinates x^j of a point $p \in M$ w.r.t. the chart $(U_\alpha, \varphi_\alpha)$ to the local coordinates $\tilde{x}^j$ w.r.t. the chart (U_β, φ_β):
$$\begin{array}{rccl} \varepsilon^j \circ \eta_{\beta\alpha}: & \varphi_\alpha(U_{\alpha\beta}) \subseteq \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & x = (x^1, \ldots, x^n) & \longmapsto & (\varepsilon^j \circ \eta_{\beta\alpha})(x). \end{array}$$
Notice that $\varepsilon^j \circ \eta_{\beta\alpha}$ are nothing but the component functions of $\eta_{\beta\alpha}$ interpreted as vector-valued functions. Instead of denoting them as $\eta^j_{\beta\alpha}$, it is usual (in particular in Physics books) to write them simply with the symbol $\tilde{x}^j$:
$$\begin{array}{rccl} \tilde{x}^j: & \varphi_\alpha(U_{\alpha\beta}) \subseteq \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & (x^i) & \longmapsto & \tilde{x}^j(x^i) = (\varepsilon^j \circ \eta_{\beta\alpha})(x^i), \end{array}$$
they are called the **local coordinate transformation functions**. The diagram below gives a graphical visualization of the objects just defined.

$$\begin{tikzcd}
& U_{\alpha\beta} \arrow[dr, "\varphi_\beta"] & \\
\mathbb{R}^n \supseteq \varphi_\alpha(U_{\alpha\beta}) \ni x \arrow[ur, "{\varphi_\alpha}^{-1}"] \arrow[rr, dashed, "\eta_{\alpha\beta}"] \arrow[drr, dashed, "\tilde{x}^j(x^i)"'] & & \tilde{x} \in \varphi_\beta(U_{\alpha\beta}) \subseteq \mathbb{R}^n \arrow[d, "\varepsilon^j"] \\
& & \tilde{x}^j \in \mathbb{R}
\end{tikzcd}$$

It should be clear from the context when $\tilde{x}^j$ represents a real number or a real-valued function, in any case, the weird notation $\tilde{x}^j = \tilde{x}^j(x^i)$ must be interpreted as follows:
$$\underset{\text{(real number)}}{\tilde{x}^j} = \underset{\text{(function } \mathbb{R}^n \to \mathbb{R})}{\tilde{x}^j} \underset{(\mathbb{R}^n \text{ vector})}{(x^i)},$$
and similarly for the inverse local coordinate transformation $x^i = x^i(\tilde{x}^j)$.

In general, a point in manifold M has always:

- a **local representation**, thought as an element of the local model $\mathbb{R}^n$, obtained by applying a local chart map;
- a **local coordinate representation**, thought as the vector of $\mathbb{R}^n$ whose components are obtained by further composing the local representation with the functionals of the canonical dual basis of the local model $\mathbb{R}^n$.

We will see that this considerations can be extended also to other objects defined on M, e.g. functions.

The compatibility between local charts can be equivalently stated in coordinates. To understand why, let us first recall the classical inverse function theorem of ordinary calculus in $\mathbb{R}^n$.

Theorem 1.1.1 (Inverse mapping theorem in $\mathbb{R}^n$) *Let:*

- $\Omega \subset \mathbb{R}^n$ *be an open set;*
- $f : \Omega \to \mathbb{R}^n$, $f \in \mathscr{C}^k(\Omega)$, $k \geqslant 1$*;*
- $x_0 \in \Omega$ *such that*[5]*:*
$$\det(Jf(x_0)) \neq 0.$$

Then there exist two neighborhoods $U \subseteq \Omega$ of x_0 and $V \subseteq \mathbb{R}^n$ of $f(x_0)$ such that $f|_U : U \to V$ is a $\mathscr{C}^k$-diffeomorphism.

We can organize the partial derivatives of the local coordinate transformation functions $\tilde{x}^j : \mathbb{R}^n \to \mathbb{R}$ in **the matrix of functions** $\frac{\partial \tilde{x}^j}{\partial x^i} : \mathbb{R}^n \to \mathbb{R}$ defined by $J^j_i := \left(\frac{\partial \tilde{x}^j}{\partial x^i}\right)_{i,j=1,\dots n}$, explicitly:

$$J := \begin{pmatrix} \frac{\partial \tilde{x}^1}{\partial x^1} & \cdots & \frac{\partial \tilde{x}^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{x}^n}{\partial x^1} & \cdots & \frac{\partial \tilde{x}^n}{\partial x^n} \end{pmatrix}.$$

If the determinant of the Jacobian matrix $J(x)$ is not null for every $x \in \varphi_\alpha(U_{\alpha\beta})$, the charts are compatible, i.e.

Compatibility condition between local charts in coordinates:

$$\det J(x) \neq 0 \qquad \forall x \in \varphi_\alpha(U_{\alpha\beta}),$$

where $J(x) \in M(n, \mathbb{R})$, $J(x) = \mathrm{ev}_x \circ J = \left(\frac{\partial \tilde{x}^j}{\partial x^i}(x)\right)_{i,j=1,\dots,n}$, ev_x being the evaluation map of the functions $\frac{\partial \tilde{x}^j}{\partial x^i}$ in x.

[5] The geometrical interpretation of this condition is the following: the fact that the Jacobian matrix of f in x_0 is non-singular guarantees that the total derivative $Df(x_0) \in \mathrm{End}(\mathbb{R}^n)$ is invertible. Since the differential map is the linear approximation of f in a neighborhood of x_0, the result of the theorem says that this is enough to guarantee that, if we consider a sufficiently small neighborhood of x_0, f itself is invertible and its inverse map has the same regularity as f.

Def. 1.1.5 (Equivalent atlases) *Two atlases of a locally Euclidean space are equivalent if all the local charts of the first atlas are compatible with all those of the second atlas.*

Many authors define **two atlases** of a locally Euclidean space **equivalent if their union is again an atlas** for the same locally Euclidean space. Of course the two definitions are equivalent because, if all the local charts of the first are compatible with those of the second, then the covering and compatibility properties are satisfied and so we get an atlas; vice-versa, if the union is an atlas, then, by definition the compatibility of charts must be satisfied.

The adjective *equivalent* is not used by chance, in fact it can be verified that being equivalent is an actual equivalence relation in the set of atlases of locally Euclidean spaces.

This fact gives us the possibility to define the concept of differential manifold without ambiguity.

Def. 1.1.6 (Differential (smooth) manifold) *A differential (smooth) manifold of dimension n is a couple $(M, \mathcal{A})$, where M is a locally Euclidean space of dimension n and $\mathcal{A}$ is an equivalence class of smooth atlases of M. A (smooth) maximal atlas, i.e. an atlas that is not contained in any other atlases, is said to provide a (smooth)* ***differential structure*** *for M.*

If the compatibility among local charts is only of class $\mathscr{C}^r$, then we will talk about a $\mathscr{C}^r$ differential manifold. If the compatibility is analytic, in symbols $\mathscr{C}^\omega$, the manifold is called **real analytic**.

Convention: in this document we will only consider smooth manifolds, so we will omit to specify the adjective 'smooth' from now on, unless otherwise explicitly stated.

This choice is not so reductive after all, in fact, a celebrated theorem due to the great geometer Hassler Whitney [25] states that **every differential manifold of class $\mathscr{C}^1$ can always be endowed with a real-analytic maximal atlas and with $\mathscr{C}^r$ maximal atlases**, for all $r \geqslant 1$, which make it either a real-analytic or a $\mathscr{C}^r$ manifold (hence also a smooth manifold when $r = \infty$). Moreover, all the $\mathscr{C}^r$ differential structures are equivalent.

Thus, the really important gap to pass is from the $\mathscr{C}^0$-compatibility between local charts to the $\mathscr{C}^1$-compatibility, the more regular compatibility being assured to exist thanks to Whitney's theorem.

If the local model is $\mathbb{C}^n$ and not $\mathbb{R}^n$, then we will talk about a **complex manifold** of dimension n, in this case the transition functions are required to be holomorphic.

1.2 Examples of manifolds

Let us discuss some example of manifold:

1. **The trivial manifold**. $\mathbb{R}^n$ is a manifold with the **canonical single chart atlas** given by $(\mathbb{R}^n, id_{\mathbb{R}^n})$.

   To give an example of non-equivalent atlases, let us consider $\mathbb{R}$ and the atlas $(\mathbb{R}, \varphi)$, where $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(x) = \begin{cases} x & x \leqslant 0 \\ 2x & x > 0 \end{cases}$. This atlas is not compatible with the canonical atlas, in fact the transition function $\eta = \varphi \circ id_{\mathbb{R}}^{-1} = \varphi \circ id_{\mathbb{R}} = \varphi$ is continuous but not derivable in $x = 0$.

2. **Open submanifold**. Any open subset $U \subseteq \mathbb{R}^n$ is a manifold with single chart atlas given by (U, id_U).

3. **Product manifold**. If M and N are manifolds of dimension m and n, respectively, with atlases:
$$\mathscr{A} = \{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}, \quad \mathscr{B} = \{(V_\beta, \psi_\beta)\}_{\beta \in B},$$
respectively, then
$$\mathscr{A} \times \mathscr{B} := \{(U_\alpha \times V_\beta, \varphi_\alpha \times \psi_\beta)\}_{(\alpha,\beta) \in A \times B},$$
where $\varphi_\alpha \times \psi_\beta$ is the Cartesian product maps[6]
$$\begin{array}{rccl} \varphi_\alpha \times \psi_\beta : & U_\alpha \times V_\beta & \longrightarrow & \mathbb{R}^m \times \mathbb{R}^n \\ & (x,y) & \longmapsto & (\varphi_\alpha \times \psi_\beta)(x,y) = (\varphi_\alpha(x), \psi_\beta(y)), \end{array}$$
is an atlas that makes the Cartesian product $M \times N$ a manifold, called the **product manifold** of M and N. Since $\mathbb{R}^n \times \mathbb{R}^m \cong \mathbb{R}^{m+n}$, **the dimension of the product manifold is the sum of the factor manifolds**: $\dim(M \times N) = m + n$.

4. **Vector spaces of finite dimension as manifolds**. Let V be a real vector space of finite dimension n. Any norm on V determines a topology, which is known to be independent of the choice of the norm. With this topology, V is a topological manifold of dimension n. A natural differential structure on V can be defined thanks to the isomorphism between V and its prototype $\mathbb{R}^n$. More precisely, if $E = (e_1, \ldots, e_n)$ is any basis of V, then $I : V \to \mathbb{R}^n$, $v = v^i e_i \mapsto (v^i)_{i=1}^n$, is a linear isomorphism and also a homeomorphism in the topology induced by the norm. It follows that (V, I) is a *global chart* for V that can be used as a single-chart atlas.

   Any other basis $\tilde{E} = (\tilde{e}_1, \ldots, \tilde{e}_n)$ will induce a new global chart for V given by $(V, \tilde{I})$, where $\tilde{I} : V \to \mathbb{R}^n$, $v = \tilde{v}^i \tilde{e}_i \mapsto (\tilde{v}^i)_{i=1}^n$. To find the transition functions between these two charts, let us first recall that the change-of-basis matrix $A = (a_i^j)$, defined by $e_i = a_i^j \tilde{e}_j$, is invertible. From the equation
$$v = \tilde{v}^j \tilde{e}_j = v^i e_i = v^i a_i^j \tilde{e}_j, \qquad \forall v \in V,$$
we deduce that $\tilde{v}^j = a_i^j v^i$, i.e. the coordinates of any $v \in V$ w.r.t. the two charts, are related by an invertible linear transformation, which is obviously a diffeomorphism in $\mathbb{R}^n$. As a consequence, V is a smooth manifold.

   The differential structure defined in this way is called the *standard differential structure of the real vector space V*.

5. **The manifold of matrices**. The group of $m \times n$ matrices with real entries $M(m \times n, \mathbb{R})$ is known to be isomorphic with $\mathbb{R}^{mn}$ via the lexicographic order of the matrix elements (ordered by either rows or columns), thus it is a manifold of dimension mn. $M(m \times n, \mathbb{C})$ is a $2mn$ dimensional real manifold.

[6] We have used the Cartesian product map, defined as follows: given $f : D_f \to R_f$ and $g : D_g \to R_g$, D and R are used for domain and range, the Cartesian product function between f and g is:
$$\begin{array}{rccl} f \times g : & D_f \times D_g & \longrightarrow & R_f \times R_g \\ & (x,y) & \longmapsto & (f \times g)(x,y) := (f(x), g(x)). \end{array}$$

6. **The manifold of invertible matrices**. $\mathrm{GL}(n,\mathbb{R}) = \{A \in M(n,\mathbb{R}),\ \det(A) \neq 0\}$ is not only a subset of $M(n,\mathbb{R}) \cong \mathbb{R}^{n^2}$, but it is also *open* w.r.t. the topology of $\mathbb{R}^{n^2}$. In fact, $\mathrm{GL}(n,\mathbb{R}) = (\det^{-1}\{0\})^c$, i.e. it is the complementary set of the inverse image of 0 via the determinant function, being $\{0\}$ a closed set, $\det^{-1}\{0\}$ is closed because det is a continuous function, thus $\mathrm{GL}(n,\mathbb{R})$ is the complementary of a closed set, so it is an open set. As open subset of $M(n,\mathbb{R}) \cong \mathbb{R}^{n^2}$, **$\mathrm{GL}(n,\mathbb{R})$ is manifold of dimension** n^2. $\mathrm{GL}(n,\mathbb{C})$ is a $2n^2$ dimensional real manifold.

7. **The sphere as a manifold**. Proving that a spherical surface in $\mathbb{R}^{n+1}$, briefly a sphere, is a manifold is a classical and beautiful computation in differential geometry. Before considering the most general case, we start with the easiest one, i.e. that of the 1-dimensional sphere of radius 1, which has the advantage of showing us in a very clear geometrical way how to build an atlas. We will then extend this same construction to the n-dimensional case and to a generic radius $R > 0$.

   Let $S^1 := \{x \in \mathbb{R}^2\ :\ \|x\| = 1\}$, where $\|\ \|$ is the Euclidean norm, be the 1-dimensional unit sphere in $\mathbb{R}^2$, i.e. with radius equal to 1. We start by considering the following identification:
$$\pi := \{x \in \mathbb{R}^2,\ x = (x^1, 0)\} \cong \mathbb{R},$$
   then we define the north pole N, south pole S and a generic point p of the 1-dimensional sphere S^1 as follows:
$$\begin{cases} N = (0,1) = e_2 \ \text{(the second element of the canonical basis of } \mathbb{R}^2) \\ S = (0,-1) = -N \\ p = (p^1, p^2). \end{cases}$$
   Let us now consider $\mathcal{A} := \{(U_1, \varphi_N), (U_2, \varphi_S)\}$, where $U_1 := S^1 \backslash \{N\}$, $U_2 := S^1 \backslash \{S\}$, and
$$\begin{array}{rccl} \varphi_N : & U_1 & \longrightarrow & \pi \\ & (p^1,p^2) & \longmapsto & \varphi_N(p^1,p^2) := \frac{1}{1-p^2}\, p^1, \end{array} \qquad \begin{array}{rccl} \varphi_S : & U_2 & \longrightarrow & \pi \\ & (p^1,p^2) & \longmapsto & \varphi_S(p^1,p^2) := \frac{1}{1+p^2}\, p^1. \end{array}$$
   The functions φ_1 and φ_2 are called **stereographic projections** from the north and the south pole, respectively. Their geometrical meaning is represented in figure 1.1.

   The (unique) intersection between $\pi \cong \mathbb{R}$ and the straight line that connects $N = (0,1)$ with $p = (p^1, p^2)$ can be determined as follows: the Cartesian equation of this straight line is of course $y(p) = 1 + \frac{1-p^2}{0-p^1}(p-0)$, i.e. $y(p) = 1 - \frac{1-p^2}{p^1}p$, so the only value of $p^* \in \pi$ such that $y(p^*) = 0$ is $p^* = \frac{1}{1-p^2}p^1 = \varphi_N(p^1,p^2)$, thus the stereographic projection from the north pole is simply the point p^*. Analogous considerations can be done for the stereographic projection from the south pole, obtaining $\varphi_S(p^1,p^2) = \overline{p}$.

   We observe that the stereographic projection from N excludes from its domain N itself and maps the south pole to the origin of $\pi \cong \mathbb{R}$, in fact: $\varphi_N(S) = \varphi_1(0,-1) = \frac{1}{2}\,0 = 0$. The same considerations hold exchanging N with S and φ_N with φ_S. Of course $U_1 \cup U_2 = S^1$, so the covering property is verified by $\mathcal{A}$, we must check the compatibility. φ_N and φ_S are of course smooth and invertible on their respective domains, let us make the transition functions between them explicit in order to check if they are smooth.

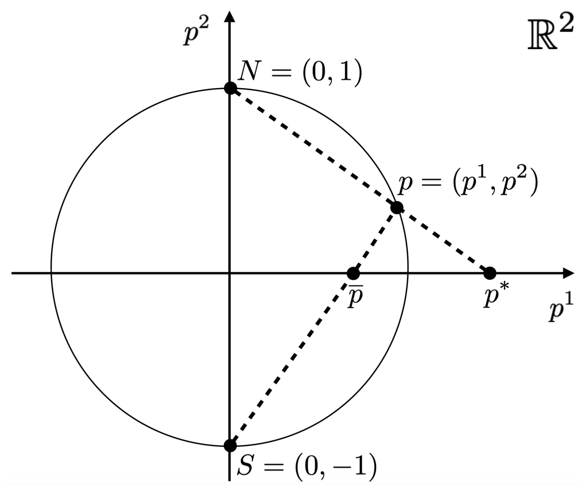


Figure 1.1: The stereographic projection from the north pole in 2D.

We start with φ_N: its inverse function is $\varphi_N^{-1} : \pi \to S^1\backslash\{N\}$, $x \mapsto \varphi_N^{-1}(x) = p = (p^1, p^2)$, with (p^1, p^2) such that $\varphi_N(p^1, p^2) = x$, i.e. $\frac{p^1}{1-p^2} = x$, or

$$p^1 = (1 - p^2)x. \tag{1.1}$$

If we manage to write $p^2(x)$, i.e. p^2 as a function of x, then p^1 also becomes a function of x, thus making φ_N^{-1} explicit. In order to do so, it is convenient to use the constraint that defines S^1, i.e. $\|p = (p^1, p^2)\| = 1 \iff \|(p^1, p^2)\|^2 = 1$, or:

$$(p^1)^2 + (p^2)^2 = 1 \iff (p^1)^2 = 1 - (p^2)^2 = (1 - p^2)(1 + p^2),$$

which, introduced in the square of eq. (1.1) gives:

$$\cancel{(1 - p^2)}(1 + p^2) = (1 - p^2)^{\cancel{2}} x^2 \iff 1 + p^2 = x^2 - x^2 p^2 \iff p^2(x) = \frac{x^2 - 1}{x^2 + 1}.$$

By introducing this expression in eq. (1.1) we obtain:

$$p^1(x) = (1 - p^2(x))x = \left(1 - \frac{x^2 - 1}{x^2 + 1}\right) x = \frac{2}{x^2 + 1} x.$$

So, we have expressed both p^1 and p^2 as a function of x and both expressions satisfy eq. (1.1), hence the explicit expression of φ_N^{-1} is:

$$\begin{array}{rccl} \varphi_N^{-1} : & \pi \cong \mathbb{R} & \longrightarrow & S^1\backslash\{N\} \\ & x & \longmapsto & \varphi_N^{-1}(x) = \big(p^1(x), p^2(x)\big) = \left(\frac{2}{x^2+1}x, \frac{x^2-1}{x^2+1}\right), \end{array}$$

analogously, we obtain:

$$\begin{array}{rccl} \varphi_S^{-1} : & \pi \cong \mathbb{R} & \longrightarrow & S^1\backslash\{S\} \\ & x & \longmapsto & \varphi_S^{-1}(x) = \left(\frac{2}{x^2+1}x, \frac{1-x^2}{x^2+1}\right). \end{array}$$

We can now compute the transition functions explicitly to test if they are smooth: first of all we notice that, since $\varphi_N(S) = \varphi_S(N) = 0$, on the intersection $U_{1,2} := U_1 \cap U_2 = S^1 \backslash \{N, S\}$ we have that $\varphi_N(U_{1,2}) = \varphi_S(U_{1,2}) = \mathbb{R}\backslash\{0\}$, so $\eta_{SN} := \varphi_S \circ \varphi_N^{-1} : \mathbb{R}\backslash\{0\} \to U_{1,2}$ and similarly for η_{NS}. By direct computation we have, for all $y \in \mathbb{R}\backslash\{0\}$,

$$\eta_{SN}(y) = \varphi_S(\varphi_N^{-1}(y)) = \varphi_S\left(\frac{2}{y^2+1}y, \frac{y^2-1}{y^2+1}\right) = \frac{\frac{2}{y^2+1}y}{1+\frac{y^2-1}{y^2+1}} = \frac{2y}{2y^2} = \frac{1}{y},$$

which is a smooth function on $\mathbb{R}\backslash\{0\}$, similarly:

$$\eta_{NS}(y) = \varphi_N(\varphi_S^{-1}(y)) = \varphi_N\left(\frac{2}{y^2+1}y, \frac{1-y^2}{y^2+1}\right) = \frac{\frac{2}{y^2+1}y}{1-\frac{1-y^2}{y^2+1}} = \frac{1}{y},$$

again, a smooth function on $\mathbb{R}\backslash\{0\}$. Thus, the transition functions between the charts defined by the stereographic projections are smooth, so $\mathcal{A}$ is an atlas for S^1, which acquires the status of **smooth manifold of dimension 1 with local model** $\mathbb{R}$.

Let us consider the general case. We call **sphere of radius** $R > 0$ the subset of $\mathbb{R}^{n+1}$ given by

$$\boxed{S^n_R = \{x \in \mathbb{R}^{n+1},\ \|x\| = R\}}\ , \tag{1.2}$$

where $\|\ \|$ is the Euclidean norm. If $R = 1$ we simply write S^n. The sphere S^n_R is a n-dimensional manifold for every $R > 0$. To prove it, let us build an atlas with two charts and show that the transition functions are smooth. As before, we use the stereographic projections of the generic point $p \in S^n_R$ from the north N and the south S pole:

$$\begin{cases} N = (0, \dots, 0, R) = Re_{n+1} \\ S = (0, \dots, 0, -R) = -N \\ p = (p^1, \dots, p^{n+1}) \end{cases}$$

onto the hyperplane

$$\pi := \{x \in \mathbb{R}^{n+1},\ x = (x^1, \dots, x^n, 0)\} \cong \mathbb{R}^n.$$

The first chart is: $(S^n_R\backslash\{N\}, \varphi_N)$, with

$$\begin{array}{rccl} \varphi_N : & S^n_R\backslash\{N\} & \longrightarrow & \pi \\ & p = (p^1, \dots, p^{n+1}) & \longmapsto & \varphi_N(p) = \frac{R}{R-p^{n+1}}(p^1, \dots, p^n). \end{array} \tag{1.3}$$

This time, to understand why the stereographic projection of p from the north pole N has this analytic form, instead of the Cartesian equation of the straight line connecting N to p, we consider (just to offer another possible view) its parametric equation, i.e. $x : \mathbb{R} \to \mathbb{R}^{n+1}$, $t \mapsto x(t) = N + t(p - N)$, notice that $x(0) = N$, $x(1) = p$. Since the coordinates of N are all zero unless the last one which is equal to R, the coordinates of $x(t)$ are

$$\begin{cases} x^1(t) = tp^1 \\ x^2(t) = tp^2 \\ \vdots \\ x^n(t) = tp^n \\ x^{n+1}(t) = R + t(p^{n+1} - R). \end{cases}$$

The point $\varphi_N(p) \in \pi$ is obtained by applying on the previous coordinates the constraint that defines π, i.e. by imposing $x^{n+1}(t) = 0$, or $x^{n+1} = R + t(p^{n+1} - R) = 0 \iff t = \frac{R}{R-p^{n+1}}$, so

$$\varphi_N(p) = (x^1(t), \ldots, x^n(t))\big|_{t=R/(R-p^{n+1})},$$

i.e. eq. (1.3).

Notice that $\varphi_N(N)$ is not defined[7] and that, if we take $p = S = (0, \ldots, 0, -R)$, then $p^i = 0$ for all $i = 1, \ldots, n$ and $p^{n+1} = -R$, so $\varphi_N(S) = (0, \ldots, 0)$, i.e. the stereographic projection from the north pole of the south pole is the origin of $\mathbb{R}^n$.

By the unicity of the intersection between the hyperplane π and the straight line passing through N and p, we have that φ_N is bijective.

The inverse of φ_N is defined as:

$$\begin{array}{rccc} \varphi_N^{-1}: & \pi & \longrightarrow & S^n_R \backslash \{N\} \\ & x = (x^1, \ldots, x^n) & \longmapsto & \varphi_N^{-1}(x) = p, \end{array}$$

where $x = \varphi_N(p)$, i.e. $(x^1, \ldots, x^n) = \frac{R}{R-p^{n+1}}(p^1, \ldots, p^n)$, thus

$$(p^1, \ldots, p^n) = \frac{R - p^{n+1}}{R}(x^1, \ldots, x^n), \tag{1.4}$$

which shows also for this general case that we just need to compute p^{n+1} as a function of x, i.e. $p^{n+1}(x^1, \ldots, x^n)$, to express also $p^1, \ldots, p^n$ as functions of $(x^1, \ldots, x^n)$ and thus finding the explicit expression of φ_N^{-1}.

As in the 1-dimensional case, we take advantage of the constraint that defines S^n_R, i.e. $p \in S^n_R$ if and only if $(p^1)^2 + \cdots + (p^n)^2 + (p^{n+1})^2 = R^2$, thus

$$(p^1)^2 + \cdots + (p^n)^2 = R^2 - (p^{n+1})^2 = (R - p^{n+1})(R + p^{n+1}). \tag{1.5}$$

If we compute the sum of the square components of both sides of eq. (1.4) we get:

$$(p^1)^2 + \cdots + (p^n)^2 = \frac{(R - p^{n+1})^2}{R^2}(x^1)^2 + \cdots + (x^n)^2 \underset{(x^{n+1}=0!)}{=} \frac{(R - p^{n+1})^2}{R^2}\|x\|^2,$$

but thanks to eq. (1.5),

$$\cancel{(R - p^{n+1})}(R + p^{n+1}) = \frac{(R - p^{n+1})^{\cancel{2}}}{R^2}\|x\|^2 \iff \frac{R + p^{n+1}}{R - p^{n+1}} = \frac{\|x\|^2}{R^2},$$

which, solved w.r.t. p^{n+1}, gives

$$p^{n+1}(x) = R\frac{\|x\|^2 - R^2}{\|x\|^2 + R^2}.$$

By inserting this expression for $p^{n+1}(x)$ in eq. (1.4) we get:

$$p^j(x) = 1 - \frac{\|x\|^2 - R^2}{\|x\|^2 + R^2}x^j = \frac{2R^2}{\|x\|^2 + R^2}x^j, \quad j = 1, \ldots, n,$$

[7]If φ_N would be defined on the whole sphere, it would create a homeomorphism between a compact set and a non compact one, $\mathbb{R}^n$, which is impossible. This observation leads to the conclusion that **it is not possible to have a single chart atlas for the sphere**, or any other compact manifold in $\mathbb{R}^n$.

thus

$$\varphi_N^{-1}(x^1,\dots,x^n) = \left(\frac{2R^2}{\|x\|^2+R^2}x^1,\dots,\frac{2R^2}{\|x\|^2+R^2}x^n, R\frac{\|x\|^2-R^2}{\|x\|^2+R^2}\right). \tag{1.6}$$

The stereographic projection from the south pole is built in the same way, we simply have to replace N with S, obtaining

$$\begin{array}{rrcl} \varphi_S : & S_R^n\backslash\{S\} & \longrightarrow & \pi \\ & p=(p^1,\dots,p^{n+1}) & \longmapsto & \varphi_S(p) = \frac{R}{R+p^{n+1}}(p^1,\dots,p^n). \end{array}$$

with $\varphi_S(N) = (0,\dots,0)$ and

$$\begin{array}{rrcl} \varphi_S^{-1} : & \pi & \longrightarrow & S_R^n\backslash\{S\} \\ & x=(x^1,\dots,x^n) & \longmapsto & \varphi_S^{-1}(x) = \frac{1}{\|x\|^2+R^2}(2R^2x^1,\dots,2R^2x^n, R(R^2-\|x\|^2)). \end{array}$$

Having at disposal the explicit expressions of φ_S, φ_N and their inverses, we can check the compatibility between them, i.e. that the transition functions are smooth on the intersection $S_R^n\backslash\{N,S\}$. Since $\varphi_N(S) = \varphi_S(N) = 0$, we have

$$\varphi_N(S_R^n\backslash\{N,S\}) = \mathbb{R}^n\backslash\{0\} = \varphi_S(S_R^n\backslash\{N,S\})$$

so $\eta_{SN} := \varphi_S\circ\varphi_N^{-1} : \mathbb{R}^n\backslash\{0\}\to\mathbb{R}^n\backslash\{0\}$, $y\mapsto\eta_{SN}(y)$. We have:

$$\eta_{SN}(y) = \frac{R}{R+R\frac{\|y\|^2-R^2}{\|y\|^2+R^2}}\frac{1}{\|y\|^2+R^2}(2R^2y^1,\dots,2R^2y^n) \iff \eta_{SN}(y) = \frac{R^2}{\|y\|^2}y,$$

which is smooth because $y\neq 0$ in the domain of η_{SN}. With analogous computations we have $\eta_{NS}(y) = \frac{R^2}{\|y\|^2}y$, smooth as well. This shows that $((U_N,\varphi_N),(U_S,\varphi_S))$ is an atlas for S_R^n, called **stereographic atlas** and that S_R^n **is a smooth manifold of dimension** n **with local model** $\mathbb{R}^n$.

8. **An alternative (but compatible) atlas on the sphere**. There are other atlases, compatible with the stereographic atlas that can be built on the sphere. For the sake of clarity, let us consider S^1 to show an alternative (very redundant) atlas that can be proven by direct computation to be compatible with the stereographic atlas. This is the atlas: $\mathcal{B} = \{(U_i,\varphi_i),\ i=1,\dots,4\}$, where:

$$\begin{cases} U_1 = \{(p^1,p^2)\in S^1 \ :\ p^1>0\}, & \varphi_1(p^1,p^2) := p^2 \\ U_2 = \{(p^1,p^2)\in S^1 \ :\ p^2>0\}, & \varphi_2(p^1,p^2) := p^1 \\ U_3 = \{(p^1,p^2)\in S^1 \ :\ p^1<0\}, & \varphi_3(p^1,p^2) := p^2 \\ U_4 = \{(p^1,p^2)\in S^1 \ :\ p^2<0\}, & \varphi_4(p^1,p^2) := p^1. \end{cases}$$

9. **The n-torus**. Thanks to example 3. we can build the product manifold:

$$\mathbb{T}^n = S^1\times\cdots\times S^1$$

which is a compact manifold.

10. **The real projective manifold**. It is defined as follows:

$$\boxed{\mathbb{RP}^n := \mathbb{R}^{n+1}\backslash\{0\}/\sim},$$

where

$$\forall x, y \in \mathbb{R}^{n+1}\backslash\{0\},\ x \sim y \iff \exists\lambda \in \mathbb{R}\backslash\{0\}\ :\ y = \lambda x,$$

i.e. with $\sim$ we **identify any two non-zero vectors in $\mathbb{R}^{n+1}$ which are multiples of each other by a non zero real coefficient**: $(x^0, \dots, x^n) = (\lambda x^0, \dots, \lambda x^n)$.

So, the elements of the projective manifold will be equivalence classes of vectors in $\mathbb{R}^{n+1}\backslash\{0\}$ that lie on the same straight line passing through the origin[8].

Endowed with the quotient topology, $\mathbb{RP}^n$ is a topological manifold, we will prove that $\mathbb{RP}^n$ is also a differential manifold of dimension n and this will provide a first example of manifold that is not made up by a subset of points in $\mathbb{R}^d$, $d \geqslant 1$, as the elements of $\mathbb{RP}^n$ can be identified with straight lines in $\mathbb{R}^{n+1}$ and not points of $\mathbb{R}^{n+1}$!

A typical notation used when dealing with the projective manifold is the following:

$$(x^0 : \cdots : x^n) := (\lambda x^0 : \cdots : \lambda x^n) \qquad \forall\lambda \neq 0,$$

$(x^0 : \cdots : x^n)$ are called **homogeneous coordinates** of an element in $\mathbb{RP}^n$.

Let us construct an atlas with compatible charts for $\mathbb{RP}^n$ by considering the following open domains:

$$U_i := \{(x^0 : \cdots : x^n) \in \mathbb{RP}^n\ :\ x^i \neq 0\},$$

i.e. the i-th homogeneous coordinate of the elements belonging to U_i is non null (the others can be null or not, but the i-th surely not). There are $n+1$ such domains and they trivially cover $\mathbb{RP}^n$, i.e. $\mathbb{RP}^n = \bigcup_{i=0}^{n} U_i$, in fact, having removed 0 from $\mathbb{R}^{n+1}$, at least one homogeneous coordinate of an arbitrary element of $\mathbb{RP}^n$ must be different from 0, but then it belongs to a suitable U_i.

The chart maps on U_i are defined as follows:

$$\begin{array}{cccl} \varphi_i : & U_i & \xrightarrow{\sim} & \mathbb{R}^n \\ & (x^0 : \cdots : x^n) & \longmapsto & \varphi_i(x^0 : \cdots : x^n) := \left(\frac{x^0}{x^i}, \dots, \frac{x^{i-1}}{x^i}, \frac{x^{i+1}}{x^i}, \dots, \frac{x^n}{x^i}\right), \end{array}$$

analytically well defined because in U_i, $x_i \neq 0$. Notice that we only have n components in the image of φ_i because the i-th component gives $\frac{x^i}{x^i} = 1$, which is a fixed value that we remove from the image. φ_i does not depend on the particular representative in the equivalence class where $(x^0 : \cdots : x^n)$ belongs, in fact:

$$U_i \ni (\lambda x^0 : \cdots : \lambda x^i : \cdots : \lambda x^n) \overset{\varphi_i}{\mapsto} \left(\frac{\lambda x^0}{\lambda x^i}, \dots, \frac{\lambda x^{i-1}}{\lambda x^i}, \frac{\lambda x^{i+1}}{\lambda x^i}, \dots, \frac{\lambda x^n}{\lambda x^i}\right) = \varphi_i(x^0 : \cdots : x^n),$$

so that $\varphi_i(\lambda x^0 : \cdots : \lambda x^n) = \varphi_i(x^0 : \cdots : x^n)\ \forall\lambda \neq 0$.

[8] Actually, since we have eliminated 0 from $\mathbb{R}^{n+1}$, the vectors belong to two opposite half lines with origin in 0, but, of course, these half lines identify in a unique way a straight lines passing through the origin of $\mathbb{R}^{n+1}$. This identification will be implicitly assumed in the main text.

φ_i is invertible, its inverse being the map that restores the value 1 after the i-th position starting from the position 1 of the index:

$$\begin{array}{cccl} \varphi_i^{-1}: & \mathbb{R}^n & \xrightarrow{\sim} & U_i \\ & (y^1, \ldots, y^n) & \longmapsto & \varphi_i(y^1, \ldots, y^n) = (y^1 : \cdots : y^i : 1 : y^{i+1} : \cdots : y^n). \end{array}$$

In fact,

$$\varphi_i^{-1}(\varphi_i(x^0 : \cdots : x^n)) = \varphi_i^{-1}\left(\frac{x^0}{x^i}, \ldots, \frac{x^{i-1}}{x^i}, \frac{x^{i+1}}{x^i}, \ldots, \frac{x^n}{x^i}\right) = \left(\frac{x^0}{x^i}, \ldots, \frac{x^{i-1}}{x^i}, 1, \frac{x^{i+1}}{x^i}, \ldots, \frac{x^n}{x^i}\right),$$

where, since in the last expression we start from the index 0, the value 1 must be restored after the $(i-1)$-th position. By definition of homogeneous coordinates we have $\left(\frac{x^0}{x^i}, \ldots, \frac{x^{i-1}}{x^i}, 1, \frac{x^{i+1}}{x^i}, \ldots, \frac{x^n}{x^i}\right) = (x^0 : \cdots : x^{i-1} : x^i : x^{i+1} : \ldots x^n)$, so $\varphi_i^{-1} \circ \varphi_i = id_{U_i}$ and, by an analogous computation, we have $\varphi_i \circ \varphi_i^{-1} = id_{\mathbb{R}^n}$.

$\{(U_i, \varphi_i),\ i = 0, \ldots, n\}$ is a $(n+1)$-charts atlas for the projective manifold if we can show that these charts are compatible on the intersections of their domains. For that, notice that, when $i \neq j$, the condition $U_i \cap U_j \neq \varnothing$ implies, by definition of the sets U_i and U_j, that the i-th and the j-th homogeneous coordinates of the elements of $\mathbb{RP}^n$ belonging to $U_i \cap U_j$ are both $\neq 0$. If $i < j$, the transition functions can be written as follows:

$$\begin{aligned} \eta_{ij} &= \varphi_i \circ \varphi_j^{-1}(y^1, \ldots, y^n) = \varphi_i(y^1 : \cdots : y^j : 1 : y^{j+1} : \cdots : y^n) \\ &= \left(\frac{y^1}{y^i}, \ldots, \frac{y^{i-1}}{y^i}, \frac{y^{i+1}}{y^i}, \ldots, \frac{y^j}{y^i}, \frac{1}{y^i}, \frac{y^{j+1}}{y^i}, \ldots, \frac{y^n}{y^i}\right), \end{aligned}$$

if $j < i$, we simply exchange i with j in the previous expression. Notice the gap between the $(i-1)$-th and the $(i+1)$-th coordinate, which guarantees the correct number of components. η_{ij} is evidently smooth because y^i and y^j are non null. Since $\eta_{ji} = \varphi_j^{-1} \circ \varphi_i$, we get exactly the same functional expression with inverted indices, thus also η_{ji} is smooth. So, $\mathbb{RP}^n$ **is a differential manifold of dimension** n.

11. **Grassmannian manifolds**. We have seen that $\mathbb{RP}^n$ can be identified with the set of vector subspaces of order 1 (the straight lines passing through the origin) of $\mathbb{R}^{n+1}$. More generally, if V is a real n-dimensional vector space, we define:

$$\boxed{\mathrm{Gr}_k(V) := \{W \ :\ W \text{ is a vector subspace of dimension } k \text{ of } V\}}\ .$$

It can be proven that $\mathrm{Gr}_k(V)$ is a differential manifold of dimension $k(n-k)$, called the Grassmannian manifold of order k of V. It is clear that:

$$\boxed{\mathbb{RP}^n = \mathrm{Gr}_1(\mathbb{R}^{n+1})}\ .$$

12. $\mathbb{RP}^n$ **as a suitable quotient of the sphere** S^n. Consider a vector $x \in \mathbb{R}^{n+1}\backslash\{0\}$, then x and $\frac{x}{\|x\|}$, $\|\ \|$ being the Euclidean norm on $\mathbb{R}^{n+1}$, define the same element of $\mathbb{RP}^n$. However, $\frac{x}{\|x\|}$ belongs to the sphere $S^n = \{x \in \mathbb{R}^{n+1},\ \|x\| = 1\}$, this very simple observation shows that we can always see $\mathbb{RP}^n$ as a subset of S^n and that the map $\pi : S^n \to \mathbb{RP}^n$, $x = (x^0, \ldots, x^n) \mapsto (x^0 : \cdots : x^n)$ is surjective.

Notice however that π is not injective, because $\pi(x) = \pi(-x)$ for all $x \in S^n$, in fact $-\frac{x}{\|x\|}$ belongs to the same equivalence class as x and $\frac{x}{\|x\|}$ in the projective manifold! x and $-x$ are called **antipodal points**. To remove the lack of injectivity, it is sufficient to identify the antipodal points on the sphere S^n, i.e. to operate the quotient $S^n/\sim$, $x \sim -x$ for all $x \in S^n$. It is not difficult to prove that $S^n/\sim$ **endowed with the quotient topology, is isomorphic, as a differential manifold, to** $\mathbb{RP}^n$.

This example shows **how much manifold can be modified by a quotient**: in this case, we pass from a spherical surface, to a set of straight lines passing through the origin!

1.2.1 Manifolds from the level-set theorem in $\mathbb{R}^{n+m}$

Noticeable examples of manifolds embedded in a Euclidean space of suitable dimension can be built thanks to the so-called level-set theorem, which is a remarkable consequence of the inverse mapping theorem.

Let us consider $f : \Omega \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ open, $f \in \mathscr{C}^1(\Omega)$.

Def. 1.2.1 *$x \in \Omega \subset \mathbb{R}^n$ is a **critical point** of f if the total derivative $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ is not onto, i.e. if $\operatorname{rank}(Df(x)) < m$. A **critical value** of f is the image via f of a critical point x of f, so $f(x) \in \mathbb{R}^m$. We denote with $\operatorname{Crit}(f) \subset \Omega$ the set of critical points of f. A **regular value** of f is an element in $f(\Omega)$ that is not critical for f.*

It is easy to see that $\operatorname{Crit}(f)$ is a closed subset of Ω.

We recall that the rank of a matrix A coincides with the maximal dimension of the square sub-matrices M which have non-zero determinant. Hence, a critical point $x_0 \in \Omega$ for f is such that all the sub-matrices of the Jacobian matrix $Jf(x_0)$ with non-zero determinant have dimension strictly less than m.

The following result gives a (not necessary) sufficient condition for a set to be a manifold.

Theorem 1.2.1 (Level set theorem in $\mathbb{R}^{n+m}$) *Let:*

- *$\Omega \subseteq \mathbb{R}^{n+m}$ open set*
- *$f : \Omega \to \mathbb{R}^m$, $f \in \mathscr{C}^\infty(\Omega)$*
- *$a \in f(\Omega)$.*

Then, the set

$$M_a = f^{-1}(a) \backslash \operatorname{Crit}(f),$$

i.e. the a-level set of f without the critical points, is a smooth manifold of dimension n (the difference between the dimension of the domain and the codomain of f), w.r.t. the differential structure inherited by $\mathbb{R}^{n+m}$.

Of course, if f does not have critical points, then $M_a = f^{-1}(a)$.

Proof. The proof is very technical and consists in building explicitly an atlas of compatible local charts on M_a.

Let $x_0 \in M_a$, then, since p_0 is not a critical point, $Df(x_0)$ is surjective and so the Jacobian matrix $Jf(x_0)$ has maximal rank, equal to m. This means that is exists a square $m \times m$ sub-matrix of $Jf(x_0)$ with determinant different than 0. Modulo a permutation of the indices, we can always suppose that this sub-matrix is given by the last m columns, i.e.

$$B := \begin{pmatrix} \frac{\partial f^1}{\partial x^{n+1}} & \cdots & \frac{\partial f^1}{\partial x^{n+m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x^{n+1}} & \cdots & \frac{\partial f^m}{\partial x^{n+m}} \end{pmatrix}(x_0), \quad \det(B) \neq 0.$$

Let us now consider the function $g : \Omega \to \mathbb{R}^{n+m}$ defined by:

$$g(x) = g(x^1, \ldots, x^n, f(x)). \tag{1.7}$$

Its Jacobian matrix in x_0 is:

$$Jg(x_0) = \left(\begin{array}{ccc|c} 1 & \ldots & 0 & \\ \vdots & \ddots & \vdots & 0 \\ 0 & \ldots & 1 & \\ \hline & \star & & B \end{array}\right),$$

in fact the partial derivatives of the first block are $\frac{\partial x^i}{\partial x^j}$, those of the block on the right are $\frac{\partial x^i}{\partial x^{n+j}}$ and those of the block $\star$ are $\frac{\partial f^i}{\partial x^j}(x_0)$.

It follows that $\det(Jg(x_0)) = \det(B) \neq 0$, thus we are allowed to apply the inverse function theorem to say that g is a local diffeomorphism: there exists a neighborhood $\tilde{U} \subset \Omega \backslash \mathrm{Crit}(f)$ of x_0 and a neighborhood $W \subset \mathbb{R}^{n+m}$ of $g(x_0)$ such that $g|_{\tilde{U}} : \tilde{U} \to W$ is a diffeomorphism.

Let now $h : W \to \tilde{U}$ be the inverse of $g|_{\tilde{U}}$ and denote with $(h^1, \ldots, h^{n+m})$ its component functions, then, for all $y \in W$, it holds that

$$y = (y^1, \ldots, y^n, y^{n+1}, \ldots, y^{n+m}) = g(h(y)) \underset{(1.7)}{=} (h^1(y), \ldots, h^n(y), f(h(y))), \tag{1.8}$$

so $h^i(y) = y^i$, for all $i = 1, \ldots, n$, which implies

$$\begin{aligned} f(h(y)) &= f(h^1(y), \ldots, h^n(y), h^{n+1}(y), \ldots, h^{n+m}(y)) = f(y^1, \ldots, y^n, h^{n+1}(y), \ldots, h^{n+m}(y)) \\ &\underset{(1.8)}{=} (y^{n+1}, \ldots, y^{n+m}). \end{aligned} \tag{1.9}$$

The set $V := \{x \in \mathbb{R}^n \ : \ (x, a) \in W\}$, i.e. the intersection between W and the hyperplane $z = a$, is open in $\mathbb{R}^n$ because W is open in $\mathbb{R}^{n+m}$.

On V we define the following function: $\psi : V \to \mathbb{R}^{n+m}$,

$$\psi(x) := (x, h^{n+1}(x, a), \ldots, h^{n+m}(x, a)),$$

ψ is smooth, because it either acts as the identity, or as h, but h is the inverse of $g|_{\tilde{U}}$ and the inverse mapping theorem assures that h has the same regularity as g, which is smooth being defined as the identity on the first n components and as f on the last ones, and f is smooth by hypothesis.

Eq. (1.9), i.e. the equality

$$f(y^1, \dots, y^n, h^{n+1}(y), \dots, h^{n+m}(y)) = (y^{n+1}, \dots, y^{n+m}), \quad \forall y \in W,$$

implies that $\psi(V) = f^{-1}(a) \cap \tilde{U} = M_a \cap \tilde{U} =: U$.

Hence we have:

$$\mathbb{R}^n \supset V \xrightarrow[\psi]{\sim} U \subset M_a \subset \mathbb{R}^{n+m}$$

which implies that the inverse of ψ, i.e. $\varphi := \psi^{-1}$, is a local chart on M_a defined in the open neighborhood U of x_0.

We remark that, since ψ acts as the identity on the first n components of its image,

$$\varphi(x^1, \dots, x^{n+m}) = (x^1, \dots, x^n),$$

i.e. φ acts as the projection on the first n coordinates[9]. Of course, the projection is smooth, so also φ is a smooth function.

Thanks to this observation, the compatibility among local charts is trivial: in fact, given any two charts (U, φ), (U', φ'), the transition function $\varphi' \circ \varphi^{-1} = \varphi' \circ \psi$ has as coordinates either x^i or $h^j(x, a)$, which are smooth.

□

Thanks to this theorem we can prove quite easily that the most important matrix groups are differential manifolds.

- We start by showing how easy it is to prove that the sphere $S^n_R = \{x \in \mathbb{R}^{n+1} \ : \ \|x\|^2 = R^2\}$ is a manifold of dimension n thanks to the level set theorem in comparison to the construction of the stereographic atlas. In fact, it is enough to consider the function that associates to each vector of $\mathbb{R}^{n+1}$ its squared Euclidean norm:

  $$\begin{array}{rccl} f: & \mathbb{R}^{n+1} & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & f(x) = \|x\|^2 = (x^1)^2 + \cdots + (x^{n+1})^2, \end{array}$$

  which is smooth and, being a scalar function, its total derivative can be identified with its gradient of f, see the appendix IV.

  In this special case, the total derivative of f calculated in a point $x_0 \in \mathbb{R}^{n+1}$ is not onto only when $\nabla f(x_0) = \mathbf{0}$. The components of the gradient are the partial derivatives and $\frac{\partial f}{\partial x^i}(x_0) = 2x_0^i$, $i = 1, \dots, n+1$, this shows that $x_0 = \mathbf{0}$ is the only critical point of f and the only critical value of f is 0. Thus, for all $R > 0$, the level set

  $$f^{-1}(R^2) = \{x \in \mathbb{R}^{n+1} \ : \ f(x) = (x^1)^2 + \cdots + (x^{n+1})^2 = R^2\} \equiv S^n_R$$

  is a smooth manifold of dimension $(n+1) - 1 = n$.

[9] The fact that exactly the first n components are involved is a consequence of the permutation of indices supposed at the beginning of the proof.

- $\mathbf{SL}(n, \mathbb{R})$ **as a manifold of dimension** $n^2 - 1$. The function to be considered here is the determinant of a $n \times n$ matrix with real entries:

$$\begin{array}{rccl} \det: & M(n,\mathbb{R}) \cong \mathbb{R}^{n^2} & \longrightarrow & \mathbb{R} \\ & A & \longmapsto & \det(A). \end{array}$$

If $A = (a^i_j) \in M(n, \mathbb{R})$, then, by Laplace's formula, $\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a^i_j \det(A^i_j)$, where $A^i_j \in M(n-1, \mathbb{R})$ is the submatrix of A obtained by eliminating the i-th row and the j-th column. Being a polynomial function, det is smooth. Again, being a scalar function, its total derivative can be identified with its gradient $\nabla \det(A)$, whose n^2-components are given by

$$\frac{\partial \det}{\partial a^i_j}(A) = (-1)^{i+j} \det(A^i_j).$$

The matrix A is a critical points of det if and only if $\nabla \det(A) = \mathbf{0} \iff \det(A^i_j) = 0$ for all $i, j = 1, \ldots, n$. This situation can happen only if A has rank strictly less than $n - 1$, so:

$$\operatorname{Crit}(\det) = \{A \in M(n, \mathbb{R}) \ : \ \operatorname{rank}(A) \leqslant n - 2\}.$$

Any $A \in \operatorname{Crit}(\det)$ has null determinant, thus the only critical value for det is 0. Since $\mathrm{SL}(n, \mathbb{R}) = \{A \in M(n, \mathbb{R}) \ : \ \det(A) = 1\} = \det^{-1}\{1\}$, and 1 is a regular value for det, it follows that $\mathbf{SL}(n, \mathbb{R})$ **is a smooth manifold of dimension** $n^2 - 1$.

As a consequence of this result, $\mathbf{SL}(n, \mathbb{C})$ **is a (real) manifold of dimension** $2n^2 - 2$.

An alternative proof consists in observing that, thanks to equation (B.14), the matrices of $SL(n, \mathbb{R})$ are not critical point for the determinant.

- With similar, but more sophisticated, techniques based on the rank theorem, it can be proven that:

  - $\mathbf{O}(n)$ **and** $\mathbf{SO}(n)$ **are manifolds of dimension** $\frac{n(n-1)}{2}$;

  - $\mathbf{U}(n)$ **and** $\mathbf{SU}(n)$ **are (real) manifolds of dimension** n^2.

  We will show how to prove that $\mathrm{O}(n)$ is a manifold through the rank theorem in section 2.9.2 after discussing the concept of differential of functions between manifolds.

1.3 Morphisms and diffeomorphims between manifolds

Manifolds are the arena of differential geometry, let us now analyze their morphisms, i.e. the transformations between manifolds that respect their properties regarding the differential structure. Smooth functions between manifolds are the morphisms of the category of smooth manifolds, while diffeomorphisms are its isomorphisms.

As usual, smoothness is defined through the use of local charts and compatibility among intersecting charts must be required.

Def. 1.3.1 *Given two manifolds M and N of dimensions m and n, respectively, and a function*

$$\begin{array}{rccl} f: & M & \longrightarrow & N \\ & p & \longmapsto & f(p) = q, \end{array}$$

*two local charts $(U_\alpha, \varphi_\alpha)$ in M and (V_β, ψ_β) in N are said to be **f-related** if $f(U_\alpha) \subseteq V_\beta$. Two atlases $\mathcal{A}$ and $\mathcal{B}$ of M and N, respectively, are f-related if every chart of one atlas is f-related with at least one chart of the other atlas.*

The following result shows that the continuity of f is sufficient to guarantee the existence of related atlases.

Theorem 1.3.1 *Given two manifolds M and N and a **continuous** function $f : M \to N$, it exists a couple of f-related atlases of M and N.*

Proof. The proof is constructive. Given any two atlases $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}_{\alpha \in I}$ and $\mathcal{B} = \{(V_\beta, \psi_\beta)\}_{\beta \in J}$ of M and N, respectively, a direct way to build an atlas $\tilde{\mathcal{A}}$ equivalent to $\mathcal{A}$ and f-related to $\mathcal{B}$ is to define $\tilde{\mathcal{A}} := \{(\tilde{U}_{\alpha\beta}, \tilde{\varphi}_{\alpha\beta})\}_{\alpha \in I,\, \beta \in J}$, with:

$$\begin{cases} \tilde{U}_{\alpha\beta} := U_\alpha \cap f^{-1}(V_\beta) \\ \tilde{\varphi}_{\alpha\beta} := \varphi_\alpha|_{\tilde{U}_{\alpha\beta}} . \end{cases}$$

In fact, thanks to the continuity of f, $f^{-1}(V_\beta)$ is an open subset of M and so $U_\alpha \cap f^{-1}(V_\beta)$ is an open subset included in (or coincident with) U_α. The charts $\tilde{\varphi}_{\alpha\beta}$ are compatible with the charts φ_α because the operation of restriction preserves the smoothness of the transition functions, thus the atlases $\mathcal{A}$ and $\tilde{\mathcal{A}}$ are equivalent.

Moreover, $f(U_\alpha \cap f^{-1}(V_\beta)) \subseteq f(U_\alpha) \cap V_\beta \subseteq V_\beta$ thanks to well-known relationships between functions and sets, which guarantees that the atlases $\tilde{\mathcal{A}}$ and $\mathcal{B}$ are f-related. □

We can now define the important concept of local representation (or expression) of a function between manifolds.

Def. 1.3.2 (Local representation of a function between manifolds) *The local representation of $f : M \to N$ w.r.t. the f-related local charts $(U_\alpha, \varphi_\alpha)$ and (V_β, ψ_β) is the function:*

$$\boxed{f_{\beta\alpha} := \psi_\beta \circ f|_{U_\alpha} \circ \varphi_\alpha^{-1}} ,$$

$$\begin{array}{rccl} f_{\beta\alpha}: & \varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^m & \longrightarrow & \psi_\beta(f(U_\alpha)) \subseteq \mathbb{R}^n \\ & x \equiv (x^i)_{i=1}^m & \longmapsto & f_{\beta\alpha}(x) = y \equiv (y^j)_{j=1}^n . \end{array}$$

The following commutative diagram visualizes the local representation of a function.

$$\begin{tikzcd}
M \supseteq U_\alpha \arrow[r, "f|_{U_\alpha}"] & f(U_\alpha) \subseteq N \arrow[d, "\psi_\beta"] \\
\varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^m \arrow[u, "{\varphi_\alpha}^{-1}"] \arrow[r, dashed, "f_{\beta\alpha}"'] & \psi_\beta(f(U_\alpha)) \subseteq \mathbb{R}^n
\end{tikzcd}$$

$f_{\beta\alpha}$ is a function between open subsets of finite-dimensional real Euclidean spaces, thus we perfectly know what it means for such a function to be smooth. Its smoothness is used to define that of the function f itself.

Def. 1.3.3 (Smooth function between manifolds) *$f : M \to N$ is smooth if it exists a couple of f-related charts, $(U_\alpha, \varphi_\alpha)$ of M and (V_β, ψ_β) of N, such that $f_{\beta\alpha}$, the local representation of f w.r.t. these charts, is smooth.*

Notation: the symbol $\mathscr{C}^\infty(M, N)$ denotes the set of all smooth functions between M and N. If $N \equiv \mathbb{R}$ we simply write $\mathscr{C}^\infty(M)$.

As in standard differential calculus, smoothness implies continuity.

Theorem 1.3.2 *If $f : M \to N$ is smooth, then f is also continuous.*

Proof. Almost immediate: if $f : M \to N$ is smooth in any point $p \in M$ then, by definition of smoothness, it exists a couple of charts $(U_\alpha, \varphi_\alpha)$ and (V_β, ψ_β) such that $p \in U_\alpha$, $f(U_\alpha) \subseteq V_\beta$ and $f_{\beta\alpha} = \psi_\beta \circ f|_{U_\alpha} \circ \varphi_\alpha^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ is smooth, and thus continuous, because it is a map between real Euclidean spaces, where we know that smoothness implies continuity. But then, since φ_α and ψ_β are homeomorphisms, $\psi_\beta^{-1} \circ f_{\beta\alpha} \circ \varphi_\alpha$ is continuous too, as composition of continuous maps, but:

$$\psi_\beta^{-1} \circ \psi_\beta \circ f|_{U_\alpha} \circ \varphi_\alpha^{-1} \circ \varphi_\alpha = f|_{U_\alpha},$$

i.e. $f|_{U_\alpha}$ is continuous in an open neighborhood of any point $p \in M$, hence it is continuous on the whole manifold M that, we recall, is a topological manifold, so it intrinsically carries the notion of continuity w.r.t. its topology. □

The *definition of smoothness* just given is intrinsic, i.e. it does not depend on the f-related local charts considered: *once it is true for one couple of f-related local charts, it holds for all f-related local charts.*

To check this, fix any local chart (V_β, ψ_β) of N and consider two f-related overlapping local charts of M, $(U_\alpha, \varphi_\alpha)$ and $(U_{\alpha'}, \varphi_{\alpha'})$, i.e. $U_\alpha \cap U_{\alpha'} = U_{\alpha\alpha'} \neq \varnothing$ and $f(U_{\alpha\alpha'}) \subseteq V_\beta$. The chart maps are related by smooth transition functions $\eta_{\alpha'\alpha} = \varphi_{\alpha'} \circ \varphi_\alpha^{-1}$, thus $\varphi_\alpha^{-1} = \varphi_{\alpha'}^{-1} \circ \eta_{\alpha'\alpha}$. Hence, the local representations $f_{\beta\alpha} = \psi_\beta \circ f|_{U_{\alpha\alpha'}} \circ \varphi_\alpha^{-1}$ and $f_{\beta\alpha'} = \psi_\beta \circ f|_{U_{\alpha\alpha'}} \circ \varphi_{\alpha'}^{-1}$ satisfy:

$$f_{\beta\alpha} = \psi_\beta \circ f|_{U_{\alpha\alpha'}} \circ \varphi_{\alpha'}^{-1} \circ \eta_{\alpha'\alpha} = f_{\beta\alpha'} \circ \eta_{\alpha'\alpha},$$

which implies, thanks to the smoothness of $\eta_{\alpha'\alpha}$, that $f_{\beta\alpha}$ is smooth if and only if $f_{\beta\alpha'}$ is. The f-related couples of local charts considered, $((U_\alpha, \varphi_\alpha), (V_\beta, \psi_\beta))$ and $((U_{\alpha'}, \varphi_{\alpha'}), (V_\beta, \psi_\beta))$, are arbitrary, thus it is enough to check the smoothness of the local representation of f w.r.t. one couple of local maps to guarantee the validity of this property w.r.t. every other couple.

By composing $f_{\beta\alpha}$ with the functionals ε^j of the dual basis of $\mathbb{R}^n$, we get the real-valued functions:

$$\begin{array}{rrcl} f^j_{\beta\alpha} \equiv \varepsilon^j \circ f_{\beta\alpha} : & \varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^m & \longrightarrow & \mathbb{R} \\ & x = (x^i) & \longmapsto & f^j_{\alpha\beta}(x) = y^j, \end{array}$$

which, as always, are nothing but the scalar components of the $\mathbb{R}^n$-valued function $f_{\beta\alpha}$.

The functions $f^j_{\beta\alpha}$, $j = 1, \dots, n$, represent **the local coordinate transformation functions between the local coordinates (x^i) of a point $p \in M$ and the local coordinates $(y^j) = (f^j_{\beta\alpha}(x^i))$ of the point $q = f(p) \in N$.**

With the usual abuse of notation, we write $f^j_{\beta\alpha} \equiv y^j$, so that:

$$y^j = y^j(x^i), \qquad i = 1, \dots, m, \; j = 1, \dots, n.$$

The following diagram shows the action of the local coordinate transformation functions.

$$\begin{tikzcd}
U_\alpha \arrow[r, "f|_{U_\alpha}"] & f(U_\alpha) \arrow[d, "\psi_\beta"] \\
x \in \varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^m \arrow[u, "{\varphi_\alpha}^{-1}"] \arrow[r, dashed, "f_{\beta\alpha}"'] \arrow[dr, dashed, "y^j(x^i)"'] & \psi_\beta(f(U_\alpha)) \subseteq \mathbb{R}^n \arrow[d, "\varepsilon^j"] \\
& y^j
\end{tikzcd}$$

Since the functionals ε^j are smooth, it follows that a function $f : M \to N$ **is smooth if and only if we can pass smoothly from a local coordinate description of a point $p \in M$ to a local coordinate description of the transformed point** $q = f(p) \in N$.

A special case is provided by functions for which $N = \mathbb{R}^n$, or an open subset of $\mathbb{R}^n$ (thus, in particular, for **scalar functions on** M when $n = 1$). In this case, the differential structure is provided by the canonical global atlas $(\mathbb{R}^n, id_{\mathbb{R}^n})$, so the composition with ψ_β is not necessary anymore and the **local representation of** $f : M \to \mathbb{R}^n$, is just $f_\alpha = f|_{U_\alpha} \circ \varphi_\alpha^{-1}$, that will be denote simply as

$$\boxed{f_\alpha = f \circ {\varphi_\alpha}^{-1}}\,. \tag{1.10}$$

The following commutative diagrams resume our considerations.

$$\begin{tikzcd}
U_\alpha \arrow[r, "f|_{U_\alpha}"] & f(U_\alpha) \subseteq \mathbb{R}^n \arrow[d, "id_{f(U_\alpha)}"] & & U_\alpha \arrow[r, "f"] & \mathbb{R}^n \\
\varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^m \arrow[u, "{\varphi_\alpha}^{-1}"] \arrow[r, dashed, "f_\alpha"'] & f(U_\alpha) \subseteq \mathbb{R}^n & \underset{\text{(simplified as)}}{\Longleftrightarrow} & \mathbb{R}^m \arrow[u, "{\varphi_\alpha}^{-1}"] \arrow[ur, dashed, "f_\alpha"'] &
\end{tikzcd}$$

In the intersection of two charts $(U_\alpha, \varphi_\alpha)$, (U_β, φ_β) it holds that:

$$f_\alpha = f_\beta \circ \eta_{\beta\alpha}, \quad f_\beta = f_\alpha \circ \eta_{\alpha\beta},$$

as shown by the following diagram for the first formula, the second being analogous.

$$\begin{tikzcd}
 & U_\alpha \cap U_\beta \arrow[d, bend right, "\varphi_\beta"'] \arrow[dr, bend left, "f"] & \\
\mathbb{R}^m \arrow[ur, bend left, "{\varphi_\alpha}^{-1}"] \arrow[r, dashed, "\eta_{\beta\alpha}"] \arrow[rr, dashed, bend right, "f_\alpha"'] & \mathbb{R}^m \arrow[u, bend right, "{\varphi_\beta}^{-1}"'] \arrow[r, dashed, "f_\beta"] & \mathbb{R}^n
\end{tikzcd}$$

Another special case is provided by functions for which $M = \mathbb{R}^m$ or an open subset of $\mathbb{R}^m$, thus, in particular, for **curves in** N when $m = 1$, as recalled in the following definition.

Def. 1.3.4 (Path, or curve, in a manifold passing through a point) *The smooth function*[10] $\gamma : (-\varepsilon, \varepsilon) \subseteq \mathbb{R} \to M$, $\varepsilon > 0$, *is said to be a path, or curve, in M passing through the point $p \in M$ if $\gamma(0) = p$.*

In this case, the differential structure is provided by the canonical global atlas $(\mathbb{R}^m, id_{\mathbb{R}^m})$, so the composition with φ_α^{-1} is not necessary anymore and the **local representation of** $f : U \subseteq \mathbb{R}^m \to N$, such that $f(U) \subseteq V_\beta$ is just $f_\beta = \psi_\beta \circ f|_U$, that will be denote simply as

$$\boxed{f_\beta = \psi_\beta \circ f}\ . \tag{1.11}$$

$$\begin{tikzcd}
U \subseteq \mathbb{R}^m \arrow[r, "f|_U"] & f(U) \subseteq V_\beta \subseteq N \arrow[d, "\psi_\beta"] \\
U \subseteq \mathbb{R}^m \arrow[u, "id_U"] \arrow[r, dashed, "f_\beta"'] & \psi_\beta(f(U)) \subseteq \mathbb{R}^n
\end{tikzcd}
\quad \underset{\text{(simplified as)}}{\Longleftrightarrow} \quad
\begin{tikzcd}
U \arrow[r, "f"] \arrow[dr, dashed, "f_\beta"'] & N \arrow[d, "\psi_\beta"] \\
 & \mathbb{R}^n
\end{tikzcd}$$

To resume, the local representations of the previous special cases of functions between manifolds are:

$$\begin{cases} f_\alpha = f \circ \varphi_\alpha^{-1} & \forall f : M \to \mathbb{R}^n \\ f_\beta = \psi_\beta \circ f & \forall f : \mathbb{R}^m \to N. \end{cases}$$

We are now ready to define the concept of diffeomorphism.

Def. 1.3.5 (Global and local diffeomorphism) *$f : M \to N$ is a diffeomorphism if it is a smooth bijective function with smooth inverse $f^{-1} : N \to M$, in this case M and N are said to be **diffeomorphic manifolds**.*

$f : M \to N$ is a local diffeomorphism if there exists an open subset $U \subset M$ such that $f(U)$ is open in N and $f|_U : U \to f(U)$ is a diffeomorphism.

[10] $(-\varepsilon, \varepsilon)$ is to be considered as an open submanifold of $\mathbb{R}$.

The most **basic example of local diffeomorphism is easily provided by any chart map** $\varphi_\alpha : U_\alpha \subseteq M \to \varphi(U_\alpha) \subseteq \mathbb{R}^n$ of a manifold M of dimension n. By definition, φ_α is bijective, thus, the only property that we must check to verify that φ_α is a local diffeomorphism is its smoothness and that of its inverse $\varphi_\alpha^{-1} : \varphi(U_\alpha) \subseteq \mathbb{R}^n \to U_\alpha \subseteq M$.

It is clear that, in both cases, we can use formulae (1.10) and (1.11) to compute the local representations of φ_α and φ_α^{-1}, respectively. As the diagram below shows, **the local representation of a chart map and its inverse is provided by the identity function** $id_{\varphi_\alpha(U_\alpha)}$, which is of course smooth.

$$\begin{tikzcd} U_\alpha \arrow[r, "\varphi_\alpha"] & \varphi_\alpha(U_\alpha) \\ \varphi_\alpha(U_\alpha) \arrow[u, "\varphi_\alpha^{-1}"] \arrow[ur, dashed, "id_{\varphi_\alpha(U_\alpha)}"'] & \end{tikzcd} \qquad \begin{tikzcd} \varphi_\alpha(U_\alpha) \arrow[r, "\varphi_\alpha^{-1}"] \arrow[dr, dashed, "id_{\varphi_\alpha(U_\alpha)}"'] & U_\alpha \arrow[d, "\varphi_\alpha"] \\ & \varphi_\alpha(U_\alpha) \end{tikzcd}$$

Thus, each local chart map allows us to **diffeomorphically identify any open chart domain of M with an open subset of $\mathbb{R}^n$**. Moreover, **the transition functions $\eta_{\beta\alpha}$ are local diffeomorphisms**, being composition of chart maps and their inverses.

We end this section by underlying the difference between identical and diffeomorphic manifolds.

Def. 1.3.6 (Identical manifolds) *Let M be a topological manifold and $(M, \mathcal{A}_1)$, $(M, \mathcal{A}_2)$ two manifolds over M with their corresponding maximal atlases. Then, $(M, \mathcal{A}_1)$ and $(M, \mathcal{A}_2)$ are said to be identical, as manifolds, if $id_M : (M, \mathcal{A}_1) \to (M, \mathcal{A}_2)$ is a diffeomorphism w.r.t. the differential structures associated to $\mathcal{A}_1$ and $\mathcal{A}_2$.*

From the point of view of manifold classification, diffeomorphic manifolds are considered as equivalent. However, as the following example shows, in the same diffeomorphic class of manifolds, we can find manifolds that are not identical.

Example of diffeomorphic non-identical manifolds. We consider:

$$\begin{cases} M_1 = (\mathbb{R}, \varphi = id_\mathbb{R}) \\ M_2 = (\mathbb{R}, \psi), \ \psi(x) = x^3 \ \forall x \in \mathbb{R}. \end{cases}$$

To check if $id_\mathbb{R}$ is a diffeomorphism w.r.t. these two monochart atlases, we have to consider, as always, the local representation:

$$\begin{tikzcd} \mathbb{R} \arrow[r, "f=id_\mathbb{R}"] \arrow[d, "\varphi=id_\mathbb{R}"'] & \mathbb{R} \arrow[d, "\psi"] \\ \mathbb{R} \arrow[r, dashed, "\tilde{f}"'] & \mathbb{R} \end{tikzcd}$$

While $\tilde{f} : \mathbb{R} \to \mathbb{R}$, $\tilde{f}(x) = (\psi \circ id_\mathbb{R} \circ \varphi^{-1})(x) = \psi(x) = x^3$ is smooth, its inverse $\tilde{f}^{-1} : \mathbb{R} \to \mathbb{R}$, $\tilde{f}^{-1}(y) = (\varphi \circ id_\mathbb{R} \circ \psi^{-1})(y) = \sqrt[3]{y}$ is not, because $(\tilde{f}^{-1})'(y) = 1/(3\sqrt[3]{y^2})$, which is not differentiable in $y = 0$.

Thus, M_1 and M_2 are not identical manifolds. However, they are diffeomorphic to each other, a simple diffeomorphism being $F : M_1 \to M_2$, $x \mapsto F(x) = \sqrt[3]{x}$. To check it, let us analyze again the local representation:

$$\begin{tikzcd} \mathbb{R} \arrow[r, "F"] \arrow[d, "\varphi = id_{\mathbb{R}}"'] & \mathbb{R} \arrow[d, "\psi"] \\ \mathbb{R} \arrow[r, dashed, "\tilde{F}"'] & \mathbb{R} \end{tikzcd}$$

Of course, $\tilde{F}(x) = (\sqrt[3]{x})^3 = x$ and $(\tilde{F})^{-1}(y) = (\sqrt[3]{y})^3 = y$, both smooth.

More generally,

- $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^n$ is not a diffeomorphism for all $n \geqslant 1$, so polynomial functions on $\mathbb{R}$ are not diffeomorphisms because their inverse functions lack of smoothness.

- $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^{1/n}$ is a diffeomorphism for all $n \in \mathbb{N}$ odd.

We list next some general interesting facts about differential structures:

- Any connected manifold M of dimension 1 is diffeomorphic to either S^1 or to $\mathbb{R}$. In particular, if M is compact (as a topological manifold), then it is diffeomorphic to S^1, otherwise it is diffeomorphic to $\mathbb{R}$.

- Every topological manifold of dimension $\leqslant 3$ admits a *unique differential structure up to diffeomorphisms.*

- For every topological manifold of dimension > 3 there exist compact topological manifolds that does not admit differentiable atlases.

- $\mathbb{R}^n$ admits a *unique differential structure up to diffeomorphisms for all* $n \neq 4$.

- Donaldson-Freedman's 1984 result: $\mathbb{R}^4$ admits *infinite non-countable non-diffeomorphic smooth structures.*

- S^7 has exactly 28 *non-diffeomorphic smooth structures* that can be explicitly written.

1.3.1 Introduction to Lie groups

We now have all the information that we need to introduce the hugely important concept of Lie group, that will be extensively treated later in this document.

We have seen that $M = \mathrm{GL}(n, \mathbb{R})$ is a smooth manifold of dimension n^2, as open subset of $M(n, \mathbb{R}^2) \cong \mathbb{R}^{n^2}$. We also know that $M \times M$ is a product manifold of dimension $2n^2$.

The matrix product function is:

$$\begin{array}{rccl} f: & M \times M & \longrightarrow & M \\ & (A, B) & \longmapsto & f(A, B) = C := A \cdot B, \end{array}$$

where $C = (c^i_j)_{i,j=1,\dots,n}$, with $c^i_j = a^i_h b^h_j$. The components of f are polynomial functions, hence they are smooth and so is f.

The inverse matrix function is:

$$\begin{array}{rccl} g: & M & \longrightarrow & M \\ & A & \longmapsto & g(A) = A^{-1} = \frac{A^*}{\det(A)}, \end{array}$$

where A^* is the *adjugate matrix* of A, i.e. the transpose of its cofactor matrix, defined by $C(A) = \Big((-1)^{i+j}\det(A^i_j)\Big)_{j=1,\dots,n}$, where A^i_j is, as we have already seen, the submatrix of A obtained by eliminating the i-th row and the j-th column. All the operations contained in A^* are smooth, plus the division by the determinant of A is smooth, so g is a smooth function.

Thus, the fundamental group operations, product and inversion, of M are smooth. Every group which has these properties is called a Lie group, as defined below.

Def. 1.3.7 (Lie group) *A topological group*[11] *G endowed with a differential structure that makes it a manifold and such that the product $G \times G \to G$, $(a,b) \mapsto a \cdot b$ and the inversion $G \to G$, $g \mapsto g^{-1}$ are smooth is called a Lie group. The dimension of a Lie group is its dimension as manifold.*

$\mathbb{R}^d$, considered as a group w.r.t. the operation of sum is a Lie group for all $d \geqslant 1$ and, thus, so is $M(n,\mathbb{R})$. Other examples of Lie groups are given by the so-called classical matrix Lie groups, which are listed below.

Classical real matrix groups

- $\mathrm{GL}(n,\mathbb{R}) = \{g \in M(n,\mathbb{R}) \ : \ \det(g) \neq 0\}$ (**general linear group**)
- $\mathrm{SL}(n,\mathbb{R}) = \{g \in \mathrm{GL}(n,\mathbb{R}) \ : \ \det(g) = 1\}$ (**special linear group**)
- $\mathrm{O}(n) = \{g \in \mathrm{GL}(n,\mathbb{R}) \ : \ \forall x,y \in \mathbb{R}^n, \ \langle gx, gy\rangle = \langle x,y\rangle\} = \{g \in \mathrm{GL}(n,\mathbb{R}) \ : \ g^t = g^{-1}\}$ (**orthogonal group**[12], it is the group of all the **isometries of** $\mathbb{R}^n$)
- $\mathrm{SO}(n) = \{g \in \mathrm{O}(n) \ : \ \det(g) = 1\}$ (**special orthogonal group**)

Classical complex matrix groups

- $\mathrm{GL}(n,\mathbb{C}) = \{g \in M(n,\mathbb{C}) \ : \ \det(g) \neq 0\}$ (**general linear complex group**)
- $\mathrm{SL}(n,\mathbb{C}) = \{g \in \mathrm{GL}(n,\mathbb{C}) \ : \ \det(g) = 1\}$ (**special linear complex group**)
- $\mathrm{U}(n) = \{g \in \mathrm{GL}(n,\mathbb{C}) \ : \ \forall x,y \in \mathbb{C}^n, \ \langle gx, gy\rangle = \langle x,y\rangle\} = \{g \in \mathrm{GL}(n,\mathbb{C}) \ : \ g^\dagger = g^{-1}\}$ (**unitary group**[13] it is the group of all the **isometries of** $\mathbb{C}^n$)
- $\mathrm{SU}(n) = \{g \in \mathrm{U}(n) \ : \ \det(g) = 1\}$ (**special unitary group**)

[11] i.e. a group G that is also a topological space such that the product and the inversion maps are continuous.
[12] In this definition $\langle \ , \ \rangle$ is the Euclidean product of $\mathbb{R}^n$.
[13] In this definition $\langle \ , \ \rangle$ is the Euclidean product of $\mathbb{C}^n$ and $g^\dagger = \overline{g}^t$ is the adjoint matrix of g.

1.3.2 S^1, SO(2) and U(1) as isomorphic mono-dimensional Lie groups

We can easily show that the unit sphere S^1, the groups SO(2) and $U(1)$ are isomorphic Lie groups by using the isomorphism between $\mathbb{R}^2$ and $\mathbb{C}$:

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\sim} & \mathbb{C} \\ (a,b) & \longmapsto & z = a + ib. \end{array} \tag{1.12}$$

In fact,

$$S^1 = \{(a,b) \in \mathbb{R}^2 \ : \ a^2 + b^2 = 1\} \subset \mathbb{R}^2,$$

is the unit circle in $\mathbb{R}^2$, and $\mathrm{U}(1) = \{z \in \mathbb{C} \ : \ \forall x, y \in \mathbb{C},\ \langle zx, zy\rangle = \langle x, y\rangle\}$, but thanks to the sequilinearity of the complex scalar product, $\langle zx, zy\rangle = |z|^2\langle x, y\rangle = \langle x, y\rangle$ if and only if $|z|^2 = 1$, i.e. $|z| = 1$, thus:

$$\mathrm{U}(1) = \{z \in \mathbb{C} \ : \ |z| = 1\} \subset \mathbb{C},$$

can be identified with the multiplicative group of complex numbers with unit modulus: if $|z_1| = |z_2| = 1$, then $|z_1 z_2| = 1$ and $|z^{-1}| = 1$ whenever $|z| = 1$, thus the multiplicative group structure of $U(1)$ is evident. Since $|z| = 1 \iff |z|^2 = a^2 + b^2 = 1$, it is clear that if we restrict the isomorphism (1.12) to S^1, we obtain the following isomorphism:

$$\begin{array}{ccc} \mathbb{R}^2 \supset S^1 & \xrightarrow{\sim} & \mathrm{U}(1) \subset \mathbb{C} \\ (a,b) & \longmapsto & z = a + ib. \end{array}$$

Thanks to this identification, S^1 inherits the group structure from U(1) and, vice-versa, U(1) inherits a manifold structure from S^1. It can be proven that the manifold and group structures are compatible, in the sense of definition 1.3.7, so S^1 and U(1) are Lie groups. Since the dimension of S^1 is 1, S^1 **and** $\mathbf{U}(1)$ **are mono-dimensional compact Lie groups**.

We can push the isomorphism even further by considering the group SO(2). We recall that the matrices of this group can be characterized very easily. In fact, given any 2×2 real matrix with unit determinant A:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad A^t = \begin{pmatrix} a & c \\ b & d \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

we have that $A^t = A^{-1} \iff a = d$ and $c = -b$, i.e. we can rewrite SO(2) as follows:

$$\mathrm{SO}(2) = \left\{ A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},\ \det(A) = a^2 + b^2 = 1 \right\},$$

but then the correspondence

$$\begin{array}{ccc} \mathbb{R}^2 \supset S^1 & \xrightarrow{\sim} & \mathrm{SO}(2) \subset \mathrm{SL}(2,\mathbb{R}) \\ (a,b) & \longmapsto & \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \end{array}$$

is an isomorphism. Moreover, for all $\vartheta \in [0, 2\pi)$, if we set $a = \cos\vartheta$ and $b = \sin\vartheta$, then $a^2 + b^2 = 1$, so we can explicitly characterize the matrices of SO(2) as follows:

$$\mathrm{SO}(2) = \left\{ \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} \ : \ \vartheta \in [0, 2\pi) \right\}.$$

As a consequence, we have **three isomorphic mono-dimensional Lie groups**, S^1, $\mathrm{U}(1)$ and $\mathrm{SO}(2)$ that can be explicitly characterized by one free parameter $\vartheta \in [0, 2\pi)$ as follows:

$$S^1 = \{(a,b) \in \mathbb{R}^2 \ : \ a = \cos\vartheta, \ b = \sin\vartheta, \ \vartheta \in [0,2\pi)\} \subset \mathbb{R}^2$$

$$\mathrm{U}(1) = \{z \in \mathbb{C} \ : \ z = \cos\vartheta + i\sin\vartheta, \ \vartheta \in [0,2\pi)\} \subset \mathbb{C}$$

$$\mathrm{SO}(2) = \left\{ A \in \mathrm{SL}(2,\mathbb{R}) \ : \ A = \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}, \ \vartheta \in [0,2\pi) \right\} \subset \mathrm{SL}(2,\mathbb{R}).$$

We now examine S^3 without considering S^2, in fact it can be proven that S^2 **is not a Lie group**. The isometries that we have discussed in this section follow from the natural identification between $\mathbb{R}^2$ and $\mathbb{C}$, those that we need to analyze S^3 follow from the natural identification between $\mathbb{R}^4$ and the non-Abelian division algebra (thus also a group) of **quaternions**. For this reason, before going through the analysis of S^3, we provide the basic information about the quaternions that we will use.

1.3.3 Interlude: rudiments about quaternions

The natural isomorphism between $\mathbb{R}^4$ and the non-Abelian division algebra of **quaternions** $\mathbb{H}$ is the following:

$$\begin{array}{ccl} \mathbb{R}^4 & \xrightarrow{\sim} & \mathbb{H} \\ (a,b,c,d) & \longmapsto & q = a + ib + jc + kd, \end{array} \tag{1.13}$$

where $i^2 = j^2 = k^2 = ijk = -1$ and the multiplication of the quaternionic units i, j, k follows this diagram:

$$\begin{tikzcd} & i \arrow[dr, bend left] & \\ k \arrow[ur, bend left] & & j \arrow[ll, bend left] \end{tikzcd}$$

if we multiply the quaternionic units in the sense of the arrows, we get as result the next quaternionic unit multiplied by $+1$, if we multiply the quaternionic units following the opposite sense w.r.t. the arrows, we obtain the next quaternionic unit multiplied by -1. For example, $ij = k$, $ji = -k$, $ik = -j$, $jk = i$, and so on.

It is useful to define the **scalar part** and the **vector part** of $q = a + ib + jc + kd$ as follows: $s_q := a$ and $\mathbf{v}_q := ib + jc + kd$. $q = s_q + \mathbf{v}_q \in \mathbb{H}$ is said to be a **pure quaternion** if $s_q = 0$, $\mathbf{v}_q \neq 0$, so[14] $q = \mathbf{v}_q$. The set of pure quaternions is denoted by

$$\mathbb{H}_0 := \{ib + jc + kd, \ b, c, d \in \mathbb{R}\} = \{\mathbf{v}_q, \ q \in \mathbb{H}\} \subset \mathbb{H}.$$

The **conjugate quaternion** of $q = a + ib + jc + kd$ is

$$\overline{q} := a - ib - jc - kd = s_q - \mathbf{v}_q = -\frac{1}{2}(q + \ iqi + \ jqj + \ kqk),$$

and it holds that $\overline{pq} = \overline{q}\,\overline{p}$ for all $p, q \in \mathbb{H}$. In terms of conjugate quaternions, the scalar and vector part of a quaternion can be written as follows:

$$s_q = \frac{1}{2}(q + \overline{q}), \ \mathbf{v}_q = \frac{1}{2}(q - \overline{q})$$

[14]this explains why some authors instead of talking about pure quaternions use the name **vector quaternion**.

The **modulus**, or **norm**, of a quaternion, or its conjugate, is the non negative real number $|q|$ such that:

$$|q|^2 = |\overline{q}|^2 := q\overline{q} = a^2 + b^2 + c^2 + d^2 = s_q^2 + |\mathbf{v}_q|^2,$$

which implies that

$$|q|^2 = |\mathbf{v}_q|^2, \qquad \forall q \in \mathbb{H}_0.$$

So, $(\mathbb{H}, |\ |)$ and $(\mathbb{H}_0, |\ |)$ can be identified, as a metric spaces, with a copy of $\mathbb{R}^4$ and $\mathbb{R}^3$, respectively, equipped with the Euclidean norm.

Another consequence of the previous formula is that every non-zero quaternion has an inverse given by

$$q^{-1} = \frac{\overline{q}}{|q|^2}, \tag{1.14}$$

taking the modulus at both sides and recalling that $|\overline{q}| = |q|$ we find

$$|q^{-1}| = \frac{|q|}{|q|^2} = |q|^{-1}.$$

The quaternion modulus is a multiplicative norm, i.e.

$$|pq| = \sqrt{(pq)(\overline{pq})} = \sqrt{pq\overline{q}\,\overline{p}} = \sqrt{p|q|^2\overline{p}} = \sqrt{|p|^2|q|^2} = |p|\,|q|, \quad \forall p, q \in \mathbb{H}.$$

With direct but tedious computations, it can be verified that the product of two quaternions $p = p_0 + ip_1 + jp_2 + kp_3$ and $q = q_0 + iq_1 + jq_2 + kq_3$ can be written as follows

$$\boxed{pq = (s_q + \mathbf{v}_p)(s_q + \mathbf{v}_q) = (s_p s_q - \langle \mathbf{v}_p, \mathbf{v}_q \rangle) + (s_p \mathbf{v}_q + s_q \mathbf{v}_p + \mathbf{v}_p \times \mathbf{v}_q)}, \tag{1.15}$$

where $\langle \mathbf{v}_p, \mathbf{v}_q \rangle = \sum_{\ell=1}^{3} p_\ell q_\ell$, the non-Abelian part of the quaternion product is encoded by

$$\mathbf{v}_p \times \mathbf{v}_q = \det \begin{pmatrix} i & j & k \\ p_1 & p_2 & p_3 \\ q_1 & q_2 & q_3 \end{pmatrix} = i(p_2q_3 - q_3p_2) - j(p_1q_3 - p_3q_1) + k(p_1q_2 - p_2q_1),$$

in fact, $pq - qp = \mathbf{v}_p \times \mathbf{v}_q - \mathbf{v}_q \times \mathbf{v}_p = \mathbf{v}_p \times \mathbf{v}_q + \mathbf{v}_p \times \mathbf{v}_q$, thanks to the anti-symmetry of the cross product, so:

$$pq - qp = 2\mathbf{v}_p \times \mathbf{v}_q, \tag{1.16}$$

which shows that **the quaternionic product is commutative if and only if** $\mathbf{v}_p$ **and** $\mathbf{v}_q$ **are collinear**, i.e. $\mathbf{v}_p = \alpha \mathbf{v}_q$, $\alpha \in \mathbb{R}$. A special case of this situation is when $p = q$, in this case eq. (1.15) gives:

$$\boxed{q^2 = (s_q^2 - |\mathbf{v}_q|^2) + (2s_q\mathbf{v}_q)}. \tag{1.17}$$

Eq. (1.15) implies that **the scalar part of the quaternion product is commutative**

$$s_{pq} = s_{qp}, \quad \forall p, q \in \mathbb{H}, \tag{1.18}$$

and, if at least one between p and q is a pure quaternion, then

$$pq + qp = -2\langle \mathbf{v}_p, \mathbf{v}_q \rangle, \quad p \text{ and/or } q \in \mathbb{H}_0. \tag{1.19}$$

Let us now introduce a paramount important map.

Def. 1.3.8 (Quaternionic Conjugation) *Given $q \in \mathbb{H}$, the q-conjugation is defined as follows:*

$$\begin{array}{rccl} C_q : & \mathbb{H} & \longrightarrow & \mathbb{H} \\ & p & \longmapsto & C_q(p) := qpq^{-1}. \end{array}$$

C_q is clearly linear and, moreover, by the multiplicative property of the quaternionic modulus, we have $|C_q(p)| = |qpq^{-1}| = |q||p||q^{-1}| = |q||p||q|^{-1} = |p|$, for all $p \in \mathbb{H}$. Recalling that $(\mathbb{H}, |\ |)$ can be identified with a copy of $\mathbb{R}^4$ endowed with the Euclidean norm, we find that **the $q-$conjugation C_q is an orthogonal transformation on** $\mathbb{R}^4$ for all $q \in \mathbb{H}$.

From eq. (1.18) it follows that any conjugation of a pure quaternion is pure, in fact if $s_p = 0$, then $s_{qpq^{-1}} = s_{q^{-1}qp} = s_p = 0$. Hence, $C_q \in \mathrm{End}(\mathbb{H}_0)$ for all $q \in \mathbb{H}$.

The set of **unit quaternions**, i.e. quaternions with unit modulus, is denoted by

$$\mathbb{H}_1 := \{q = a + ib + jc + kd \in \mathbb{H} \ : \ |q| = 1 \iff a^2 + b^2 + c^2 + d^2 = 1\} \subset \mathbb{H}.$$

From eq. (1.14), we have that, for unit quaternion, the inverse coincides with the conjugate:

$$q^{-1} = \overline{q}, \qquad \forall q \in \mathbb{H}_1.$$

Contrary to $\mathbb{C}$, where there are only two square roots of -1, i.e. i and $-i$, **there are infinitely many square roots of -1 in $\mathbb{H}$**: they can be identified with S^2, **the unit sphere in** $\mathbb{R}^3$, which coincides with the **pure unit quaternions**, as we prove below.

Proposition 1.3.1 *Let $q \in \mathbb{H}$, then $q^2 = -1$ if and only if $q \in (\mathbb{H}_0 \cap \mathbb{H}_1) = S^2$.*

Proof.

$\boxed{\Longleftarrow}$: let $q = a + ib + jc + kd = s_q + \mathbf{v}_q \in \mathbb{H}_0 \cap \mathbb{H}_1$, then $s_q = 0$ and $|q|^2 = |\mathbf{v}_q|^2 = 1$, so, by eq. (1.17), $q^2 = -|\mathbf{v}_q|^2 = -1$.

$\boxed{\Longrightarrow}$: let now $q \in \mathbb{H}$ satisfy $q^2 = -1$, then, again by eq. (1.17), we have:

$$\begin{cases} s_q^2 - |\mathbf{v}_q|^2 = a^2 - b^2 - c^2 - d^2 = -1 \\ 2ab = 0 \\ 2ac = 0 \\ 2ad = 0. \end{cases}$$

If a were non null, then the last three equations would imply $b = c = d = 0$ and the first equation would give $a^2 = -1$, which cannot be true because $a \in \mathbb{R}$. Hence $a = s_q = 0$, so $q \in \mathbb{H}_0$ and, by the first equation of the system above we get $|q| = |\mathbf{v}_q| = 1$, so $q \in \mathbb{H}_1$, i.e. $q \in \mathbb{H}_0 \cap \mathbb{H}_1$. □

1.3.4 S^3, $\mathbb{H}_1$ and $\mathrm{SU}(2)$ as isomorphic Lie groups of dimension 3

By recalling that the sphere S^3 is defined as:

$$S^3 = \{(a, b, c, d) \in \mathbb{R}^4 \ : \ a^2 + b^2 + c^2 + d^2 = 1\} \subset \mathbb{R}^4,$$

it is clear that if we restrict the identification defined by (1.13) to $S^3 \subset \mathbb{R}^4$ we get a natural identification between S^3 and $\mathbb{H}_1$:

$$\begin{array}{rcl} \mathbb{R}^4 \supset S^3 & \xrightarrow{\sim} & \mathbb{H}_1 \subset \mathbb{H} \\ (a,b,c,d) & \longmapsto & q = a + ib + jc + kd. \end{array}$$

Thanks to this isomorphism, S^3 inherits the group structure from $\mathbb{H}_1$ and, vice-versa, $\mathbb{H}_1$ inherits a manifold structure from S^3. As for the case of S^1 and $\mathrm{U}(1)$, it can be proven that the manifold and group structures are compatible, thus making S^3 and $\mathbb{H}_1$ Lie groups. Since the dimension of S^3 is 3, S^3 and $\mathbb{H}_1$ are Lie groups of dimension 3.

As before, we can find a further isomorphism with a matrix group: $\mathrm{SU}(2)$. In order to formalize this, we need to introduce the **Pauli matrices**:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.20}$$

For $\ell = 1,2,3$, the matrices σ_ℓ are complex, Hermitian ($\overline{\sigma_\ell}^t = \sigma_\ell$) and unitary ($\overline{\sigma_\ell}^t = \sigma_\ell^{-1}$), so it also holds that $\sigma_\ell = \sigma_\ell^{-1}$ and $\sigma_\ell^2 = I_2$. Actually, the set $(I_2, \sigma_1, \sigma_2, \sigma_3)$ is a **basis** for the *real* vector space of 2×2 Hermitian matrices.

By direct computation, we get that:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = I_2, \quad \sigma_1\sigma_2 = i\sigma_3, \quad \sigma_2\sigma_1 = -i\sigma_3, \ldots$$

These properties are reminiscent of those of the quaternionic units, they perfectly agree with them if we multiply the Pauli matrices by i, for in that case we get:

$$\tilde{\sigma}_1 = i\sigma_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \tilde{\sigma}_2 = i\sigma_2 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \tilde{\sigma}_3 = i\sigma_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

and

$$\tilde{\sigma}_1^2 = \tilde{\sigma}_2^2 = \tilde{\sigma}_3^2 = -I_2, \quad \tilde{\sigma}_1\tilde{\sigma}_2 = -\tilde{\sigma}_3, \quad \tilde{\sigma}_2\tilde{\sigma}_1 = \tilde{\sigma}_3, \quad \tilde{\sigma}_2\tilde{\sigma}_3 = -\tilde{\sigma}_1, \ldots$$

By comparison with the quaternions, we can establish these correspondences:

$$\begin{cases} 1 \leftrightarrow I_2 \\ i \leftrightarrow \tilde{\sigma}_3 \\ j \leftrightarrow \tilde{\sigma}_2 \\ k \leftrightarrow \tilde{\sigma}_1. \end{cases}$$

This allows us to represent the quaternions via matrices, in fact:

$$q = 1 \cdot a + i \cdot b + j \cdot c + k \cdot d \iff q = I_2 a + \tilde{\sigma}_3 b + \tilde{\sigma}_2 c + \tilde{\sigma}_1 d = \begin{pmatrix} a + ib & c + id \\ -c + id & a - ib \end{pmatrix} =: A_q.$$

Moreover, by direct computation, we have:

$$\det(A_q) = a^2 + b^2 + c^2 + d^2 = |q|^2, \quad \frac{1}{2}\mathrm{Tr}(A_q) = s_q. \tag{1.21}$$

We notice that A_q is a matrix of the type:

$$M = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix},$$

with $\det(M) = |\alpha|^2 + |\beta|^2$ and $\overline{M}^t M = \det(M) I_2$, so:

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix},\ |\alpha|^2 + |\beta|^2 = 1 \right\} = \left\{ \begin{pmatrix} a+ib & c+id \\ -c+id & a-ib \end{pmatrix},\ a^2+b^2+c^2+d^2 = 1 \right\}.$$

From eq. (1.21) we get the (group) isomorphism

$$\mathbb{H}_1 \cong \mathrm{SU}(2).$$

The matrices $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3)$ are **anti-Hermitian**, i.e. $\overline{\tilde{\sigma}_\ell}^t = -\tilde{\sigma}_\ell$, $\ell = 1, 2, 3$. We will show that they constitute a basis for the Lie algebra of SU(2): $\mathfrak{su}(2) \cong T_e\,\mathrm{SU}(2)$, the tangent space to SU(2) at e, the unit element of the group. Since SU(2) has dimension 3 as a manifold, its tangent space has dimension 3 as well and so does its Lie algebra.

1.4 Covering and universal covering

The concept of covering (or cover) is very important in differential geometry, in particular in Lie group theory.

The definition of covering can be puzzling at first sight, thus we prefer to discuss a very simple example that will serve as a motivation for the definition.

Consider $S^1_R \subset \mathbb{R}^2$, $R > 0$, and $\mathbb{R}$, then the map

$$\begin{array}{rccl} \pi: & \mathbb{R} & \longrightarrow & S^1_R \\ & t & \longmapsto & \pi(t) = (R\cos t, R\sin t) \end{array}$$

is smooth and surjective, i.e. via π we can *cover* smoothly the whole manifold S^1_R. However, π is not injective, thus, if we consider any open subset $U \subset S^1_R$, the counter-image $\pi^{-1}(U)$ will be composed by infinitely many open subsets of $\mathbb{R}$. For example, to fix the ideas, consider the open arc A of the circle of radius R which goes from $(R, 0)$ to $(0, R)$, then $\pi^{-1}(A)$ is the following union of disjoint open intervals in $\mathbb{R}$:

$$\pi^{-1}(A) = \bigcup_{k\in\mathbb{Z}} (2k\pi, \pi/2 + 2k\pi).$$

For a fixed value $k \in \mathbb{Z}$, the interval $I_k = (2k\pi, \pi/2 + 2k\pi)$ is a connected set in $\mathbb{R}$ and the restriction of π on I_k is a diffeomorphism between I_k and A.

These considerations motivate the definition of *covering as a smooth surjective map that, locally, strengthens to become a diffeomorphism.*

Def. 1.4.1 (Covering) *Given the manifold M, a covering of M is the couple $(\tilde{M}, \pi)$, where $\tilde{M}$ is a manifold and $\pi : \tilde{M} \to M$ verifies the following properties:*

1. *π is smooth and surjective*

2. *for all $p \in M$ it exists an open connected neighborhood $U \subset M$ of p such that the restriction of π to all the connected components $\tilde{U} \subset \tilde{M}$ of $\pi^{-1}(U)$ is a diffeomorphism between $\tilde{U}$ and U.*

If $\tilde{M}$ is simply connected[15]*, then we say that $(\tilde{M}, \pi)$ is the* ***universal covering***[16] *of M.*

The components of $\pi^{-1}(U)$ are called the ***sheets of the covering****.*

[15] i.e. M, as topological space, is such that any continuous loop contained in M is homotopic to a point.

[16] If the universal covering exists, it can be proven to be unique up to homeomorphisms.

1.4.1 $\mathbb{R}$ and $\mathbb{R}^n$ as the universal covering of S^1_R and the torus $\mathbb{T}^n$

$\mathbb{R}$ is simply connected and we have seen that it is a covering of S^1_R, it follows that $\mathbb{R}$ **is the universal covering of** S^1_R. This is the 1-dimensional case of a more general covering involving $\mathbb{R}^n$ and the torus[17] $\mathbb{T}^n$.

Fixed any lattice $\Lambda \subset \mathbb{R}^n$, we can define an equivalence relation $\sim_\Lambda$ in $\mathbb{R}^n$ by identifying the elements of $\mathbb{R}^n$ that belong to the opposite edges, as depicted in Figure 1.2.

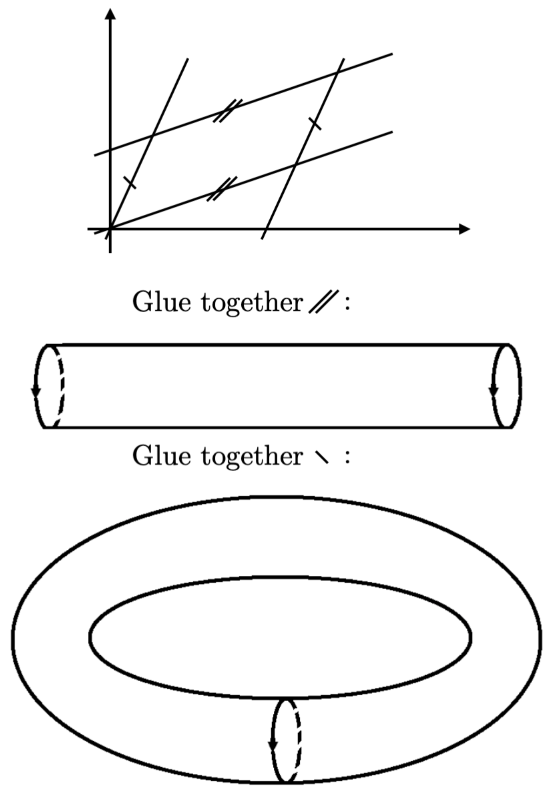


Figure 1.2: The construction of the torus $\mathbb{T}^2$.

$(\mathbb{R}^n, \pi)$, where $\pi := \mathbb{R}^n \to \mathbb{T}^n := \mathbb{R}^n / \sim_\Lambda$, $x \mapsto \pi(x) = [x]$, is the universal covering of the n-dimensional torus $\mathbb{T}^n$.

1.4.2 SU(2) as the two-sheets universal covering of SO(3)

The Lie group SU(2) is diffeomorphic to S^3, which is simply connected, thus it is simply connected itself. We prove that it is the universal covering of SO(3), the Lie group of proper rotations in $\mathbb{R}^3$.

To this aim, quaternions will help again, in fact, we shall prove that **all rotation in** $\mathbb{R}^3$ **can be accomplished by performing a conjugation with a unit quaternion**, analogously to the fact that all rotation in $\mathbb{R}^2$ can be performed through the multiplication by a 'phase factor', i.e. a unit complex number. To this aim, we identify $\mathbb{R}^3$ with the pure quaternions $\mathbb{H}_0$

[17] This peculiar name is derived from Latin, in which this word is used to describe a donuts-shaped pillow.

via the natural isomorphism:

$$\begin{array}{rcl} \mathbb{R}^3 & \xrightarrow{\sim} & \mathbb{H}_0 \\ x = (x, y, z) & \longmapsto & \mathbf{v} := ix + jy + kz, \end{array}$$

next, fixed any unit quaternion $q \in \mathbb{H}_1$ ($|q| = 1$, so $\overline{q} = q^{-1}$), the linear map

$$\begin{array}{rrcl} R_q : & \mathbb{H}_0 & \longrightarrow & \mathbb{H}_0 \\ & \mathbf{v} & \longmapsto & R_q(\mathbf{v}) := q\,\mathbf{v}\,\overline{q}, \end{array}$$

is well-posed because, as we have seen in section 1.3.3, the conjugation is an endomorphism of $\mathbb{H}_0$. Moreover, thanks to the multiplicative nature of the modulus in $\mathbb{H}$ we have:

$$|R_q(\mathbf{v})| = |q\,\mathbf{v}\,\overline{q}| = |q|\,|\mathbf{v}|\,|\overline{q}| = |\mathbf{v}|\,|qq^{-1}| = |\mathbf{v}|,$$

i.e. R_q is an isometry of $\mathbb{H}_0 \cong \mathbb{R}^3$, hence $R_q \in \mathrm{O}(3)$. Moreover, the matrix associated to R_q, $q = a + ib + jc + kd \in \mathbb{H}_1$, i.e. $a^2 + b^2 + c^2 + d^2 = 1$, w.r.t. the canonical basis of $\mathbb{R}^3$ is:

$$\mathbf{R}_q = \begin{pmatrix} a^2 + b^2 - c^2 - d^2 & 2bc - 2ad & 2bd + 2ac \\ 2bc + 2ad & a^2 - b^2 + c^2 - d^2 & 2cd - 2ab \\ 2bd - 2ac & 2cd + 2ab & a^2 - b^2 - c^2 + d^2 \end{pmatrix}.$$

By direct computation, we find $\det(\mathbf{R}_q) = (a^2 + b^2 + c^2 + d^2)^3 = 1$, so $R_q \in \mathrm{SO}(3)$, i.e. R_q is a proper rotation.

Actually, it can be proven that all matrix of SO(3) can be written as the matrix above, so the map $\mathbb{H}_1 \ni z \mapsto \mathbf{R}_q \in \mathrm{SO}(3)$ is onto.

Finally, from the fact that each entry of the matrix $\mathbf{R}_q$ is a polynomial of order two of the coefficients of q, it follows with simple calculations that $\mathbf{R}_q = \mathbf{R}_{q'} \iff q = q'$ or $q = -q'$. Thus, the correspondence $\mathbb{H}_1 \ni q \mapsto \mathbf{R}_q \in \mathrm{SO}(3)$ is $2:1$.

To resume, we have proven that SU(2) is the universal covering[18] of SO(3). We can say more: the onto map $q \mapsto \mathbf{R}_q$ is also a homomorphism of groups:

$$\begin{array}{rrcl} \pi : & \mathbb{H}_1 \cong \mathrm{SU}(2) & \twoheadrightarrow & \mathrm{SO}(3) \\ & q \cong A_q & \longmapsto & \pi(q) := \mathbf{R}_q, \end{array}$$

with

$$\ker(\pi) = \{I_2, -I_2\},$$

so that, by the homomorphism theorem, we have the isomorphism:

$$\mathrm{SU}(2)/\{I_2, -I_2\} \cong \mathrm{SO}(3).$$

Thinking about SU(2) and SO(3) as Lie groups, $\pi : \mathrm{SU}(2) \twoheadrightarrow \mathrm{SO}(3)$ defines a two-sheets covering (since the counter-image of $\mathbf{R}_q$ by π is $\pi^{-1}(\mathbf{R}_q) = \{q, -q\}$).

Finally, if we identify SU(2) with S^3, the quotient $\mathrm{SU}(2)/\{I_2, -I_2\}$ becomes the quotient of S^3 w.r.t. the equivalence relation $\sim_{\leftrightarrow}$ given by antipodal points identification on the 3-sphere. However, as we have seen in section 1.2, this quotient procedure gives rise to the real 3-dimensional projective space $\mathbb{RP}^3$, thus:

$$\mathbb{RP}^3 \;\cong\; S^3/\sim_{\leftrightarrow} \;\cong\; \mathrm{SU}(2)/\{I_2, -I_2\} \;\cong\; \mathrm{SO}(3)$$

and, thanks to these identifications, even the **3-dimensional real projective space $\mathbb{RP}^3$ acquires a Lie group structure**.

[18]In Physics, and in the formalism of Clifford algebras, the universal covering of SO(3) is called **Spin**(3).

1.4.3 Quaternions and Rodrigues' rotation formula

We end this section by unveiling the explicit link between the rotation that can be performed via conjugation with a unit pure quaternion and the well-known **Rodrigues' rotation formula**, . This formula establishes that if $\mathbf{v}$ is a vector in $\mathbb{R}^3$ and $\mathbf{n}$ is a unit vector describing an axis of rotation about which $\mathbf{v}$ rotates counterclockwise by an angle ϑ, then the rotated vector $\mathbf{v}_{\text{rot}}$ can be written as:

$$\boxed{\mathbf{v}_{\text{rot}} = \cos\vartheta\,\mathbf{v} + (1-\cos\vartheta)\,\langle \mathbf{n}, \mathbf{v}\rangle\,\mathbf{n} + \sin\vartheta\,(\mathbf{n}\times\mathbf{v})}\,. \tag{1.22}$$

In order to understand this quite involved formula, let us first notice that, given a vector $\mathbf{v} = (x, y)^t \in \mathbb{R}^2$ and a rotation matrix $R_\vartheta \in \mathrm{SO}(2)$, then

$$\begin{aligned} R_\vartheta \mathbf{v} &= \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\vartheta\, x - \sin\vartheta\, y \\ \sin\vartheta\, x + \cos\vartheta\, y \end{pmatrix} = \cos\vartheta \begin{pmatrix} x \\ y \end{pmatrix} + \sin\vartheta \begin{pmatrix} -y \\ x \end{pmatrix} \\ &= \cos\vartheta\,\mathbf{v} + \sin\vartheta\, R_{\frac{\pi}{2}}\mathbf{v}, \end{aligned}$$

these simple calculations show that a 2D rotation of a vector $\mathbf{v}$ can be seen as a linear combination between $\mathbf{v}$ and the vector $R_{\frac{\pi}{2}}\mathbf{v}$, whose defining features are the following the fact of being orthogonal to $\mathbf{v}$ and of having the same length.

Let us now increase the dimension and consider two vectors $\mathbf{v}, \mathbf{n} \in \mathbb{R}^3$ such that $\mathbf{n}$ is a unit vector and it is orthogonal to $\mathbf{v}$. Then, $\mathbf{n}\times\mathbf{v}$ is orthogonal to the plane defined by $\mathbf{n}$ and $\mathbf{v}$ and

$$|\mathbf{n}\times\mathbf{v}| = \cancelto{1}{|\mathbf{n}|}\,|\mathbf{v}|\,\cancelto{1}{\sin\frac{\pi}{2}} = |\mathbf{v}|.$$

Thus, thanks to the previous considerations, the 2D counterclockwise rotation by the angle ϑ of the 3D vector $\mathbf{v}$ *in the plane orthogonal to the axis defines by* $\boldsymbol{n}$ is:

$$\mathrm{Rot}(\mathbf{v}) = \cos\vartheta\,\mathbf{v} + \sin\vartheta\,(\mathbf{n}\times\mathbf{v}). \tag{1.23}$$

Finally, let us consider the 3D counterclockwise rotation by the angle ϑ of the 3D vector $\mathbf{v}$ around the axis defined by a unit vector $\mathbf{n}$, this time in a generic position w.r.t. $\mathbf{v}$. Taking advantage of what we have just learned, a valid strategy to obtain the formula for the rotated vector $\mathbf{v}_{\text{rot}}$ is to decompose $\mathbf{v}$ into its parallel

$$\mathbf{v}_\parallel = \langle \mathbf{n}, \mathbf{v}\rangle\mathbf{n}$$

and orthogonal

$$\mathbf{v}_\perp = \mathbf{v} - \langle \mathbf{n}, \mathbf{v}\rangle\mathbf{n}$$

components w.r.t. $\mathbf{n}$, then perform the 2D rotation of the perpendicular component (as we have learned above) and finally add the result to $\mathbf{v}_\parallel$ to recover the correct height of $\mathbf{v}_{\text{rot}}$. Figure 1.3 visualizes this construction.

Using eq. (1.23), the 2D rotation of $\mathbf{v}_\perp$ gives:

$$\begin{aligned} \mathrm{Rot}(\mathbf{v}_\perp) &= \cos\vartheta\,\mathbf{v}_\perp + \sin\vartheta\,(\mathbf{n}\times\mathbf{v}_\perp) = \cos\vartheta\,(\mathbf{v} - \langle \mathbf{n}, \mathbf{v}\rangle\mathbf{n}) + \sin\vartheta\,(\mathbf{n}\times(\mathbf{v} - \langle \mathbf{n}, \mathbf{v}\rangle\mathbf{n})) \\ &= \cos\vartheta\,\mathbf{v} - \cos\vartheta\,\langle \mathbf{n}, \mathbf{v}\rangle\mathbf{n} + \sin\vartheta\,(\mathbf{n}\times\mathbf{v}) - \sin\vartheta\,\langle \mathbf{n}, \mathbf{v}\rangle\cancelto{\mathbf{0}}{(\mathbf{n}\times\mathbf{n})} \\ &= \cos\vartheta\,\mathbf{v} - \cos\vartheta\,\langle \mathbf{n}, \mathbf{v}\rangle\mathbf{n} + \sin\vartheta\,(\mathbf{n}\times\mathbf{v}). \end{aligned}$$

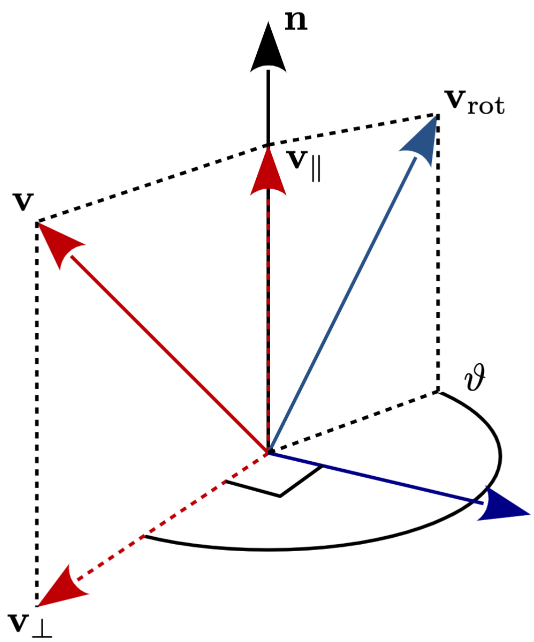


Figure 1.3: Interpretation of Rodrigues' rotation formula. Adapted from Wikipedia.

So, finally,

$$\mathbf{v}_{\text{rot}} = \mathbf{v}_{\|} + \text{Rot}(\mathbf{v}_{\perp}) = \cos\vartheta\,\mathbf{v} + (1-\cos\vartheta)\,\langle\mathbf{n},\mathbf{v}\rangle\,\mathbf{n} + \sin\vartheta\,(\mathbf{n}\times\mathbf{v}),$$

thus confirming eq. (1.22).

From the Rodrigues rotation formula we can obtain a useful parameterization of a rotation matrix $R \in \text{SO}(3)$. In fact, by expanding the formula we obtain:

$$\begin{aligned}
\mathbf{v}_{\text{rot}} &= \cos\vartheta \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} + (1-\cos\vartheta)(n_1v_1 + n_2v_2 + n_3v_3) \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} + \sin\vartheta \begin{pmatrix} n_2v_3 - v_3n_2 \\ n_3v_1 - n_1v_3 \\ n_1v_2 - n_2v_1 \end{pmatrix} \\
&= \begin{pmatrix} [\cos\vartheta + (1-\cos\vartheta)n_1^2]v_1 + [(1-\cos\vartheta)n_1n_2 - \sin\vartheta n_3]v_2 + [(1-\cos\vartheta)n_1n_3 + \sin\vartheta n_2]v_3 \\ [(1-\cos\vartheta)n_1n_2 + \sin\vartheta n_3]v_1 + [\cos\vartheta + (1-\cos\vartheta)n_2^2]v_2 + [(1-\cos\vartheta)n_2n_3 - \sin\vartheta n_1]v_3 \\ [(1-\cos\vartheta)n_1n_3 - \sin\vartheta n_2]v_1 + [(1-\cos\vartheta)n_2n_3 + \sin\vartheta n_1]v_2 + [\cos\vartheta + (1-\cos\vartheta)n_3^2]v_3 \end{pmatrix},
\end{aligned}$$

i.e. $\mathbf{v}_{\text{rot}} = R_{\vartheta,\mathbf{n}}\mathbf{v}$ with

$$R_{\vartheta,\mathbf{n}} = \begin{pmatrix} \cos\vartheta + (1-\cos\vartheta)n_1^2 & (1-\cos\vartheta)n_1n_2 - \sin\vartheta n_3 & (1-\cos\vartheta)n_1n_3 + \sin\vartheta n_2 \\ (1-\cos\vartheta)n_1n_2 + \sin\vartheta n_3 & \cos\vartheta + (1-\cos\vartheta)n_2^2 & (1-\cos\vartheta)n_2n_3 - \sin\vartheta n_1 \\ (1-\cos\vartheta)n_1n_3 - \sin\vartheta n_2 & (1-\cos\vartheta)n_2n_3 + \sin\vartheta n_1 & \cos\vartheta + (1-\cos\vartheta)n_3^2 \end{pmatrix}. \tag{1.24}$$

Rodrigues' formula can be written is a much more elegant and compact form thanks to the conjugation of $\mathbf{v}$ by a suitable unit quaternion.

Theorem 1.4.1 (Quaternionic Rodrigues' formula) *The vector $\mathbf{v}_{rot}$, representing the counterclockwise rotation by an angle ϑ of a vector $\mathbf{v} \in \mathbb{R}^3$ around the axis defined by the unit vector $\mathbf{u} \in \mathbb{R}^3$, can be expressed as the conjugation by the unit quaternion*

$$q = \cos(\vartheta/2) + \sin(\vartheta/2)\mathbf{u}, \tag{1.25}$$

i.e.

$$\mathbf{v}_{\text{rot}} = q\,\mathbf{v}\,\overline{q}. \tag{1.26}$$

Before going through the proof, we notice that $|q|^2 = \cos^2(\vartheta/2) + \sin^2(\vartheta/2)|\mathbf{u}|^2 = 1$, moreover, any *unit* quaternion $q = s_q + \mathbf{v}_q$ can be decomposed as in eq. (1.25), in fact we can re-write q as:

$$q = s_q + |\mathbf{v}_q|\frac{\mathbf{v}_q}{|\mathbf{v}_q|},$$

hence we can define $\mathbf{u} := \frac{\mathbf{v}_q}{|\mathbf{v}_q|}$ and select the angle ϑ such that $\cos(\vartheta/2) = s_q$ and $\sin(\vartheta/2) = |\mathbf{v}_q|$, which is possible because $s_q^2 + |\mathbf{v}_q|^2 = |q|^2 = 1$.

Proof. Let us write eq. (1.26) explicitly:

$$\begin{aligned} q\,\mathbf{v}\,\overline{q} &= (\cos(\vartheta/2) + \sin(\vartheta/2)\mathbf{u})\,\mathbf{v}\,(\cos(\vartheta/2) - \sin(\vartheta/2)\mathbf{u}) \\ &= \cos^2(\vartheta/2)\mathbf{v} + \sin(\vartheta/2)\cos(\vartheta/2)(\mathbf{uv} - \mathbf{vu}) - \sin^2(\vartheta/2)\mathbf{uvu} \\ &\underset{(1.16)}{=} \cos^2(\vartheta/2)\mathbf{v} + \sin(\vartheta/2)\cos(\vartheta/2)2(\mathbf{u}\times\mathbf{v}) - \sin^2(\vartheta/2)\mathbf{uvu} \end{aligned},$$

in fact (1.16) holds also by replacing the quaternions with their vector part since the terms with the scalar part cancel out in the difference. We see that in the first two terms of the expression above we have the same vectors that appear in Rodrigues' rotation formula, so we must re-write $\mathbf{uvu}$ in a suitable way.

To this aim, we start noticing that eq. (1.19) applied in our situation allows us writing $\langle \mathbf{u}, \mathbf{v}\rangle = -\frac{1}{2}(\mathbf{uv} + \mathbf{vu})$, so $\mathbf{uv} = -2\langle \mathbf{u}, \mathbf{v}\rangle - \mathbf{vu}$ and

$$\begin{aligned} \mathbf{uvu} &= (-2\langle \mathbf{u}, \mathbf{v}\rangle - \mathbf{vu})\mathbf{u} = -2\langle \mathbf{u}, \mathbf{v}\rangle\mathbf{u} - \mathbf{vuu} \\ &\underset{(1.17)}{=} -2\langle \mathbf{u}, \mathbf{v}\rangle\mathbf{u} - \mathbf{v}(\cancelto{0}{\mathbf{v}\times\mathbf{v}} - \cancelto{1}{\|\mathbf{u}\|^2}) \\ &= -2\langle \mathbf{u}, \mathbf{v}\rangle\mathbf{u} + \mathbf{v}. \end{aligned}$$

Replacing this expression of $\mathbf{uvu}$ in the formula for $q\,\mathbf{v}\,\overline{q}$ written above we find:

$$\begin{aligned} q\,\mathbf{v}\,\overline{q} &= \cos^2(\vartheta/2)\mathbf{v} + 2\sin(\vartheta/2)\cos(\vartheta/2)(\mathbf{u}\times\mathbf{v}) + 2\sin^2(\vartheta/2)\langle \mathbf{u}, \mathbf{v}\rangle\mathbf{u} - \sin^2(\vartheta/2)\mathbf{v} \\ &= (\cos^2(\vartheta/2) - \sin^2(\vartheta/2))\mathbf{v} + 2\sin(\vartheta/2)\cos(\vartheta/2)(\mathbf{u}\times\mathbf{v}) + 2\sin^2(\vartheta/2)\langle \mathbf{u}, \mathbf{v}\rangle\mathbf{u}. \end{aligned}$$

Thanks to the identities $\cos^2(\vartheta/2) - \sin^2(\vartheta/2) = \cos(\vartheta)$, $2\sin(\vartheta/2)\cos(\vartheta/2) = \sin(\vartheta)$ and $2\sin^2(\vartheta/2) = 1 - \cos(\vartheta)$, we find that $q\,\mathbf{v}\,\overline{q}$ agrees precisely with Rodrigues' formula. $\square$

Remark: any rotation in $\mathbb{R}^2 \cong \mathbb{C}$ is performed simply via the multiplication by a unit complex number, but in $\mathbb{R}^3 \cong \mathbb{H}_0$ this cannot happen: the *multiplication by a unit quaternion* is a rotation in $\mathbb{R}^4$, not $\mathbb{R}^3$! Instead, as the previous theorem shows, the *conjugation by a unit quaternion* is a well-defined rotation in $\mathbb{R}^3$. However, due to the appearance of q and $\overline{q}$ in the conjugation, the angle must be divided by 2, as explicitly underlined by formula (1.25).

The equations appearing in the previous theorem can also be written in an exponential form. In fact, the unit vector $\mathbf{u} \in \mathbb{R}^3$ can also be thought as belonging to $\mathbb{H}_0 \cap \mathbb{H}_1$, i.e. as a pure unit quaternion, but then, thanks to proposition 1.3.1, $\mathbf{u}$ is a square root of -1 in $\mathbb{H}$. These objects can be used to obtain the analogous of Euler's formula for $\mathbb{H}$.

Proposition 1.4.1 (Euler's formula for $\mathbb{H}$) *Let $q \in \mathbb{H}$ be a square root of -1. Then, for all $t \in \mathbb{R}$,*

$$e^{tq} = \cos(t) + \sin(t)q.$$

Proof. The property $q^2 = -1$ entails $q^{2n} = (-1)^n$ and $q^{2n+1} = (-1)^n q$ for all $n \in \mathbb{N}$. Hence,

$$\begin{aligned} e^{tq} &= \sum_{n=0}^{+\infty} \frac{t^n q^n}{n!} = \sum_{n=0}^{+\infty} \frac{t^{2n} q^{2n}}{(2n)!} + \sum_{n=0}^{+\infty} \frac{t^{2n+1} q^{2n+1}}{(2n+1)!} = \sum_{n=0}^{+\infty} \frac{t^{2n}(-1)^n}{(2n)!} + \sum_{n=0}^{+\infty} \frac{t^{2n+1}(-1)^n}{(2n+1)!} q \\ &= \cos(t) + \sin(t)q \end{aligned}$$

□

It follows that, for all unit vector $\mathbf{u} \in \mathbb{R}^3$ we can write:

$$q = \cos(\vartheta/2) + \sin(\vartheta/2)\mathbf{u} = e^{\frac{\vartheta}{2}\mathbf{u}}, \tag{1.27}$$

and so the quaternionic Rodrigues' formula has the following exponential form:

$$\mathbf{v}_{\text{rot}} = e^{\frac{\vartheta}{2}\mathbf{u}}\, \mathbf{v}\, e^{-\frac{\vartheta}{2}\mathbf{u}}. \tag{1.28}$$

1.5 Partition of the unity

Partitions of the unity are very important in differential geometry, because they allows us to extend the definition of objects from a local neighborhood of a point to the whole manifold. This is used, just to give an idea, for connections and Riemannian metrics.

Let us start with the following useful function displayed in Figure 1.4:

$$h : \mathbb{R} \to \mathbb{R}, \quad h(t) = \begin{cases} 0 & \text{if } t \leqslant 0 \\ e^{\frac{1}{t}} & \text{if } t > 0 \end{cases}$$

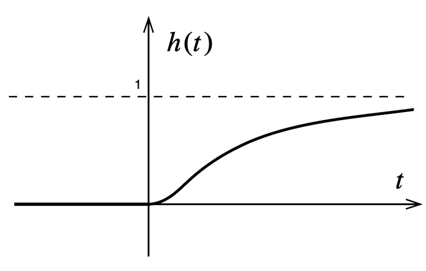


Figure 1.4: The h function.

The properties of h are listed below:

- $h(t) \in [0, 1)\ \forall t \in \mathbb{R}$ and $h(t) \underset{t \to +\infty}{\longrightarrow} 0$;
- h is increasing;
- $h \in \mathscr{C}^\infty(\mathbb{R})$.

With this smooth function h, we can cook up other one, depicted on the left hand side of Figure 1.5:

$$\eta : \mathbb{R} \to \mathbb{R}, \quad \eta(t) = \frac{h(1 - |t|^2)}{h(1 - |t|^2) + h(|t|^2 - \frac{1}{4})}$$

with the following characteristics:

- $\eta(t) \geqslant 0 \ \forall t \in \mathbb{R}$;
- $h \in \mathscr{C}^\infty(\mathbb{R})$;
- $\eta(t) = 1$ (exactly 1) in $[-\frac{1}{2}, \frac{1}{2}]$, in fact, for $t \in [-\frac{1}{2}, \frac{1}{2}]$, $|t|^2 \leqslant \frac{1}{4}$, so $h(|t|^2 - \frac{1}{4}) = 0$, because of the definition of h;
- $\eta(t) = 0$ for $t \geqslant 1$ or $t \leqslant -1$, in fact, in this case $1 - |t|^2 \geqslant 0$, so that $h(1 - |t|^2) = 0$ by definition in these intervals.

The extension to $\mathbb{R}^n$ is the following (for $n = 2$ the graph is depicted on the right hand side of Figure 1.5):

$$\eta : \mathbb{R}^n \to \mathbb{R}, \quad \eta(x) = \frac{h(1 - \|x\|^2)}{h(1 - \|x\|^2) + h(\|x\|^2 - \frac{1}{4})}, \quad \eta \in \mathscr{C}^\infty(\mathbb{R}^n),$$

- $\eta(t) \geqslant 0 \ \forall t \in \mathbb{R}$;
- $\eta(x) = 1$ (exactly 1) in $\overline{B(0, \frac{1}{2})}$;
- $\eta(x) = 0 \ \forall x \in \mathbb{R}^n \backslash B(0, 1)$.

η is called the **bump function**.

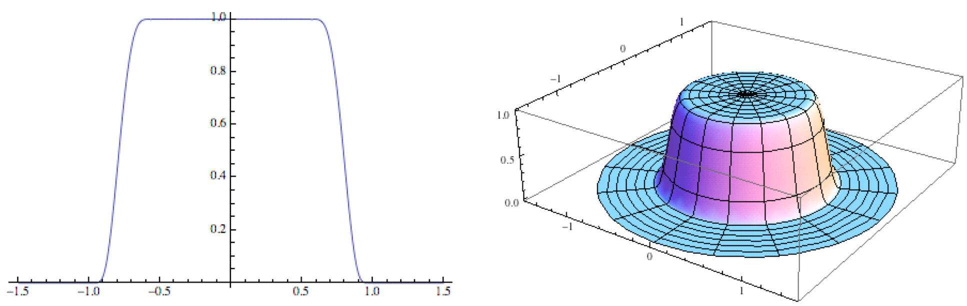


Figure 1.5: From left to right, the bump function η function for $n = 1$ and $n = 2$.

We recall that, given a topological space X, the support of a function $f : X \to \mathbb{R}$ is the closed subset of X defined by $\mathrm{supp}(f) = \overline{\{x \in X \ : \ f(x) \neq 0\}}$.

The following result is central in the theory of partitions of unity.

Theorem 1.5.1 *Let M be a smooth manifold and:*

- *$K \subset M$ a compact subset of M;*
- *$V \subset M$ an open subset of M containing K: $K \subset V$.*

Then, there exists a smooth function $g : M \to \mathbb{R}$ such that

$$\begin{cases} g|_K \equiv 1 \\ \operatorname{supp}(g) \subset V \implies g|_{M\setminus V} \equiv 0. \end{cases}$$

Thus, g is a generalization of the bump function to M: g is identically 1 on K, identically 0 on $M\setminus V$ and it takes intermediate (unknown) values on $V\setminus K$.

The proof is constructive.

Corollary 1.5.1 *For every point $p \in M$ and every open neighborhood $V \subseteq M$ of p, it exist $f, g \in \mathscr{C}^\infty(M)$ such that*

$$\begin{cases} f(p) = 0 \\ f|_{M\setminus V} \equiv 1 \end{cases}, \quad \begin{cases} g(p) = 1 \\ g|_{M\setminus V} \equiv 0 \end{cases}.$$

Proof. It is enough to choose $K = \{p\}$, obviously compact, in the previous theorem: we obtain a function $g \in \mathscr{C}^\infty(M)$ such that $g(p) = 1$ and $g|_{M\setminus V} \equiv 0$. Then, by setting $f(x) = 1 - g(x)$ for all $x \in M$, we obtain the thesis. □

Let us now introduce a handy symbol that will give a sort of generalization of *smooth functions for maps not necessarily defined on open sets.*

Def. 1.5.1 *Let $S \subset M$ be* ***any subset*** *of M. Then we denote with $\mathscr{C}^\infty(S)$ the set of continuous real-valued functions $f : S \to \mathbb{R}$ that can be obtained by restriction of a smooth function $\tilde{f} : V \to \mathbb{R}$, V open and $S \subset V$, i.e. $f = \tilde{f}\big|_S$.*

We use immediately this concept to show that **any $\mathscr{C}^\infty$ function defined on a compact subset of a manifold M can be extended to a smooth function on the whole M**... with *a sort of smooth padding with zeros*!

Theorem 1.5.2 (Extension theorem for smooth functions) *Let $K \subset M$ be* ***a compact*** *subset of the smooth manifold M and let $f \in \mathscr{C}^\infty(M)$. Let also $K \subset W$, W open in M. Then, it exists $\hat{f} \in \mathscr{C}^\infty(M)$ such that:*

- $\hat{f}\big|_K = f$;
- $\operatorname{supp}(\hat{f}) \subset W$, *so that* $\hat{f}\big|_{M\setminus W} \equiv 0$.

Proof. By definition, f extends to $\tilde{f} \in \mathscr{C}^\infty(U)$, for some U open in M, $K \subset U$.

We set $V = U \cap W$ and we consider $g \in \mathscr{C}^\infty(M)$ such that $g|_K \equiv 1$ and $\operatorname{supp}(g) \subset V$, which exists thanks to the previous result.

We define

$$\begin{aligned} \hat{f} :\ & M \longrightarrow \mathbb{R} \\ & q \longmapsto \hat{f}(q) = \begin{cases} g(q)\tilde{f}(q) & q \in V \\ 0 & q \in M\setminus V. \end{cases} \end{aligned}$$

$\hat{f}$ is smooth and $\hat{f}\Big|_K \equiv f$ because $g(q) = 1$ for all $q \in K$. Moreover, $\tilde{f}(q) = f(q)$ for all $q \in K$ and, finally, $\hat{f}\Big|_{M\setminus W} = 0$, because either $\hat{f}$ is evaluated outside V, or, in any case, g is 0. □

The last concept that we need is that of cover.

Def. 1.5.2 (Cover) *Let X be a topological space. A cover of X is a family of subsets $\mathcal{U} = \{U_\alpha\}_{\alpha \in I}$ of X such that $X = \bigcup\limits_{\alpha \in A} U_\alpha$. The cover is said to be:*

- ***open**, if all the sets U_α are open;*
- ***locally finite**, if every $p \in X$ has a neighborhood $U \subset X$ such that $U \cap U_\alpha \neq \varnothing$ only for a finite number of indices α.*

*Another covering $\mathcal{V} = \{V_\beta\}_{\beta \in J}$ is a **refining** of $\mathcal{U}$ if $\forall \beta \in J$ $\exists\, \alpha \in I$ such that $V_\beta \subset U_\alpha$, i.e. if the subsets of $\mathcal{V}$ are smaller than those of $\mathcal{U}$.*

Def. 1.5.3 (Partition of unity) *Let M be a smooth manifold. A partition of unity on M is a family of functions $\{\rho_\alpha : M \to \mathbb{R}\}_{\alpha \in I}$, I finite or infinite set, such that:*

1. *$\rho_\alpha \in \mathscr{C}^\infty(M)$;*
2. *$\rho_\alpha(p) \in [0,1]$ $\forall p \in M$, $\forall \alpha \in I$;*
3. *$\{\mathrm{supp}(\rho_\alpha)\}_{\alpha \in I}$ is a locally finite covering of M;*
4. *$\sum\limits_{\alpha \in I} \rho_\alpha(p) = 1$ $\forall p \in M$.*

The partition of unity is subordinated to the open covering $\mathcal{U} = \{U_\alpha\}_{\alpha \in I}$ of M if $\mathrm{supp}(\rho_\alpha) \subset U_\alpha$ $\forall \alpha \in I$.

The last property explains the name. The third property implies that $\sum\limits_{\alpha \in I} \rho_\alpha(p)$ is always a finite sum of real numbers, and not a series.

The fundamental result about partitions of unity is the following. The proof relies on the fact that the topological space underlying a smooth manifold is required to be second countable.

Theorem 1.5.3 *Every open covering of a smooth manifold admits a partition of unity subordinated to it.*

For the very technical proof see e.g. [10].

1.6 Manifolds with boundary

We start by defining the basic space underlying the theory of manifolds with border.

Def. 1.6.1 (Upper half space, its border and its interior) *The upper half space in $\mathbb{R}^n$ is the set*

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

The border of $\mathbb{H}^n$ is the hyperplane $\partial\mathbb{H}^n$ defined by the equation $x^n = 0$, i.e. it is the set

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

which can of course be identified with $\mathbb{R}^{n-1}$.

The interior of $\mathbb{H}^n$ is the set

$$\mathrm{int}(\mathbb{H}^n) := \mathbb{H}^n \backslash \partial\mathbb{H}^n = \{(x^1, \ldots, x^n) \in \mathbb{R}^n \ : \ x^n > 0\}.$$

The comprehension of these definitions is clear when $n = 2$: $\mathbb{H}^2$ and $\mathrm{int}(\mathbb{H}^2)$ are the half planes $\{(x, y) : x \in \mathbb{R},\ y \geqslant 0\}$ and $\{(x, y) : x \in \mathbb{R},\ y > 0\}$, respectively, while $\partial\mathbb{H}^2$ is the horizontal axis defined by $\{(x, y) : x \in \mathbb{R},\ y = 0\}$.

The construction of a manifold with boundary is totally analogous to that of a manifold, the only difference being that the local model this time is $\mathbb{H}^n$ instead of $\mathbb{R}^n$.

The first step consists of course in the definition of suitable charts in which the boundary makes its appearance.

Def. 1.6.2 (n-boundary and inner charts) *Let M be a topological space. A n-boundary chart for M is a couple (U, φ) where $U \subset M$ is an open subset (resp. a subset) and $\varphi : U \to V$ is a homeomorphism with an open subset $V \subset \mathbb{H}^n$ which intersects the boundary of $\mathbb{H}^n$ non-trivially, i.e. such that $V \cap \partial\mathbb{H}^n \neq \varnothing$. Otherwise, when $V \cap \partial\mathbb{H}^n = \varnothing$, we will call (U, φ) a n-inner chart for M, hence if (U, φ) a n-inner chart for M, then φ takes values in* $\mathrm{int}(\mathbb{H}^n)$.

Recall that $\mathbb{R}^n$ and $\mathrm{int}(\mathbb{H}^n)$ are homeomorphic via, e.g. the function

$$\mathbb{R}^n \ni (x^1, \ldots, x^{n-1}, x^n) \longmapsto (x^1, \ldots, x^{n-1}, \exp(x^n)) \in \mathbb{R}^{n-1} \times (0, +\infty),$$

which is of course continue and invertible with continuous inverse given by

$$\mathbb{R}^{n-1} \times (0, +\infty) \ni (x^1, \ldots, x^{n-1}, \exp(x^n) > 0) \longmapsto (x^1, \ldots, x^{n-1}, \log(\exp(x^n)) = x^n) \in \mathbb{R}^n,$$

thus an n-inner chart is a chart in the usual sense.

The fundamental difference with an n-inner chart and a n-boundary chart is thus the fact that, in the first case the chart homeomorphism is between two open sets of $\mathbb{R}^n$, while in the second the homeomorphism is between an open set in $\mathbb{R}^n$ and an open set in the topology induced on $\mathbb{H}^n$, which *is not* an open set of $\mathbb{R}^n$ because it contains a part of the boundary of $\mathbb{H}^n$. We will not repeat the definitions and properties of usual charts for n-inner charts and we will only focus on the boundary ones. Figure 1.6 gives a graphical representation of the difference between the two kinds of maps.

The next step is the definition of an atlas for a manifold with boundary, which needs the concept of compatibility between n-boundary charts. To define them we first need to extend the concept of smooth function to $\mathbb{H}^n$

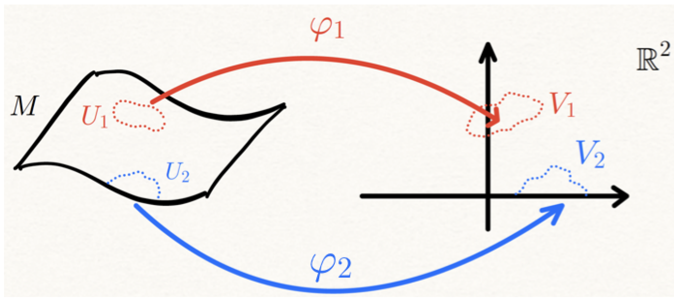


Figure 1.6: The difference between an 2-inner and a 2-border chart map.

Def. 1.6.3 *A function $f : W_1 \to W_2$, where W_1, W_2 are open sets in $\mathbb{H}^n$ is smooth if there exist two open sets U_1, U_2 in $\mathbb{R}^n$ and a smooth function $F : U_1 \to U_2$ such that $W_j \subset U_j$, $j = 1, 2$ and $F|_{W_1} \equiv f$.*

Def. 1.6.4 (Compatibility between n-boundary charts) *Two n- boundary charts $(U_\alpha, \varphi_\alpha)$ and (U_β, φ_β) are compatible if $\varphi_\alpha(U_\alpha \cap U_\beta)$ and $\varphi_\beta(U_\alpha \cap U_\beta)$ and the transition function*

$$\varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(U_\alpha \cap U_\beta) \to \varphi_\alpha(U_\alpha \cap U_\beta)$$

is a smooth diffeomorphism.

Of course, the smoothness must be interpreted as in the previous definition, i.e. $\varphi_\alpha \circ \varphi_\beta^{-1}$ must admit a smooth diffeomorphism in $\mathbb{R}^n$ between an open set (in $\mathbb{R}^n$) containing $\varphi_\beta(U_\alpha \cap U_\beta)$ and $\varphi_\alpha(U_\alpha \cap U_\beta)$, as depicted in Figure 1.7.

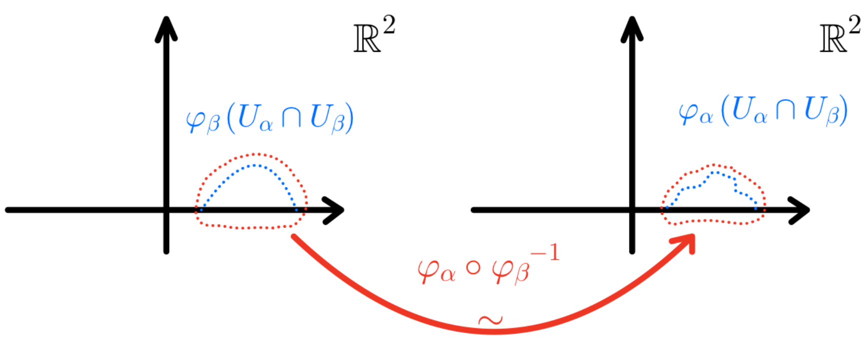


Figure 1.7: Smoothness between 2-border chart maps.

We can finally define a manifold with border.

Def. 1.6.5 (Manifold with border) *A manifold with border is the couple given by a Hausdorff topological manifold M with a countable basis and an atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}_\alpha$ of n-inner charts compatible with each other, in the usual sense, and also n-boundary charts, compatible with each other in the sense defined above, such that $M = \bigcup_\alpha U_\alpha$.*

Differently from usual manifolds, we can now distinguish between border and interior of a manifold with border.

Def. 1.6.6 (Border and interior of a manifold with border) *Let M be a manifold with border, we define its border as the set of points of M which are inverse-images of points belonging to $\partial\mathbb{H}^n$ via a n-border chart, formally:*

$$\partial M := \{p \in M \ : \ \exists \text{ an } n\text{-border chart } (U_\alpha, \varphi_\alpha) \text{ such that } p \in \varphi_\alpha^{-1}(\partial\mathbb{H}^n)\}.$$

The interior of M is

$$\text{int}(M) := M \backslash \partial M.$$

The previous definition is well-posed thanks to the following result, whose proof is not difficult, but quite long and technical, so we avoid its reproduction here.

Theorem 1.6.1 (Invariance of the border w.r.t. n-border charts) *Let M be a manifold with border and let p be a generic point of M. If it exists a n-border chart (U, φ) of M such that $\varphi(U) \subseteq \mathbb{H}^n$ and $\varphi(p) \in \partial\mathbb{H}^n$, then the same holds true for all other chart whose domain contains p.*

Corollary 1.6.1 *The border ∂M of a n-dimensional manifold with border M is a manifold (without border) of dimension $n - 1$.*

Proof. From the previous theorem it follows that if $(U_\alpha, \varphi_\alpha)$ and (U_β, φ_β) are two compatible n-border charts of M, then the border $\partial\mathbb{H}^n$ is stable w.r.t. the transition function $\varphi_\alpha \circ \varphi_\beta^{-1}$.

Hence, we can build an atlas for ∂M by considering the restriction of all the n-border charts of M to ∂M, they will surely be compatible because $\varphi_\alpha \circ \varphi_\beta^{-1}$ is a smooth function, so its restriction to $\mathbb{H}^n$ is still smooth. Since the local model of ∂M is $\partial\mathbb{H}^n \simeq \mathbb{R}^{n-1}$, its dimension is $n-1$. □

It is also clear that the n-inner charts of a n-dimensional manifold with boundary M provide a differential structure of manifold without boundary for $\text{int}(M)$ and that its dimension is still n.

Hence, a n-dimensional manifold with boundary M always comes with two companion manifolds without boundaries: the $n - 1$-dimensional manifold given by its border ∂M (so that $\partial(\partial M) = \varnothing$) and the n-dimensional manifold given by its interior $\text{int}(M)$.

An explicit example of manifold with boundary will be exhibited in chapter 2 because its proof needs the so-called rank theorem that will be introduced later.

Chapter 2

Tangent vector and tangent space to a manifold at a point

Inspirational epithap wanted...

...

Disclaimer: the reader is invited to get acquainted with the notations and concepts discussed in Appendix B about ordinary differential calculus in $\mathbb{R}^n$ before reading this chapter.

A firm understanding of the concept of tangent vector and tangent space to a point of a manifold is the most important step towards the comprehension of more advanced concepts of differential geometry.

There are at least five different, but (of course) equivalent[1] ways to define a tangent vector to a point of a manifold. Each one has advantages and disadvantages, but all of them must be known. A thorough analysis of the equivalence between these definitions is available in [7].

1. **Geometrical definition**: **tangent vectors as equivalent class of curves**. It is an intuitive definition, but not the easiest one to use in proofs or for its notation;
2. **Algebraic definition ♯ 1**: **tangent vectors as derivations of smooth scalar functions**. It is probably the most widely used in the literature, thanks to its notational and conceptual simplicity. It is the one that we will use more commonly throughout this document.
3. **Algebraic definition ♯ 2**: **tangent vectors as derivations of germs of smooth functions**. It is similar to the previous one, it has the advantage to make the local nature of tangent vectors even clearer and of being extendable to real-analytic and complex manifolds, but it has the disadvantages of being even more abstract and with a less simple notation.
4. **Physicists' definition**: **tangent vectors as equivalence classes of n-tuples**. It is mainly used by physicists and engineers, it uses the fact that tangent vectors verify a peculiar way of transforming under coordinate transformations.

[1] A perfect equivalence holds only for finite-dimensional manifolds. If the manifold dimension is infinite, the situation is trickier.

5. **Jets definition**: it is a quite abstract definition, that we will not discuss here, but it as a great importance in modern versions of calculus of variations, covariant geometric field theory and general relativity.

2.1 Geometric definition of tangent vectors

We start introducing tangent vectors with the most geometrical way. Later, we will discuss the algebraic and the physicists' way and prove their equivalence.

Following [6], let us be guided by the very easy example of the unit spheres S^1 and S^2 depicted in Fig. 2.1 (courtesy of Eric Shapiro) to understand how to define tangent vectors.

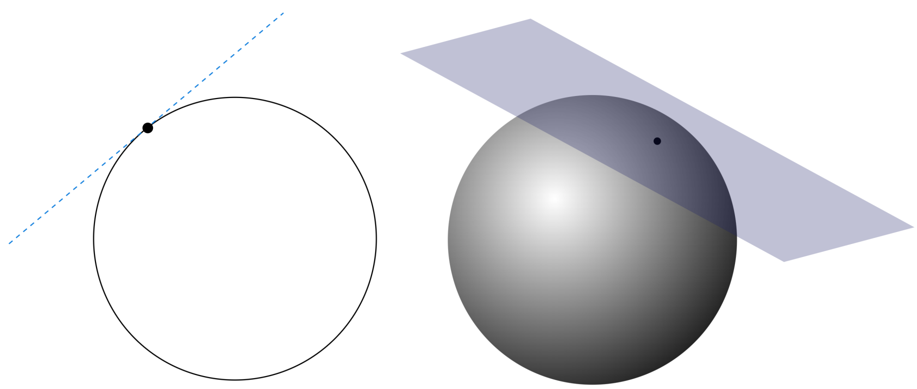


Figure 2.1: Intuitive depiction of tangent line to a circle (left) and tangent plane to a sphere (right).

We see that, while S^1 and S^2 are manifolds of dimension 1 and 2, respectively, the tangent line to a point of S^1 and the tangent plane to a point of S^2 live in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. While this may not be a problem for manifolds naturally embedded in $\mathbb{R}^{n+1}$ as the sphere S^n, for a generic abstract manifold[2] M of dimension n it is desirable to have an *intrinsic* definition of tangent vector and space, that does not make use of a larger structure.

It turns out that manifold-valued paths are exactly what we need to provide such an intrinsic definition.

Given a path γ passing through $p \in M$, the tangent vector to γ in p, i.e. the velocity at which γ passes through p, will be also tangent to M at p, since the image of γ lies in M, as shown in Figure 2.1.

To make this intuition precise, we must first define what the tangent vector to a path in M is. As always, since we know how to compute the tangent vector of a path in the local model $\mathbb{R}^n$, we can consider any local chart (U, φ) in p and build a path in $\mathbb{R}^n$ simply by composing γ with $\varphi|_{\gamma(-\varepsilon,\varepsilon)\cap U}$, that we will still denote with φ for simplicity:

$$\begin{array}{rccl} \varphi \circ \gamma : & (-\varepsilon, \varepsilon) & \longrightarrow & \mathbb{R}^n \\ & t & \longmapsto & (\varphi \circ \gamma)(t), \end{array}$$

[2]By **Whitney's embedding theorem** [10], every n-dimensional manifold M can be embedded in $\mathbb{R}^{2n+1}$, however, the fact that this embedding exists, does not mean that it is convenient to think about M as an embedded submanifold of $\mathbb{R}^{2n+1}$. For example, in general relativity, spacetime is a 4-dimensional manifold and it meaningless to embed it into $\mathbb{R}^9 \ldots$

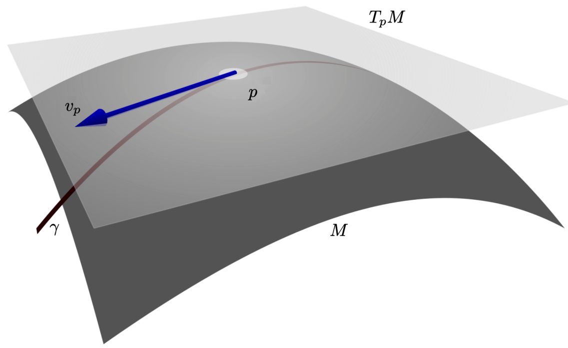


Figure 2.2: Tangent vector to a path in M at a given point p.

since $(\varphi \circ \gamma)(0) = \varphi(p) = x \in \mathbb{R}^n$, $\varphi \circ \gamma$ is a path in $\mathbb{R}^n$ passing through $x = \varphi(p)$.

Using the standard definition of calculus, the tangent vector to the curve $\varphi \circ \gamma$ at x is:

$$(\varphi \circ \gamma)^{\cdot}(0) := \frac{(\varphi \circ \gamma)(t) - (\varphi \circ \gamma)(0)}{t} \equiv \left.\frac{d}{dt}\right|_{t=0} (\varphi \circ \gamma)(t).$$

Of course, in general, there may be other curves passing through p with the property that their local representations via φ have the same tangent vector as γ.

The following basic lemma shows that, remarkably, if the local representations any two curves passing through p have the same tangent vector in $\mathbb{R}^n$ w.r.t. a given local chart in p, then this holds for any other local chart in p.

Lemma 2.1.1 *Let $(U_\alpha, \varphi_\alpha), (U_\beta, \varphi_\beta)$ be two overlapping charts in p and γ, σ two paths passing through p. Define:*

$$\gamma_\alpha := \varphi_\alpha \circ \gamma,\ \gamma_\beta := \varphi_\beta \circ \gamma \ \ \textit{and} \ \ \sigma_\alpha := \varphi_\alpha \circ \sigma,\ \sigma_\beta := \varphi_\beta \circ \sigma.$$

Then:

$$\dot{\gamma_\alpha}(0) = \dot{\sigma_\alpha}(0) \iff \dot{\gamma_\beta}(0) = \dot{\sigma_\beta}(0).$$

Proof. With the notations of the Lemma we have:

$$\dot{\gamma_\beta}(0) = (\varphi_\beta \circ \gamma)^{\cdot}(0) = (\varphi_\beta \circ \varphi_\alpha^{-1} \circ \varphi_\alpha \circ \gamma)^{\cdot}(0) = (\eta_{\beta\alpha} \circ \gamma_\alpha)^{\cdot}(0),$$

where $\eta_{\beta\alpha}$ is the (smooth) transition function between charts. Thanks to eq. (B.7) we have:

$$\dot{\gamma_\beta}(0) = D(\eta_{\beta\alpha} \circ \gamma_\alpha)(0)1 \underset{\text{(chain rule)}}{=} D\eta_{\beta\alpha}(\gamma_\alpha(0))D\gamma_\alpha(0)1 = D\eta_{\beta\alpha}(x)\dot{\gamma_\alpha}(0). \tag{2.1}$$

Of course, the same holds for the path η, i.e. $\dot{\sigma_\beta}(0) = D\eta_{\beta\alpha}(x)\dot{\sigma_\alpha}(0)$. By the linearity of the operator $D\eta_{\beta\alpha}(x)$, it follows that:

$$\dot{\gamma_\beta}(0) - \dot{\sigma_\beta}(0) = D\eta_{\beta\alpha}(x)(\dot{\gamma_\alpha}(0) - \dot{\sigma_\alpha}(0)).$$

Now, $\eta_{\beta\alpha}$ is a local diffeomorphism, thus $D\eta_{\beta\alpha}(x)$ is a linear isomorphism (the Jacobian matrix of $\eta_{\beta\alpha}$ in x has non null determinant), thus $\dot{\gamma_\alpha}(0) - \dot{\sigma_\alpha}(0) = 0$ if and only if $\dot{\gamma_\beta}(0) - \dot{\sigma_\beta}(0) = 0$, which proves the theorem. □

This lemma implies that the equality of the tangent vector in $\mathbb{R}^n$ for the local representation of two curves in M passing through the same point is an *intrinsic property of the manifold* M, meaning that it does not depend on the local chart chosen. This property allows us to define an equivalence relationship in the set of curves and also the first, geometric, definition of tangent vector to a manifold at a certain point.

Def. 2.1.1 (Tangentially equivalent, or tangent, curves) *Let M be an n-dimensional manifold and $p \in M$ fixed. Two paths γ, σ in M passing through p are tangent, or tangentially equivalent, if there exists a local chart φ in p such that they identify the same tangent vector in $\mathbb{R}^n$ when composed with φ, i.e.*

$$(\varphi \circ \gamma)^{\boldsymbol{\cdot}}(0) = (\varphi \circ \sigma)^{\boldsymbol{\cdot}}(0).$$

Being defined via an equality, the fact of being tangentially equivalent is easily seen to be indeed an equivalence relationship in the set of curves in M passing through p.

Def. 2.1.2 (Geometric tangent vectors and tangent space to M at p) *A (geometric) tangent vector to M at p is a tangentially equivalence class of curves passing through p, denoted with $[\gamma]$. The (geometric) tangent space to M at p, denoted with $T_p^{\text{geom}}M$ is the set of all tangent vectors to M at p.*

Remark: a slightly different definition of tangent vector can be obtained in a similar manner, replacing the local charts with smooth scalar functions, in this case we define two paths γ and η to be equivalent if, for all $f \in \mathscr{C}^\infty(M)$, $(f \circ \gamma)^{\boldsymbol{\cdot}}(0) = (f \circ \sigma)^{\boldsymbol{\cdot}}(0)$, where both $f \circ \gamma$ and $f \circ \sigma$ are scalar functions of a real variable. In this case we say that γ and η have a **contact of first order** in p (a **contact of order zero** being simply the fact that the pass through the same point, i.e. $\gamma(0) = \eta(0) = p$).

The set of curves in M passing through p quotiented w.r.t. the tangential equivalence turns out to be a copy of $\mathbb{R}^n$, as stated in the following result.

Theorem 2.1.1 *Fixed a local chart (U, φ) in $p \in M$, the map*

$$\begin{array}{rccl} I_{p,\varphi} : & T_p^{\text{geom}}M & \xrightarrow{\sim} & \mathbb{R}^n \\ & [\gamma] & \longmapsto & I_{p,\varphi}([\gamma]) = (\varphi \circ \gamma)^{\boldsymbol{\cdot}}(0), \end{array}$$

which associates to a tangentially equivalence class of paths passing through p their common tangent vector $(\varphi \circ \gamma)^{\boldsymbol{\cdot}}(0)$ in $\mathbb{R}^n$ w.r.t. the local chart φ, is a bijection.

Proof. Injectivity is obvious: different tangential classes of curves are associated to different tangent vectors in $\mathbb{R}^n$.

To prove surjectivity, fixed any $v \in \mathbb{R}^n$, we must prove that there exists $[\gamma] \in T_p^{\text{geom}}M$ such that $I_{p,\varphi}([\gamma]) = (\varphi \circ \gamma)^{\boldsymbol{\cdot}}(0) = v$. This can be done very simply by lifting to M via φ^{-1} the segment of straight line (restricted to $(-\varepsilon, \varepsilon)$) passing through $x = \varphi(p)$ and directed as v, i.e. $r_{x,v} : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$, $r_{x,v}(t) := x + tv$, where $\varepsilon > 0$ is small enough so that $r_{x,v}(-\varepsilon, \varepsilon)$ is contained in $\varphi(U)$:

$$\gamma = \varphi^{-1} \circ r_{x,v} : (-\varepsilon, \varepsilon) \to M, \quad \boxed{\gamma(t) = \varphi^{-1}(x + tv)}, \qquad t \in (-\varepsilon, \varepsilon). \tag{2.2}$$

γ is such that $\gamma(0) = \varphi^{-1} \circ r_{x,v}(0) = \varphi^{-1}(x) = \varphi^{-1}(\varphi(p)) = p$, hence, to prove surjectivity it remains only to check that the tangent vector in $\mathbb{R}^n$ of the local representation of γ associated to φ, i.e. $(\varphi \circ \gamma)^{\boldsymbol{\cdot}}(0)$, coincides with v:

$$I_{p,\varphi}([\gamma]) = (\varphi \circ \gamma)^{\boldsymbol{\cdot}}(0) = (\varphi \circ \varphi^{-1}(x + tv))^{\boldsymbol{\cdot}}(0) = (x + tv)^{\boldsymbol{\cdot}}(0) = v.$$

$\square$

$T_p^{\text{geom}}M$ cannot be canonically identified with $\mathbb{R}^n$ because $I_{p,\varphi}$ depends both on the point p and the local chart φ: changing the point p on M and/or the local chart φ changes the identification with $\mathbb{R}^n$.

Since the elements of $T_p^{\text{geom}}M$ are called tangent vectors, we expect $T_p^{\text{geom}}M$ to be a vector space, this is actually the case. The linear structure of $T_p^{\text{geom}}M$ is borrowed from that of $\mathbb{R}^n$ thanks to the bijection provided by $I_{p,\varphi}$.

Linear structure of $T_p^{\mathbf{geom}}M$:

$$[\gamma] + [\sigma] := I_{p,\varphi}^{-1}(I_{p,\varphi}([\gamma]) + I_{p,\varphi}([\sigma])), \quad [\gamma], [\sigma] \in T_p^{\text{geom}}M$$
$$k[\gamma] := I_{p,\varphi}^{-1}(kI_{p,\varphi}([\gamma])), \quad k \in \mathbb{R}.$$

This definition of linear structure seems to depend on φ, however it does not, it is intrinsic. We prove this for the sum, an analogous proof holds for the product by a real coefficient.

Using the hypotheses and notations of Lemma 2.1.1, we have:

$$I_{p,\varphi_\beta}([\gamma]) = \dot{\gamma_\beta}(0) \underset{(2.1)}{=} D\eta_{\beta\alpha}(x)\dot{\gamma_\alpha}(0) = (D\eta_{\beta\alpha}(x) \circ I_{\varphi_\alpha,p})([\gamma]),$$

since this holds for all $[\gamma] \in T_p^{\text{geom}}M$, we have:

$$I_{p,\varphi_\beta} = D\eta_{\beta\alpha}(x) \circ I_{\varphi_\alpha,p} \iff I_{p,\varphi_\beta}^{-1} = I_{\varphi_\alpha,p}^{-1} \circ (D\eta_{\beta\alpha}(x))^{-1}, \qquad x = \varphi_\alpha(p). \tag{2.3}$$

If we denote temporarily with $+_\alpha$ and $+_\beta$ the sum brought to $T_p^{\text{geom}}M$ by the local charts φ_α and φ_β, respectively, then:

$$\begin{aligned}
[\gamma] +_\beta [\sigma] &= I_{p,\varphi_\beta}^{-1}(I_{p,\varphi_\beta}([\gamma]) + I_{p,\varphi_\beta}([\sigma])) \\
&\underset{(2.3)}{=} I_{\varphi_\alpha,p}^{-1} \circ (D\eta_{\beta\alpha}(x))^{-1}((D\eta_{\beta\alpha}(x) \circ I_{\varphi_\alpha,p})([\gamma]) + (D(\eta_{\beta\alpha}(x)) \circ I_{\varphi_\alpha,p})([\sigma])) \\
&\underset{(\text{lin. of } D\eta_{\beta\alpha}(x)}{=} (I_{\varphi_\alpha,p}^{-1} \circ (D\eta_{\beta\alpha}(x))^{-1} \circ D\eta_{\beta\alpha}(x))(I_{\varphi_\alpha,p}([\gamma]) + I_{\varphi_\alpha,p}([\sigma])) \\
&= I_{\varphi_\alpha,p}^{-1}(I_{\varphi_\alpha,p}([\gamma]) + I_{\varphi_\alpha,p}([\sigma])) \\
&= [\gamma] +_\alpha [\sigma].
\end{aligned}$$

The bijection $I_{p,\varphi}$ becomes a linear isomorphism between $T_p^{\text{geom}}M$ and $\mathbb{R}^n$ and thus it can be used to transport a basis of $\mathbb{R}^n$ to a basis of $T_p^{\text{geom}}M$. The easiest one is of course the canonical basis of $\mathbb{R}^n$, thanks to the proof of the surjectivity of $I_{p,\varphi}$, we have that the **basis of geometric tangent vectors of $T_p^{\mathbf{geom}}M$ associated to the canonical basis of $\mathbb{R}^n$** is:

$$\left([\varphi^{-1} \circ r_{x,e_1}], \dots, [\varphi^{-1} \circ r_{x,e_n}]\right) \equiv \left([t \mapsto \varphi^{-1}(x + te_1)], \dots, [t \mapsto \varphi^{-1}(x + te_n)]\right), \tag{2.4}$$

where φ is any local chart in p and $r_{x,e_i} : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ is the straight line segment passing through $x = \varphi(p)$, contained in $\varphi(U)$ and parallel to the i-th coordinate axis, $i = 1, \ldots, n$.

Finally, $I_{p,\varphi}$ can be used also to transport any norm of $\mathbb{R}^n$ to $T_p^{\text{geom}}M$. In this case, the norm on $T_p^{\text{geom}}M$ depends on φ, however, topologically speaking, this creates no problem at all because it is well-known that all norms on finite-dimensional vector spaces are equivalent, in particular, they are equivalent to the Euclidean norm.

To resume, $T_p^{\textbf{geom}}M$ **is a normed vector space isomorphic to a copy of** $\mathbb{R}^n$ **but not canonically**.

Before passing to the algebraic definition of tangent vectors and tangent space, we introduce the concept of differential of push forward of a smooth function.

Def. 2.1.3 (Differential (or push-forward, or tangent map) of a smooth function) *Given the smooth function $f : M \to N$, the differential (or push forward, or tangent map) of f at p is the map that transforms a tangentially equivalence class of paths passing through p to a tangentially equivalence class of paths passing through $f(p)$ simply by composition, i.e. :*

$$\begin{array}{rccl} df_p \equiv f_* : & T_p^{\text{geom}}M & \longrightarrow & T_{f(p)}^{\text{geom}}N \\ & [\gamma] & \longmapsto & df_p([\gamma]) \equiv f_*([\gamma]) = [f \circ \gamma]. \end{array} \tag{2.5}$$

With a quite technical computation that uses the definition of the linear structure of $T_p^{\text{geom}}M$, it can be proven that df_p is a linear operator. The non manifestly linear nature of $T_p^{\text{geom}}M$ and of the differential of smooth functions is one of the main reasons why mathematicians searched for an alternative definition, the algebraic one, which makes linearity manifest, as we are going to discuss in the next section.

2.2 Algebraic definition of tangent vectors

The link between the geometric and algebraic definition of tangent vectors on a manifold passes through the following considerations. An element of $\mathbb{R}^n$ can be interpreted either as a point, say $x \in \mathbb{R}^n$, and as a vector $v \in \mathbb{R}^n$, once these ones are fixed, there is just one way to define the directional derivative $D_v f(x)$ of a scalar valued function $f : \mathbb{R}^n \to \mathbb{R}$ in x along the direction defined by v, as discussed in Appendix B.

Directional derivatives are linear, they satisfy Leibniz's rule when applied to the product of two functions and they are null when a function is constant along the direction of derivation.

It turns out that these properties are necessary and sufficient to identify a tangent vector on a manifold in an algebraic way. This alternative vision, as it will be proven, is fully equivalent to the geometric one previously discussed.

This algebraic abstraction of a tangent vector is typical in modern mathematics and it can be considered as the analogous of the algebraic abstraction that leads to the definition of scalar product in an arbitrary vector space: in that case, the properties of bilinearity, symmetry and positive-definiteness are necessary and sufficient to identify a form on a real vector space as a scalar product. The advantages of this abstractions are known, e.g. the possibility to define a scalar product for vector spaces of any dimension and whose elements are not necessarily vectors in the Euclidean sense, but also polynomials, functions and so on.

The algebraic abstraction of the concept of tangent vector starts with the definition of a derivation on the set of smooth real scalar functions from M to $\mathbb{R}$, denoted with $\mathscr{C}^\infty(M)$, in a point $p \in M$. $\mathscr{C}^\infty(M)$ is a real algebra w.r.t. the point-wise linear operations and multiplication.

Def. 2.2.1 (Derivation on $\mathscr{C}^\infty(M)$ in a point) *Let $f, g \in \mathscr{C}^\infty(M)$. Fixed $p \in M$, a derivation on $\mathscr{C}^\infty(M)$ in $p \in M$ is a linear functional $v : \mathscr{C}^\infty(M) \to \mathbb{R}$ that satisfies the following* ***Leibniz rule****:*

$$\boxed{v(fg) = f(p)v(g) + g(p)v(f)}\,, \qquad \forall f, g \in \mathscr{C}^\infty(M).$$

The set of derivations on $\mathscr{C}^\infty(M)$ in $p \in M$ is easily proven to be a real vector space, w.r.t. the point-wise linear operations, that we denote with $\mathrm{Der}_p(M)$.

The Leibniz rule implies two basic properties of derivations.

Lemma 2.2.1 *Let $v \in Der_p(M)$ and $f, g \in \mathscr{C}^\infty(M)$. Then:*

1. *v sets to 0 constant functions: if $k_c \equiv c$, i.e. $k_c(q) = c \in \mathbb{R}$ for all $q \in M$, then $v(k_c) = 0$;*
2. *If f, g take null values in the application point p, i.e. $f(p) = g(p) = 0$, then $v(fg) = 0$.*

Proof.

1.: let $k_1 \equiv 1$, then:

$$\begin{aligned} v(k_c) &= v(k_c k_1) = v(ck_1) \underset{v\text{ lin.}}{=} cv(k_1) = cv(k_1 \cdot k_1) \underset{\text{Leibniz}}{=} c(k_1(p)v(k_1) + k_1(p)v(k_1)) \\ &= 2cv(k_1) = 2v(ck_1) = 2v(k_c), \end{aligned}$$

i.e. $v(k_c) = 0$.

2.: $v(fg) = f(p)v(g) + g(p)v(f) = 0v(g) + 0v(f) = 0$. □

The following property is of fundamental importance: it says that **derivations act locally**, in the sense that only the values taken by a smooth function on an arbitrarily small open neighborhood of the application point matter to define the action of the derivation.

Theorem 2.2.1 *Let $v \in Der_p(M)$ and $f, g \in \mathscr{C}^\infty(M)$. If there exists any open neighborhood $U \subseteq M$ of p such that $f|_U = g|_U$, then $v(f) = v(g)$.*

Proof. By hypothesis, $f - g$ is a smooth function on M that vanishes in U. Thanks to proposition 1.5.1, we know that it exists a smooth function $h \in \mathscr{C}^\infty(M)$ such that $h(p) = 1$ and $h|_{M\setminus U} \equiv 0$, then the product function $(f-g)h$ is zero, thus, thanks to the Leibniz property:

$$0 = v((f-g)h) = v(f-g)\cancelto{1}{h(p)} + (f-g)(p)v(h) = v(f) - v(g) + \cancelto{0}{(f(p) - g(p))}\, v(h),$$

i.e. $v(f) = v(g)$. □

We are now ready to define the concept of algebraic tangent vector.

Def. 2.2.2 (Algebraic tangent vector and space) *The vector space $T_p^{\mathrm{alg}}M$, called algebraic tangent space to the manifold M at the point $p \in M$, is the vector space $\mathrm{Der}_p(M)$ of derivations on $\mathscr{C}^\infty(M)$ in $p \in M$:*

$$\boxed{T_p^{\mathrm{alg}}M := \mathrm{Der}_p(M)}\,.$$

An element of $T_p^{\mathrm{alg}}M$, i.e. a derivation on $\mathscr{C}^\infty(M)$ in $p \in M$, will be called an algebraic tangent vector to M in p.

v is a linear functional, i.e. $v \in \mathscr{C}^\infty(M)^*$, the dual space of $\mathscr{C}^\infty(M)$ (interpreted as a real vector space). $\mathscr{C}^\infty(M)^*$ is an infinite-dimensional vector space, however, as we will show, the Leibniz property is such a strong constraint to be satisfied that the linear functionals that satisfy it, i.e. those composing the subspace $\mathrm{Der}_p(M) \subset \mathscr{C}^\infty(M)^*$, form a n-dimensional vector space, n being the dimension of M.

When we have discussed the case of geometric tangent vectors, we have proven an analogous dimensional reduction, in that case it was operated by the quotient w.r.t. the tangential equivalence between paths on the set of paths in M passing through a point. This is a first indication of the fact that geometric and algebraic tangent vectors are equivalent concepts.

Proving that the algebraic tangent space to a manifold at a point is a n-dimensional vector space is more difficult than for its geometric counterpart. Multiple proofs are available in the literature, the line of reasoning that we have chosen to follow in this document is not the shortest, but it has the advantage that the intermediate steps are fairly easy to prove:

1. first of all, we prove the result in the trivial case of $M = \mathbb{R}^n$;

2. then, we define the algebraic version of the differential (or push-forward) of a smooth function and analyze its remarkable properties;

3. by fusing the previous steps, the proof that $T_p^{\mathrm{alg}}M$ is (not canonically) isomorphic to $\mathbb{R}^n$ will be almost immediate.

To prove that the algebraic tangent space to $\mathbb{R}^n$, or an open subset of $\mathbb{R}^n$, at a point x_0 is isomorphic to a copy of $\mathbb{R}^n$, we need the following intermediate result, which says that every smooth function f on $\mathbb{R}^n$ is associated to a n-tuple of smooth functions that coincide with the partial derivatives of f in x_0 and, moreover, this n-tuple of smooth functions allows for a sort of first order expansion of f in a sufficiently small open neighborhood of x_0.

Lemma 2.2.2 *Let $x_0 = (x_0^1, \dots, x_0^n) \in \mathbb{R}^n$ and $f \in \mathscr{C}^\infty(\mathbb{R}^n)$, then there exist n smooth functions $g_1, \dots, g_n \in \mathscr{C}^\infty(V)$, where $V \subset \mathbb{R}^n$ is an open neighborhood of x_0, such that:*

$$g_j(x_0) = \frac{\partial f}{\partial x^j}(x_0)$$

and

$$f(x) = f(x_0) + \sum_{j=1}^{n} (x^j - x_0^j) g_j(x),$$

for all $x \in V$.

Proof. V can be considered as *star-shaped*, i.e. the straight line segment between x_0 and $x \in V$ defined by $x_0 + t(x - x_0)$ for all $t \in [0, 1]$ is entirely included in V; if it is not, then we can restrict it to a star-shaped open subset of $\mathbb{R}^n$ and work on this new neighborhood of x_0. Thanks to this remark, the expression $f(x_0 + t(x - x_0))$ is well-defined for all $x \in V$ and we can re-write the difference $f(x) - f(x_0)$ as follows:

$$f(x) - f(x_0) = [f(x_0 + t(x - x_0))]_{t=0}^{t=1} = \int_0^1 d(f(x_0 + t(x - x_0))),$$

thanks to the fundamental theorem of integral calculus. On the other side we have

$$\int_0^1 d(f(x_0 + t(x - x_0))) = \int_0^1 \frac{\partial}{\partial t} f(x_0 + t(x - x_0))dt,$$

we can expand the derivative under the integral by using the chain rule:

$$\frac{\partial}{\partial t} f(x_0+t(x-x_0)) = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}(x_0+t(x-x_0))\frac{\partial(x_0 + t(x^j - x_0))}{\partial t} = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}(x_0+t(x-x_0))(x^j - x_0)$$

so that

$$f(x) - f(x_0) = \int_0^1 \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}(x_0 + t(x - x_0))(x^j - x_0)dt = \sum_{j=1}^{n}(x^j - x_0) \int_0^1 \frac{\partial f}{\partial x^j}(x_0 + t(x - x_0))dt.$$

Since f is smooth, the integral exists and it is finite, and (since integration increases of one degree the regularity of the integrand) the functions g_j defined as follows:

$$g_j(x) = \int_0^1 \frac{\partial f}{\partial x^j}(x_0 + t(x - x_0))dt, \qquad \forall x \in V,$$

are smooth on V. Each g_j verifies both $f(x) = f(x_0) + \sum_{j=1}^{n}(x^j - x_0)g_j(x)$ and

$$g(x_0) = \int_0^1 \frac{\partial f}{\partial x^j}(x_0 + t(x_0 - x_0))dt = \int_0^1 \frac{\partial f}{\partial x^j}(x_0)dt = \frac{\partial f}{\partial x^j}(x_0) \int_0^1 dt = \frac{\partial f}{\partial x^j}(x_0),$$

thus proving the lemma. □

Theorem 2.2.2 *Let $V \subseteq \mathbb{R}^n$ be an open set and $x_0 \in V$, then the following map is an isomorphism of vector spaces:*

$$\begin{array}{rccc} \iota : & \mathbb{R}^n & \xrightarrow{\sim} & T^{\mathrm{alg}}_{x_0}V \\ & v = (v^j) & \longmapsto & \iota(v) := \sum_{j=1}^{n} v^j \left.\frac{\partial}{\partial x^j}\right|_{x_0} \equiv \left. D_v\right|_{x_0}, \end{array}$$

where the derivation $\left. D_v\right|_{x_0} : \mathscr{C}^\infty(V) \to \mathbb{R}$ is nothing but the linear functional on $\mathscr{C}^\infty(V)$ that, when applied to a smooth scalar function f on V, provides its directional derivative along v in x_0:

$$\left. D_v\right|_{x_0}(f) = D_v f(x_0) = \sum_{j=1}^{n} v^j \frac{\partial f}{\partial x^j}(x_0).$$

In particular, $T^{\mathrm{alg}}_{x_0}V$ is a n-dimensional vector space.

Proof. The fact that ι is linear can be checked directly and it follows from the linearity of the operations involved in its definition.

ι is one-to-one: since ι is linear, to prove that it is injective we simply have to check that $\ker(\iota) = \{0_{\mathbb{R}^n}\}$. For that, it is sufficient to show that, if $v = (v^j) \neq 0_{\mathbb{R}^n}$, i.e. at least one component is non null, say, $v^k \neq 0$, then the corresponding derivation $\iota(v)$ is not the null derivation, i.e. the derivation that sets all smooth scalar functions on V to 0.

To verify that, it is enough to consider the k-th canonical element of the dual basis of $\mathbb{R}^n$, i.e. $\varepsilon^k : V \to \mathbb{R}$, $\varepsilon^k(x) := x^k$. Of course $\varepsilon^k \in \mathscr{C}^\infty(V)$ because the projection on the k-th axis is smooth, so we can apply $\iota(v)$ to it, obtaining:

$$\iota(v)(\varepsilon^k) := \sum_{j=1}^n v^j \left.\frac{\partial}{\partial x^j}\right|_{x_0} \varepsilon^k = \sum_{j=1}^n v^j \frac{\partial \varepsilon^k(x_0)}{\partial x^j} = \sum_{j=1}^n v^j \frac{\partial x_0^k}{\partial x^j} = \sum_{j=1}^n v^j \delta_j^k = v^k \neq 0,$$

and so ι is one-to-one (note that x_0 is not a constant, but a variable, for this reason $\frac{\partial x_0^k}{\partial x^j} = \delta_k^j$).
<u>ι is onto</u>: we must show that, for every $D \in \mathrm{Der}_p(V)$, it exists a vector $v = (v^j) \in \mathbb{R}^n$ such that $D = \iota(v) = \sum\limits_{j=1}^n v^j \left.\frac{\partial}{\partial x^j}\right|_{x_0}$. To this aim, we use the previous lemma, expanding an arbitrary $f \in \mathscr{C}^\infty(V)$ as follows:

$$f(x) = f(x_0) + \sum_{j=1}^n (x^j - x_0^j) g_j(x)$$

in a neighborhood of x_0 inside V. This expression can be re-written as a functional equation, namely:

$$f = k_{f(x_0)} + \sum_{j=1}^n (\varepsilon^j - k_{x_0^j}) g_j,$$

where $k_{f(x_0)}(x) \equiv f(x_0)$ and $k_{x_0^j}(x) \equiv x_0^j$ are constant functions, and $\varepsilon^j(x) = x^j$. Applying D on f we get, by linearity, $D(f) = \cancelto{0}{D(k_{f(x_0)})} + \sum\limits_{j=1}^n D\left((\varepsilon^j - k_{x_0^j}) g_j\right)$, having used the fact that a derivation sets to 0 constant functions. Now, by using Leibniz's rule:

$$D(f) = \sum_{j=1}^n \left[D(\varepsilon^j - k_{x_0^j}) g_j(x_0) + \cancelto{0}{(\varepsilon^j - k_{x_0^j})(x_0)} \, D g_j \right],$$

where the second term between square brackets vanishes because $(\varepsilon^j - k_{x_0^j})(x_0) = \varepsilon^j(x_0) - k_{x_0^j}(x_0) = x_0^j - x_0^j = 0$. So, using again the linearity of D, the nullification of constant functions and the fact that $g_j(x_0) = \frac{\partial f}{\partial x^j}(x_0)$ (thanks to the previous lemma), we have:

$$D(f) = \sum_{j=1}^n (D(\varepsilon^j) - \cancelto{0}{D(k_{x_0^j})}) g_j(x_0) = \sum_{j=1}^n D(\varepsilon^j) g_j(x_0) = \sum_{j=1}^n D(\varepsilon^j) \frac{\partial f}{\partial x^j}(x_0),$$

since f is arbitrary, we can write $D = \sum\limits_{j=1}^n D(\varepsilon^j) \left.\frac{\partial}{\partial x^j}\right|_{x_0}$, thus, to obtain $D = \iota(v)$ we simply have to consider the vector $v = (v^j) \in \mathbb{R}^n$ whose components satisfy:

$$\boxed{v^j := D(\varepsilon^j)}, \qquad j = 1, \ldots, n. \tag{2.6}$$

□

Corollary 2.2.1 *For any fixed $x_0 \in \mathbb{R}^n$, the n derivations on $\mathscr{C}^\infty(\mathbb{R}^n)$ given by*

$$\left(\frac{\partial}{\partial x^1}\bigg|_{x_0}, \dots, \frac{\partial}{\partial x^n}\bigg|_{x_0} \right) \equiv \left(D_{e_1}|_{x_0}, \dots, D_{e_n}|_{x_0} \right)$$

form a basis of $T^{\text{alg}}_{x_0}\mathbb{R}^n$.

Proof. Almost immediate: since the linear isomorphism ι of the previous theorem maps basis to basis, if we apply it to $(e_1, \dots, e_n)$, the canonical basis of $\mathbb{R}^n$, we obtain a basis of $T^{\text{alg}}_{x_0}\mathbb{R}^n$. Since the components of the canonical basis elements are all 0 except for only one, the images of $(e_1, \dots, e_n)$ are exactly the evaluation in x_0 of the directional derivatives along each Cartesian axis, i.e. $ev_{x_0} \circ \frac{\partial}{\partial x^j} \equiv \frac{\partial}{\partial x^j}\big|_{x_0}$, $j = 1, \dots, n$. $\square$

As a very special case, when $n = 1$, we have that the only element of the canonical basis of $\mathbb{R}$, i.e. 1, can be identified with the derivative w.r.t. x in x_0:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\sim} & T^{\text{alg}}_{x_0}\mathbb{R} \\ 1 & \longmapsto & \frac{d}{dx}\big|_{x_0} . \end{array}$$

2.2.1 The (algebraic) differential of a smooth function between manifolds

As we have already seen in the case of geometric tangent vectors, every smooth map f between manifolds M and N can be 'lifted' to a linear map between the tangent spaces of M and N called either differential, tangent map or (point-wise) push forward.

Here we provide the definition of differential when the tangent spaces are defined algebraically. Its properties will prove to be of fundamental importance.

Def. 2.2.3 (Differential of a smooth function – algebraic case) *Let $f : M \to N$ be a smooth function and $p \in M$, the differential of f in p is the linear function defined in this way:*

$$\begin{array}{rccc} df_p : & T^{alg}_p M & \longrightarrow & T^{alg}_{f(p)} N \\ & v & \longmapsto & df_p(v), \end{array}$$

where $df_p(v)$ is the derivation at $f(p)$ defined as follows:

$$\begin{array}{rccc} df_p(v) : & \mathscr{C}^\infty(N) & \longrightarrow & \mathbb{R} \\ & \phi & \longmapsto & \boxed{df_p(v)(\phi) = v(\phi \circ f)} . \end{array} \tag{2.7}$$

The composition between a scalar function with a map between manifolds appears often in differential geometry, for this reason it bears a special name and symbol.

Def. 2.2.4 (Pull-back of scalar functions) *Let $f : M \to N$ be a smooth function and $\phi : N \to \mathbb{R}$ a scalar function on N. Then, we can define a scalar function on M simply by composition with f:*

$$\begin{array}{rccc} f^* : & \mathscr{C}^\infty(N) & \longrightarrow & \mathscr{C}^\infty(M) \\ & \phi & \longmapsto & \boxed{f^*(\phi) = \phi \circ f} . \end{array}$$

f^* is called the pull-back via f because it pulls-back a scalar function on N, the codomain of f, to a scalar function on M, the domain of f. Of course, $(id_M)^*(\phi) = \phi$ for all $\phi \in \mathscr{C}^\infty(M)$, so:

$$(id_M)^*(\phi) = id_{\mathscr{C}^\infty(M)}. \tag{2.8}$$

Note that, if $f \in \mathscr{C}^\infty(M, N)$, $g \in \mathscr{C}^\infty(N, P)$ and $\phi \in \mathscr{C}^\infty(P)$, then $(g \circ f)^*(\phi) := \phi \circ (g \circ f)$, but $(f^* \circ g^*)(\phi) = f^*(g^*(\phi)) = f^*(\phi \circ g) = \phi \circ (g \circ f)$, thus

$$\boxed{(g \circ f)^* = f^* \circ g^*}\ . \tag{2.9}$$

When using the pull-back, the differential of a smooth map becomes:

$$df_p(v)(\phi) = (v \circ f^*)(\phi) \qquad \Longleftrightarrow \qquad df_p(v)(\phi) = v(f^*(\phi)), \qquad \forall \phi \in \mathscr{C}^\infty(N),$$

or, since the previous equation holds for every $\phi \in \mathscr{C}^\infty(N)$,

$$df_p(v) = v \circ f^* \qquad \Longleftrightarrow \qquad df_p(v) = v(f^*). \tag{2.10}$$

If we use the push forward notation (in which the point p is omitted) to push a tangent vector to M at p towards a tangent vector to N at $f(p)$, then we get the evocative expression below:

$$\begin{array}{rccl} f_* : & T_p^{\text{alg}} M & \longrightarrow & T_{f(p)}^{\text{alg}} N \\ & v & \longmapsto & \boxed{f_*(v) = v(f^*)}\ . \end{array}$$

The principal properties of the differential are listed in the following proposition.

Theorem 2.2.3 (Properties of the differential) *For all $p \in M$ the following properties hold.*

1. $d(id_M)_p = id_{T_p^{alg} M}$;

2. ***Chain rule for differential:*** *if $f \in \mathscr{C}^\infty(M, N)$ and $g \in \mathscr{C}^\infty(N, P)$, then the differential of the composed function $g \circ f : M \to P$ is the linear map $d(g \circ f)_p : T_p^{alg} M \to T_{g(f(p))}^{alg} P$ such that:*

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

3. *If $U \subseteq M$ is an open set containing p and $\iota : U \hookrightarrow M$ is the canonical inclusion in M, then $d\iota_p : T_p^{alg} U \to T_p^{alg} M$ is a canonical linear isomorphism.*

4. *If f is a* <u>local</u> *diffeomorphism defined on an open subset $U \subseteq M$ with values in $f(U) \subseteq N$, then $df_p : T_p^{alg} M \to T_{f(p)}^{alg} N$ is a (globally defined) linear isomorphism and*

$$(df_p)^{-1} = d(f^{-1})_{f(p)}. \tag{2.11}$$

Proof.

1. By (2.10) we get $d(id_M)_p(v) = v \circ (id_M)^*$ for all $v \in T_p^{\text{alg}} M$, moreover, thanks to (2.8) we have $(id_M)^* = id_{\mathscr{C}^\infty(M)}$, thus $d(id_M)_p(v) = v$, i.e. $d(id_M)_p = id_{T_p^{\text{alg}} M}$.

2. Let $v \in T_p^{\mathrm{alg}}M$, arbitrary, then:

$$\begin{aligned} d(g\circ f)_p(v) &\underset{(2.10)}{=} v\circ(g\circ f)^* \underset{(2.9)}{=} v\circ(f^*\circ g^*) = (v\circ f^*)\circ g^* \\ &\underset{(2.10)}{=} df_p(v)\circ g^* \underset{(2.10)}{=} dg_{f(p)}(df_p(v)) \\ &= (dg_{f(p)}\circ df_p)(v). \end{aligned}$$

3. We will prove injectivity and surjectivity of $d\iota_p$.

Injectivity: we must show that the kernel of the linear map $d\iota_p$ is reduced to the zero derivation. For that, let us consider an arbitrary $v \in T_p^{\mathrm{alg}}U$ and suppose that $d\iota_p(v) = 0$, we must show that this implies $v = 0$. To this aim, let B be an open neighborhood of p such that $\overline{B} \subseteq U$, then the extension theorem for smooth function (th. 1.5.2) assures us that any $f \in \mathscr{C}^\infty(U)$ can be extended to $\tilde{f} \in \mathscr{C}^\infty(M)$ in such a way that $\tilde{f}\big|_{\overline{B}} \equiv f|_{\overline{B}}$. This implies that f and $\tilde{f}\big|_U$ are smooth functions on U that agree on B, which is an open neighborhood of p, but then theorem 2.2.1 implies $v(f) = v(\tilde{f}\big|_U)$. Now, $\tilde{f}\big|_U$ is nothing but $\tilde{f}\circ\iota$, so

$$v(f) = v(\tilde{f}\circ\iota) \underset{(2.7)}{=} d\iota_p(v)(\tilde{f}) = 0,$$

because, by hypothesis, $d\iota_p(v) = 0$, the null derivation. Since $f \in \mathscr{C}^\infty(U)$ is arbitrary, $v = 0$ and so $d\iota_p$ is injective.

Surjectivity: consider an arbitrary $w \in T_p^{\mathrm{alg}}M$, we must prove that it exists $v \in T_p^{\mathrm{alg}}U$ such that $w = d\iota_p(v)$. We define such derivation as follows: $v(f) := w(\tilde{f})$ where $\tilde{f}$ in any smooth function on M such that $\tilde{f}\big|_{\overline{B}} = f|_{\overline{B}}$.

Thanks to theorem 2.2.1, this definition of v does not depend on the choice of $\tilde{f}$ and it is of course a derivation of $\mathscr{C}^\infty(U)$ at p, thanks to the fact that w is linear and verifies the Leibniz rule. Finally, fixed any arbitrary function $g \in \mathscr{C}^\infty(M)$, we have that g, $g\circ\iota$ and $\widetilde{g\circ\iota}$ agree on B, thus:

$$d\iota_p(v)(g) \underset{(2.7)}{=} v(g\circ\iota) := w(\widetilde{g\circ\iota}) = w(g),$$

since g is arbitrary, we have that $w = d\iota_p(v)$ and so $d\iota_p$ is also surjective.

4. It is an easy consequence of the previous points. In fact, if f is a local diffeomorphism between U and $f(U)$, then it exists $f^{-1} : f(U) \to U$, such that $f^{-1}\circ f = id_U$, thus:

$$d(id_U)_p = d(f^{-1}\circ f)_p \underset{(2.)}{=} d(f^{-1})_{f(p)}\circ df_p.$$

On the other side, thanks to 1., $d(id_U)_p = id_{T_p^{\mathrm{alg}}U}$ and, thanks to 3., $T_p^{\mathrm{alg}}U \cong T_p^{\mathrm{alg}}M$, thus $d(id_U)_p = id_{T_p^{\mathrm{alg}}M}$. So, equating the two expressions for $d(id_U)_p$ that we have determined, we find $d(f^{-1})_{f(p)}\circ df_p = id_{T_p^{\mathrm{alg}}M}$. Exchanging f with f^{-1} we get, with analogous considerations, $df_p\circ d(f^{-1})_{f(p)} = id_{T_p^{\mathrm{alg}}N}$, thus proving 4. □

Property 3. allows us to identify in a canonical way the tangent space at a point to an open neighborhood of p with the tangent space at the same point to the whole manifold: the derivation $d\iota_p(v)$ is the same derivation as v in p acting on smooth scalar functions defined on the whole manifold M instead of those defined on U.

This is not surprising at all, since, as proven in proposition 2.2.1, the action of a derivation in a given point on a scalar function depends only on the values of the function in an arbitrarily small neighborhood of that point. From now on, we will implicitly accept the following natural identification:

$$\boxed{T_p^{\text{alg}}U \cong T_p^{\text{alg}}M} \qquad U \subseteq M,\ U \text{ open.}$$

As previously stated, thanks to $T_x^{\text{alg}}\mathbb{R}^n \cong \mathbb{R}^n$ for all $x \in \mathbb{R}^n$ and to the properties of the differential, we can very easily prove the isomorphism between $T_p^{\text{alg}}M$ and $\mathbb{R}^n$ just by considering the differential of an arbitrary chart map.

Theorem 2.2.4 *If M is a manifold of dimension n, then, fixed any $p \in M$, it exists a non-canonical linear isomorphism of vector spaces such that:*

$$\boxed{T_p^{\text{alg}}M \cong \mathbb{R}^n}$$

so, in particular, $\dim(T_p^{\text{alg}}M) = n$.

Proof. If (U, φ) is an arbitrary chart in p such that $\varphi(p) = x$, then $\varphi : U \subseteq M \to \varphi(U) \subseteq \mathbb{R}^n$ is a local diffeomorphism. By property 4. of the differential, $d\varphi_p : T_p^{\text{alg}}M \to T_x^{\text{alg}}\mathbb{R}^n$ is a linear isomorphism of vector spaces. Since this isomorphism depends on the chart φ, it is not canonical. $\square$

2.2.2 A basis for $T_p^{\text{alg}}M$

Since $T_p^{\text{alg}}M$ is a n-dimensional vector space, it is natural to search for an explicit basis of tangent vectors.

In proposition 2.2.1 we have seen that, in the identification between $\mathbb{R}^n$ and $T_x^{\text{alg}}\mathbb{R}^n$, the canonical basis of $\mathbb{R}^n$ is identified with the basis of evaluations in x of the partial derivatives:

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\sim} & T_x^{\text{alg}}\mathbb{R}^n \\ e_j & \longleftrightarrow & \frac{\partial}{\partial x^j}\big|_x , \end{array} \qquad \forall j \in \{1, \dots, n\}.$$

Now, once selected a point $p \in M$ and a local chart (U, φ) in p such that $\varphi(p) = x \in \mathbb{R}^n$, we have just seen that the differential of φ in p is a linear isomorphism between $T_p^{\text{alg}}M$ and $T_p^{\text{alg}}\mathbb{R}^n \cong \mathbb{R}^n$, thus its inverse $(d\varphi_p)^{-1} : T_p^{\text{alg}}\mathbb{R}^n \cong \mathbb{R}^n \to T_p^{\text{alg}}M$ is a linear isomorphism too and, as such, it maps bases to bases.

As a consequence, **we can use $(d\varphi_p)^{-1}$ to transport the canonical basis of $\mathbb{R}^n$ (or, equivalently, the basis of $T_x^{\text{alg}}\mathbb{R}^n$ given by the evaluations in x of the partial derivatives) to a basis of $T_p^{\text{alg}}M$.**

So, for all $j = 1, \dots, n$:

$$\begin{array}{rccc} (d\varphi_p)^{-1} : & \mathbb{R}^n \cong T_x^{\text{alg}}\mathbb{R}^n & \xrightarrow{\sim} & T_p^{\text{alg}}M \\ & (e_j) \cong \left(\frac{\partial}{\partial x^j}\big|_x \right) & \longleftrightarrow & (d\varphi_p)^{-1} \left(\frac{\partial}{\partial x^j}\big|_x \right), \end{array}$$

the explicit action of the derivation $(d\varphi_p)^{-1}\left(\frac{\partial}{\partial x^j}\big|_x\right)$ on an arbitrary smooth scalar function $f \in \mathscr{C}^\infty(M)$ can be computed thanks to eq. (2.11), that in this case gives $(d\varphi_p)^{-1} = d(\varphi^{-1})_{\varphi(p)} = d(\varphi^{-1})_x$, so that:

$$(d\varphi_p)^{-1}\left(\left.\frac{\partial}{\partial x^j}\right|_x\right)(f) = d(\varphi^{-1})_x\left(\left.\frac{\partial}{\partial x^j}\right|_x\right)(f) := \left.\frac{\partial}{\partial x^j}\right|_x (f\circ\varphi^{-1}) \equiv \frac{\partial(f\circ\varphi^{-1})}{\partial x^j}(x),$$

but $f\circ\varphi^{-1} : \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ is nothing but **the local expression of f w.r.t. the chart** (U,φ) and the real numbers $\frac{\partial(f\circ\varphi^{-1})}{\partial x^j}(x)$, $j = 1,\dots,n$, represent the value of the directional derivatives of $f\circ\varphi^{-1}$ in the point $x \in \mathbb{R}^n$ along the unit canonical basis vectors e_j of $\mathbb{R}^n$.

The derivations $(d\varphi_p)^{-1}\left(\frac{\partial}{\partial x^j}\big|_x\right)_{j=1}^n$ constitute a basis of $T_p^{\text{alg}}M$ and, to simplify the heavy notation, they are usually written as follows:

$$\partial_j\big|_p \equiv \left.\frac{\partial}{\partial x^j}\right|_p = (d\varphi_p)^{-1}\left(\left.\frac{\partial}{\partial x^j}\right|_x\right) = d(\varphi^{-1})_x\left(\left.\frac{\partial}{\partial x^j}\right|_x\right), \quad x = \varphi(p). \tag{2.12}$$

We resume what just discussed in the following theorem.

Theorem 2.2.5 (Coordinate tangent vectors to M at p) *Fixed $p \in M$ and a local chart (U,φ) in it such that $\varphi(p) = x$, the derivations of $T_p^{\text{alg}}M$ given by $(\partial_j|_p)_{j=1}^n$, or $\left(\frac{\partial}{\partial x^j}\big|_p\right)_{j=1}^n$, defined by:*

$$\begin{array}{rccl} \partial_j\big|_p : & \mathscr{C}^\infty(M) & \longrightarrow & \mathbb{R} \\ & f & \longmapsto & \boxed{\partial_j\big|_p(f) = \dfrac{\partial(f\circ\varphi^{-1})}{\partial x^j}(x)}\,, \end{array} \tag{2.13}$$

or,

$$\begin{array}{rccl} \frac{\partial}{\partial x^j}\big|_p : & \mathscr{C}^\infty(M) & \longrightarrow & \mathbb{R} \\ & f & \longmapsto & \boxed{\left.\dfrac{\partial}{\partial x^j}\right|_p(f) = \dfrac{\partial(f\circ\varphi^{-1})}{\partial x^j}(x)}\,. \end{array} \tag{2.14}$$

form a basis of $T_p^{\text{alg}}M$. They are called coordinate tangent vectors to M at p.

Both notations are further simplified by writing:

$$\partial_j\big|_p(f) \equiv \partial_j f\big|_p = \frac{\partial(f\circ\varphi^{-1})}{\partial x^j}(x) \quad \text{and} \quad \left.\frac{\partial}{\partial x^j}\right|_p(f) \equiv \left.\frac{\partial f}{\partial x^j}\right|_p = \frac{\partial(f\circ\varphi^{-1})}{\partial x^j}(x).$$

We remark again that **the real value obtained by applying th j-th coordinate tangent vector to M at p on a smooth function f on M is just the value of the partial derivative of the local expression of f** (and not of f itself!) **along the j-th axis in** $x = \varphi(p)$.

This is the reason why the expression $\left.\frac{\partial f}{\partial x^j}\right|_p$ **must not not be intepreted as the partial derivative of f in p in the usual sense**, because f is defined on M, not on $\mathbb{R}^n$! The notation $\partial_j f\big|_p$ may be used to avoid this misinterpretation, however, the notation $\left.\frac{\partial f}{\partial x^j}\right|_p$ has the advantage to make the chain rule 'visually easier' to handle, as we will see later.

The basis of coordinate tangent vectors will be **the key to understand the link between the algebraic definition of tangent vectors and the physicist's one**.

Remark: the derivations $(\partial_1|_p, \ldots, \partial_n|_p)$ are defined by applying the linear isomorphism $(d\varphi_p)^{-1}$ to the canonical basis of $\mathbb{R}^n$, so they cannot be considered as a canonical basis for $T_p^{\text{alg}}M$ (which does not exist), because **they depend on the choice of the local chart** φ *in* p! Different charts will produce, in general, different basis for $T_p^{\text{alg}}M$.

Moreover, **as** p **varies in** M**, the tangent spaces** T_pM, in spite of being isomorphic to $\mathbb{R}^n$, are not canonically isomorphic to each other and they **must be considered as different copies of** $\mathbb{R}^n$ **attached to each point** p **of the manifold** M.

2.2.3 Coordinate formula for the differential

Fixed a local chart $(U, \varphi \equiv (x^j))$ in $p \in M$ such that $\varphi(p) = x$, a tangent vector $v \in T_p^{\text{alg}}M$ can be written *uniquely* as a linear combination of the basis of coordinate tangent vectors as follows:

$$v = v^j\, \partial_j|_p \equiv v^j \left.\frac{\partial}{\partial x^j}\right|_p ,$$

the real numbers v^j, $j = 1, \ldots, n$ are called the **components** of v on the basis of coordinate tangent vectors of $T_p^{\text{alg}}M$.

The following result establishes the main properties of the components v^j.

Theorem 2.2.6 *With the notations of this section, it hold that:*

$$T_p^{\text{alg}}M \ni v = v^j \left.\frac{\partial}{\partial x^j}\right|_p \iff d\varphi_p(v) = v^j \left.\frac{\partial}{\partial x^j}\right|_x \in T_x^{\text{alg}}\mathbb{R}^n, \tag{2.15}$$

moreover, the components of v *are obtained by applying the derivation* v *to the local coordinate functions* $x^j = \varepsilon^j \circ \varphi : U \to \mathbb{R}$, *i.e.*

$$\boxed{v^j = v(x^j)}\ . \tag{2.16}$$

Proof.

$\boxed{\Longrightarrow}$: let $v = v^j \left.\frac{\partial}{\partial x^j}\right|_p \in T_p^{\text{alg}}M$, then the isomorphism $d\varphi_p$ allows us to obtain the tangent vector $d\varphi_p(v) \in T_x^{\text{alg}}\mathbb{R}^n$, whose action on smooth scalar functions $\phi \in \mathscr{C}^\infty(\mathbb{R}^n)$ is $d\varphi_p(v)(\phi) = v(\phi \circ \varphi)$, but, by definition of differential and by linearity, we have:

$$d\varphi_p\left(v^j \left.\frac{\partial}{\partial x^j}\right|_p\right)(\phi) = v^j \left.\frac{\partial}{\partial x^j}\right|_p (\phi \circ \varphi) \underset{(2.14)}{=} v^j \left.\frac{\partial(\phi \circ \varphi \circ \varphi^{-1})}{\partial x^j}\right|_x = v^j \frac{\partial \phi}{\partial x^j}(x) = v^j \left.\frac{\partial}{\partial x^j}\right|_x (\phi),$$

since this holds for all $\phi \in \mathscr{C}^\infty(\mathbb{R}^n)$, we have proven that $v = v^j \left.\frac{\partial}{\partial x^j}\right|_p \in T_p^{\text{alg}}M$ implies $d\varphi_p(v) = v^j \left.\frac{\partial}{\partial x^j}\right|_x \in T_x^{\text{alg}}\mathbb{R}^n$.

$\boxed{\Longleftarrow}$: suppose that $d\varphi_p(v) = v^j \left.\frac{\partial}{\partial x^j}\right|_x \in T_x^{\text{alg}}\mathbb{R}^n$, then, applying the inverse linear isomorphism $(d\varphi_p)^{-1} = d(\varphi^{-1})_x : T_x^{\text{alg}}\mathbb{R}^n \to T_p^{\text{alg}}M$ we get:

$$v = (d\varphi_p)^{-1}(d\varphi_p(v)) = v^j\, d(\varphi^{-1})_x \left(\left.\frac{\partial}{\partial x^j}\right|_x\right),$$

i.e. for all $\phi \in \mathscr{C}^\infty(U)$,

$$v(\phi) = v^j\, d(\varphi^{-1})_x \left(\frac{\partial}{\partial x^j}\bigg|_x \right)(\phi) = \frac{\partial}{\partial x^j}\bigg|_x (\phi \circ \varphi^{-1}) = v^j \frac{\partial(\phi \circ \varphi^{-1})}{\partial x^j}(x) \underset{(2.14)}{=} v^j\, \frac{\partial}{\partial x^j}\bigg|_p (\phi),$$

since this holds for all $\phi \in \mathscr{C}^\infty(U)$, we have proven that $d\varphi_p(v) = v^j\, \frac{\partial}{\partial x^j}\big|_x \in T_x^{\mathrm{alg}}\mathbb{R}^n$ implies $v = v^j\, \frac{\partial}{\partial x^j}\big|_p \in T_p^{\mathrm{alg}}M$.

Finally, since the coefficients v^j appear in two formulas, let us show (redundantly, by instructively) how to recover eq. (2.16) from both expressions. One strategy is to recall formula (2.6), which says that the real coefficients v^j are computed by applying the derivation $d\varphi_p(v)$ to the elements of the dual canonical basis of $\mathbb{R}^n$, i.e. $v^j = d\varphi_p(v)(\varepsilon^j) := v(\varepsilon^j \circ \varphi) = v(x^j)$.

Another strategy consists in the following brute force computation:

$$v(x^j) = v^k\, \frac{\partial}{\partial x^k}\bigg|_p (x^j) = v^k\, \frac{\partial(\varepsilon^j \circ \varphi \circ \varphi^{-1})}{\partial x^k}(x) = v^k\, \frac{\partial x^j}{\partial x^k} = v^k\, \delta^j_k = v^j.$$

□

2.2.4 Differential of scalar functions and curves

Two special cases must be examined in relation with the differential: the first is when f is a scalar function, so that its *codomain* is a subset of $\mathbb{R}$, the other is when f is a path, so that its *domain* is a subset of $\mathbb{R}$.

Differential of a scalar function

Let us start with the case of a scalar function $\phi \in \mathscr{C}(M)$. Since ϕ is already a scalar function, we do not need to resort to other auxiliary scalar functions as in definition 2.2.3 and we can simply write:

$$\begin{array}{rccl} d\phi_p: & T_p^{\mathrm{alg}}M & \longrightarrow & T_{\phi(p)}^{\mathrm{alg}}\mathbb{R} \cong \mathbb{R} \\ & v & \longmapsto & \boxed{d\phi_p(v) := v(\phi)}\ . \end{array} \tag{2.17}$$

Since $T_{\phi(p)}^{\mathrm{alg}}\mathbb{R}$ is a tangent space to $\mathbb{R}$ at a point and $v(\phi)$ is a real number, an explanation is needed to justify the previous definition. Note that $T_{\phi(p)}^{\mathrm{alg}}\mathbb{R} = \mathrm{span}\left(\frac{d}{dt}\big|_{\phi(p)}\right)$ and $\mathbb{R} = \mathrm{span}(1)$, thus $T_{\phi(p)}^{\mathrm{alg}}$ and $\mathbb{R}$ can be canonically identified via the following correspondence:

$$\begin{array}{ccc} T_{\phi(p)}^{\mathrm{alg}}\mathbb{R} & \overset{\sim}{\longrightarrow} & \mathbb{R} \\ \frac{d}{dt}\big|_{\phi(p)} & \longleftrightarrow & 1, \end{array}$$

so:

$$T_{\phi(p)}^{\mathrm{alg}}\mathbb{R} \ni v(\phi)\, \frac{d}{dt}\bigg|_{\phi(p)} \cong v(\phi)1 = v(\phi) \in \mathbb{R}.$$

It is custom to avoid specifying this canonical identification and to write the differential of a scalar function simply as in eq. (2.17).

Differential of a curve

Let us now consider $\gamma : (-\varepsilon, \varepsilon) \to M$ and $t_0 \in (-\varepsilon, \varepsilon)$. This time, via the identification $\mathbb{R} \simeq T_{t_0}\mathbb{R}$, we can identify t_0 with $\frac{d}{dt}\big|_{t_0}$, so that

$$\begin{array}{rccc} d\gamma_{t_0} : & T^{\text{alg}}_{t_0}\mathbb{R} \cong \mathbb{R} & \longrightarrow & T^{\text{alg}}_{\gamma(t_0)}M \\ & \frac{d}{dt}\big|_{t_0} & \longmapsto & d\gamma_{t_0}(\frac{d}{dt}\big|_{t_0}), \end{array}$$

where the action of $d\gamma_{t_0}(\frac{d}{dt}\big|_{t_0})$ on smooth functions on M is the canonical one for the differential, i.e.

$$\begin{array}{rccl} d\gamma_{t_0}(\frac{d}{dt}\big|_{t_0}) : & \mathscr{C}^\infty(M) & \longrightarrow & \mathbb{R} \\ & \phi & \longmapsto & d\gamma_{t_0}\left(\frac{d}{dt}\big|_{t_0}\right)(\phi) = \frac{d}{dt}\big|_{t_0}(\phi \circ \gamma) \equiv (\phi \circ \gamma)^\bullet(t_0). \end{array}$$

It us common to simplify the quite heavy notation as follows:

$$\boxed{d\gamma_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right) \equiv \dot{\gamma}(t_0)}\,, \tag{2.18}$$

so that, when it is applied to a scalar function $\phi \in \mathscr{C}^\infty(M)$ it verifies:

$$\boxed{\dot{\gamma}(t_0)(\phi) := (\phi \circ \gamma)^\bullet(t_0)}\,.$$

Def. 2.2.5 (Velocity vector of a curve at a point) *The tangent vector $\dot{\gamma}(t_0) \in T^{\text{alg}}_{\gamma(t_0)}M$ is called the velocity vector of γ at t_0.*

As we saw previously, if we want to find out the components of $\dot{\gamma}(t_0) \in T^{\text{alg}}_{\gamma(t_0)}M$ w.r.t. the coordinate tangent vectors to M at $\gamma(t_0)$, we must fix a local chart $(U, \varphi \equiv (x^i))$ in $p = \gamma(t_0) \in M$, such that $\gamma(-\varepsilon, \varepsilon) \subseteq U$ and then we have to apply $\dot{\gamma}(t_0)$ to the local coordinate functions $x^j = \varepsilon^j \circ \varphi$.

Since $\dot{\gamma}(t_0) = d\gamma_{t_0}\left(\frac{d}{dt}\big|_{t_0}\right)$, by the algebraic definition of differential we obtain

$$\dot{\gamma}(t_0)(x^j) = d\gamma_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right)(x^j) = \left.\frac{d}{dt}\right|_{t_0}(x^j \circ \gamma),$$

we notice that $\varphi \circ \gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ is a curve in $\mathbb{R}^n$ and $x^j \circ \gamma = \varepsilon^j \circ \varphi \circ \gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}$ are nothing but its n component functions which are usually indicated with γ^j, thus, the coordinate expression for the tangent vector of the curve γ in t_0 is:

$$\boxed{\dot{\gamma}(t_0) = \frac{d\gamma^j}{dt}(t_0)\left.\frac{\partial}{\partial x^j}\right|_{\gamma(t_0)}} \iff \boxed{\dot{\gamma}(t_0) = \dot{\gamma}^j(t_0)\,\partial_j\big|_{\gamma(t_0)}}\,, \quad \gamma^j \equiv \varepsilon^j \circ \varphi \circ \gamma. \tag{2.19}$$

We finish this section by proving a result which shows that velocity vectors behave as expected under composition with smooth maps.

Theorem 2.2.7 (Velocity vector of a composite curve) *Let $f : M \to N$ be a smooth map, $\gamma : (-\varepsilon, \varepsilon) \to M$ be a smooth curve in M and $f \circ \gamma : (-\varepsilon, \varepsilon) \to N$ the composite curve in N. The velocity vector of $f \circ \gamma$ at any $t_0 \in (-\varepsilon, \varepsilon)$ satisfies:*

$$(f \circ \gamma)^{\boldsymbol{\cdot}}(t_0) = df_{\gamma(t_0)}(\dot{\gamma}(t_0)).$$

Proof. By definition of velocity vector and thanks to the chain rule for the differential we have:

$$(f \circ \gamma)^{\boldsymbol{\cdot}}(t_0) = d(f \circ \gamma)_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right) = df_{\gamma(t_0)} \circ d\gamma_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right) = df_{\gamma(t_0)}(\dot{\gamma}(t_0)).$$

□

This seemingly innocent result has a very useful consequence: from left to right, it tells us how to compute the velocity vector of a composite curve via the differential. But, if read the other way round, it allows us **to compute the differential of a function in terms of the velocity vector of a curve**! Let us see under which condition this is true: given a smooth function $f : M \to N$ and a point $p \in M$, to compute $df_p(v)$, $v \in T_p^{\mathrm{alg}}M$ with this technique we need a curve γ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = v$.

If such γ exists then, by using the previous result, the computation of the differential of f in p can be performed in terms of velocity vector of the composite curve $f \circ \gamma$ as follows:

$$df_p(v) = (f \circ \gamma)^{\boldsymbol{\cdot}}(0), \quad v = \dot{\gamma}(0). \tag{2.20}$$

Actually, we are going to prove that the condition that we have pointed out is always verified. This result has a major importance also because it provides the bridge between the geometric and the algebraic definition of tangent vectors in differential geometry.

2.2.5 Equivalence between geometric and algebraic tangent vectors

We can finally prove that the definition of geometric and algebraic tangent vectors to a manifold at a point are completely equivalent.

Theorem 2.2.8 *Let $p \in M$ and let γ be a curve in M passing through p, i.e. $\gamma(0) = p$. Then, the map*

$$\begin{array}{rccl} I: & T_p^{\mathrm{geom}}M & \xrightarrow{\sim} & T_p^{\mathrm{alg}}M \\ & [\gamma] & \longmapsto & I[\gamma] := \dot{\gamma}(0), \end{array}$$

*where $\dot{\gamma}(0) \equiv d\gamma_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right)$ is the velocity vector of any $\gamma \in [\gamma]$, is an isomorphism of vector spaces. Thus, **all tangent vectors to a manifold at a point are the velocity vector of a curve passing through that point**.*

Proof. First of all, let us prove that I is <u>well-defined</u>. Consider $\gamma_1, \gamma_2 \in [\gamma]$, we must verify that $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$. To this aim, consider a local chart $(U, \varphi \equiv (x^j))$ in $p = \gamma(0)$ and the local coordinate expressions of γ_1 and γ_2 given by $\gamma_1^j := \varepsilon^j \circ \varphi \circ \gamma_1$ and $\gamma_2^j := \varepsilon^j \circ \varphi \circ \gamma_2$. Then, by using the coordinate expression of the velocity vector, eq. (2.19), we get:

$$\dot{\gamma}_1(0) = \frac{d\gamma_1^j}{dt}(0)\left.\frac{\partial}{\partial x^j}\right|_p \quad \text{and} \quad \dot{\gamma}_2(0) = \frac{d\gamma_2^j}{dt}(0)\left.\frac{\partial}{\partial x^j}\right|_p .$$

γ_1 and γ_2 belong to the same tangentially equivalence class of curves (cfr. section 2.1), thus, by definition, $(\varphi \circ \gamma_1)^\cdot(0) = (\varphi \circ \gamma_2)^\cdot(0)$, i.e. $\frac{d\gamma_1^j}{dt}(0) = \frac{d\gamma_2^j}{dt}(0)$, for all $j = 1, \ldots, n$, since these values are nothing but the components of $(\varphi \circ \gamma_1)^\cdot(0)$ and $(\varphi \circ \gamma_2)^\cdot(0)$, respectively. It follows that $\dot\gamma_1(0)$ and $\dot\gamma_2(0)$ have the same decomposition on the coordinate tangent vector basis, hence, by the uniqueness of this decomposition, $\dot\gamma_1(0) = \dot\gamma_2(0)$.

This argument also shows that I is injective: if $[\gamma] \neq [\sigma]$, $\sigma(0) = p$, then, by definition, $\dot\gamma(0) \neq \dot\sigma(0)$. The linearity of I can be verified by direct computation and follows easily from the linearity of $\dot\gamma(0)$.

The only property that remains to be checked in the surjectivity of I, i.e. that for every $v \in T_p^{\text{alg}}M$, $v = v^j \left.\frac{\partial}{\partial x^j}\right|_p$, it exists $[\gamma] \in T_p^{\text{geom}}M$ such that $I([\gamma]) = v$. We have already proven that I is well-defined, thus we can concentrate just on searching a representative curve $\gamma : (-\varepsilon, \varepsilon) \to M$ such that $\gamma(0) = p$ and $\dot\gamma(0) = v$, i.e. $\frac{d\gamma^j}{dt}(0) = v^j$ for all $j = 1, \ldots, n$.

To solve this problem we take inspiration from eq. (2.2) and we define the curve

$$\begin{array}{rccl} \gamma: & (-\varepsilon, \varepsilon) & \longrightarrow & U \\ & t & \longmapsto & \gamma(t) = \varphi^{-1}(x + t(v^1, \ldots, v^n)), \quad x = \varphi(p), \end{array}$$

which satisfies $\gamma(0) = p$ and, $\forall t \in (-\varepsilon, \varepsilon)$, $\forall j = 1, \ldots, n$:

$$\gamma^j(t) = (\varepsilon^j \circ \varphi \circ \gamma)(t) = \varepsilon^j(\varphi(\varphi^{-1}(x + t(v^1, \ldots, v^n))) = \varepsilon^j(x + t(v^1, \ldots, v^n)) = x^j + tv^j.$$

Finally, thanks to eq. (2.19), we get:

$$\dot\gamma(0) = \frac{d\gamma^j}{dt}(0) \left.\frac{\partial}{\partial x^j}\right|_p = \frac{d(x^j + tv^j)}{dt}(0) \left.\frac{\partial}{\partial x^j}\right|_p = v^j \left.\frac{\partial}{\partial x^j}\right|_p = v.$$

□

Starting from now, we will drop the specification 'geom' and 'alg' from the notation of tangent space and we will **write simply T_pM for the tangent space to M at p.**

It will be clear from the context which kind of vector we are considering and, in any case, we know how to pass from one to the other and vice-versa. In particular, we have made the observations that led to eq. (2.20) rigorous and we can resume them in the following proposition.

Theorem 2.2.9 *Let $f \in \mathscr{C}^\infty(M, N)$ and $p \in M$. Let also $(U, \varphi \equiv (x^j))$ be a local chart in p such that $\varphi(p) = x \in \mathbb{R}^n$. If $v \in T_pM$ has the following local coordinate expression $v = v^j \left.\frac{\partial}{\partial x^j}\right|_p$ w.r.t. this local chart, then it holds that:*

$$\boxed{df_p(v) = (f \circ \gamma)^\cdot(0)}\,, \tag{2.21}$$

with $\gamma(t) = \varphi^{-1}(x + t(v^1, \ldots, v^n))$, for all $t \in \mathbb{R}$ such that $\gamma(t) \in U$.

In particular, the class of tangentially equivalent paths that are in one-to-one correspondence with the coordinate tangent vectors $\left.\frac{\partial}{\partial x^j}\right|_p$ is:

$$\left.\frac{\partial}{\partial x^j}\right|_p \cong [t \mapsto \varphi^{-1}(x + te^j)], \qquad x = \varphi(p),$$

where e^j is the j-th element of the canonical basis of $\mathbb{R}^n$. This result confirms what we have already established in eq. (2.4) and underlines once more that the tangent vectors $\frac{\partial}{\partial x^j}\big|_p$ are locally defined, they depend on the choice of the coordinate system defined by the chart (U, φ) and they are associated to the vectors of the canonical basis of $\mathbb{R}^n$.

In the particular case $M = \mathbb{R}^n$ or of a real vector space V we have global charts and we can state the previous result as follows.

Corollary 2.2.2 *Let V, W be two finite dimensional real vector spaces, $f \in \mathscr{C}^\infty(V, W)$ and $x \in V$. Then it holds that:*

$$\boxed{df_x(v) = \frac{d}{dt}\bigg|_{t=0} f(x + tv)}\,, \qquad \forall v \in T_x V. \tag{2.22}$$

2.2.6 Relationship between the differential and the total derivative on vector spaces

We are now going to show that the differential of a function $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, Ω open, coincides with its total derivative in the sense defined in Appendix B.

Suppose $x \in \Omega$, then, by proposition 2.2.1, we can write any $v \in T_x\Omega \cong T_x\mathbb{R}^n$ as $v = v^j\, \frac{\partial}{\partial x^j}\big|_x$. By definition of differential of a scalar function, i.e. (2.17), we get:

$$df_x(v) = v(f) = v^j\, \frac{\partial}{\partial x^j}\bigg|_x (f) = v^j \frac{\partial f}{\partial x^j}(x) = D_v f(x),$$

where $D_v f(x)$ is the directional derivative of f in the direction defined by v, identified with a vector of $\mathbb{R}^n$ thanks to the canonical isomorphism $T_x\mathbb{R}^n \cong \mathbb{R}^n$.

However, in Appendix B it is proven that the $D_v f(x)$ is obtained by applying the total derivative of f in x to the vector v: $Df(x)(v) = D_v f(x)$ and this holds for all $v \in \mathbb{R}^n$.

As a consequence, **we can canonically identify the differential of a scalar function defined on an open $\Omega \subseteq \mathbb{R}^n$ at any point with its total derivative in the same point**:

$$\boxed{df_x = Df(x)}\,, \qquad \forall f \in \mathscr{C}^\infty(\Omega).$$

The same identification holds for functions as $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, Ω open: as always, one considers the component functions of $f = (f^1, \dots, f^m)$, that are scalar functions to which one can apply the result just proven.

We will use this result to compute some remarkable differentials in section 2.9.

2.3 Matrix expression of the differential in coordinates

$df_p : T_pM \to T_{f(p)}N$ is a linear operator between finite dimensional vector spaces, thus we can represent it as a matrix. To understand how to do it, we first examine the trivial case of $M = \mathbb{R}^m$ and $N = \mathbb{R}^n$.

Given $f : U \subseteq \mathbb{R}^m \to V \subseteq \mathbb{R}^n$, U open and f smooth, once we fix any $x \in U$, we have just seen that the differential operator $df_x : \mathbb{R}^n \to \mathbb{R}^m$ coincides with the total derivative of f in x, which is represented in matrix form by the Jacobian matrix of f in x. It is an instructive exercise to explicitly verify that this is actually the case.

If we denote with $(x^i)_{i=1}^m$ and $(y^j)_{j=1}^n$ the coordinates in U and V respectively, then the coordinate tangent vectors $\left(\frac{\partial}{\partial x^i}\Big|_x\right)_{i=1}^m$ and $\left(\frac{\partial}{\partial y^j}\Big|_{f(x)}\right)_{j=1}^n$ form a basis of $T_x\mathbb{R}^m$ and $T_{f(x)}\mathbb{R}^n$, respectively.

To find the matrix expression of df_x w.r.t. these bases we know that we must apply df_x to the vectors of the first basis and the express the results as a linear combination of the vectors of the second basis. The coefficients of this combination are the columns of the matrix that represents df_x w.r.t. the chosen bases.

Note that $df_x\left(\frac{\partial}{\partial x^i}\Big|_x\right) \in T_{f(x)}V$, thus it is a derivation on $\mathscr{C}^\infty(V)$, so, to make its action explicit, we have to fix an arbitrary smooth scalar function $g \in \mathscr{C}^\infty(V)$ and write:

$$df_x\left(\frac{\partial}{\partial x^i}\bigg|_x\right)(g) := \frac{\partial}{\partial x^i}\bigg|_x (g \circ f) \underset{\text{(chain rule)}}{=} \frac{\partial g}{\partial y^j}(f(x))\frac{\partial f^j}{\partial x^i}(x) = \frac{\partial f^j}{\partial x^i}(x)\frac{\partial g}{\partial y^j}(f(x)),$$

re-writing conveniently $\frac{\partial g}{\partial y^j}(f(x)) = \frac{\partial}{\partial y^j}\Big|_{f(x)}(g)$ to make the coordinate tangent vector basis of $T_{f(x)}\mathbb{R}^n$ appear explicitly, we get:

$$df_x\left(\frac{\partial}{\partial x^i}\bigg|_x\right)(g) = \left(\frac{\partial f^j}{\partial x^i}(x)\ \frac{\partial}{\partial y^j}\bigg|_{f(x)}\right)(g),$$

since g is arbitrary, we have:

$$\boxed{df_x\left(\frac{\partial}{\partial x^i}\bigg|_x\right) = \frac{\partial f^j}{\partial x^i}(x)\ \frac{\partial}{\partial y^j}\bigg|_{f(x)}}\,. \tag{2.23}$$

We have obtained what we wanted: the explicit expression of the coordinate tangent vector basis of $T_x\mathbb{R}^m$ transformed by df_x and expressed as a linear combination of the coordinate tangent vector basis of $T_{f(x)}\mathbb{R}^n$.

The coefficients of the linear combination are the partial derivatives of the component functions f^j of f in x, it follow that **the matrix expression of df_x is exactly the Jacobian matrix of f in** x:

$$Jf(x) = \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(x) & \dots & \frac{\partial f^1}{\partial x^m}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^n}{\partial x^1}(x) & \dots & \frac{\partial f^n}{\partial x^m}(x) \end{pmatrix} = \begin{pmatrix} \nabla f^1(x) \\ \vdots \\ \nabla f^n(x) \end{pmatrix}.$$

Let us now consider the more general situation of a smooth function $f : M \to N$ between manifolds of dimension m and n, respectively.

As always, the idea is to select a couple of f-related charts (U, φ) in M containing p and (V, ψ) in N containing $f(p)$ and to consider the local representation of f, i.e. $\tilde{f} = \psi \circ f \circ \varphi^{-1}$, as in the following diagram:

$$\begin{tikzcd} M \supseteq U \arrow[r, "f"] \arrow[d, bend right, "\varphi"'] & V \subseteq N \arrow[d, bend right, "\psi"'] \\ \mathbb{R}^m \supseteq \varphi(U) \arrow[u, bend right, "\varphi^{-1}"'] \arrow[r, dashed, "\tilde{f}"] & \psi(V) \subseteq \mathbb{R}^n. \arrow[u, bend right, "\psi^{-1}"'] \end{tikzcd}$$

We write $\varphi(p) = x \equiv (x^i) \in \varphi(U)$ and $\tilde{f}(x) = \psi(f(p)) = \psi(f(\varphi^{-1}(x))) = y \equiv (y^j) \in \psi(V)$.

Eq. (2.23) implies:

$$d\tilde{f}_x\left(\left.\frac{\partial}{\partial x^i}\right|_x\right) = \frac{\partial \tilde{f}^j}{\partial x^i}(x)\ \left.\frac{\partial}{\partial y^j}\right|_{\tilde{f}(x)}. \tag{2.24}$$

Moreover, by definition of $\tilde{f}$ we get: $f \circ \varphi^{-1} = \psi^{-1} \circ \tilde{f}$, thus $d(f \circ \varphi^{-1}) = d(\psi^{-1} \circ \tilde{f})_x$, so, by the chain rule:

$$df_p \circ d(\varphi^{-1})_x = d(\psi^{-1})_{\tilde{f}(x)} \circ d\tilde{f}_x \tag{2.25}$$

and, thanks to property 4. of the differential (cfr. theorem 2.2.3), $d(\psi^{-1})_{\tilde{f}(x)} = d(\psi_{\psi^{-1}(\tilde{f}(x))})^{-1} = d(\psi_{f(\varphi^{-1}(x))})^{-1} = d(\psi_{f(p)})^{-1}$, i.e.

$$d(\psi^{-1})_{\tilde{f}(x)} = d(\psi_{f(p)})^{-1} \tag{2.26}$$

so:

$$\begin{aligned}
df_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) &\underset{(2.12)}{=} df_p\left((d\varphi_p)^{-1}\left(\left.\frac{\partial}{\partial x^j}\right|_x\right)\right) \underset{((d\varphi_p)^{-1}=d(\varphi^{-1})_x)}{=} df_p\left(d(\varphi^{-1})_x\left(\left.\frac{\partial}{\partial x^j}\right|_x\right)\right) \\
&= df_p \circ d(\varphi^{-1})_x\left(\left.\frac{\partial}{\partial x^j}\right|_x\right) \underset{(2.25)}{=} d(\psi^{-1})_{\tilde{f}(x)} \circ d\tilde{f}_x\left(\left.\frac{\partial}{\partial x^j}\right|_x\right) \\
&= d(\psi^{-1})_{\tilde{f}(x)}\left(d\tilde{f}_x\left(\left.\frac{\partial}{\partial x^j}\right|_x\right)\right) \underset{(2.24)}{=} d(\psi^{-1})_{\tilde{f}(x)}\left(\frac{\partial \tilde{f}^j}{\partial x^i}(x)\ \left.\frac{\partial}{\partial y^j}\right|_{\tilde{f}(x)}\right) \\
&\underset{\text{(linearity)}}{=} \frac{\partial \tilde{f}^j}{\partial x^i}(x)\, d(\psi^{-1})_{\tilde{f}(x)}\left(\left.\frac{\partial}{\partial y^j}\right|_{\tilde{f}(x)}\right) \underset{(2.26)}{=} \frac{\partial \tilde{f}^j}{\partial x^i}(x)\, d(\psi_{f(p)})^{-1}\left(\left.\frac{\partial}{\partial y^j}\right|_{\tilde{f}(x)}\right) \\
&\underset{(2.12)}{=} \frac{\partial \tilde{f}^j}{\partial x^i}(x)\ \left.\frac{\partial}{\partial y^j}\right|_{f(p)},
\end{aligned}$$

so:

$$\boxed{df_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right) = \frac{\partial \tilde{f}^j}{\partial x^i}(x)\ \left.\frac{\partial}{\partial y^j}\right|_{f(p)}}\ . \tag{2.27}$$

If we compare eqs. (2.23) and (2.27), we see that the only difference is that the real coefficients in the latter are given by the partial derivatives of the local expression f w.r.t. the charts chosen. Thus, also in the general case of a smooth function between manifolds, **the matrix expression of the differential of f in p (relative to the coordinate tangent vectors) is given by a Jacobian matrix, but, in this case, of the local expression of f computed in $x = \varphi(p)$**:

$$J\tilde{f}(x) = \begin{pmatrix} \frac{\partial \tilde{f}^1}{\partial x^1}(x) & \dots & \frac{\partial \tilde{f}^1}{\partial x^n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}^n}{\partial x^1}(x) & \dots & \frac{\partial \tilde{f}^n}{\partial x^n}(x) \end{pmatrix} = \begin{pmatrix} \nabla \tilde{f}^1(x) \\ \vdots \\ \nabla \tilde{f}^n(x) \end{pmatrix}.$$

2.4 The inverse mapping and implicit function theorems for manifolds

The result just obtained has a powerful consequence: **all the properties of standard differential calculus on $\mathbb{R}^n$ that are based on hypotheses made on the Jacobian**

matrix of a smooth function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, Ω open, **are also valid, locally, for smooth functions between manifolds**.

In this section we concentrate on two of the most important results of standard differential calculus on $\mathbb{R}^n$: the inverse mapping and the implicit function theorems.

We have already quoted the first, its extension can be stated as follows.

Theorem 2.4.1 (Inverse mapping theorem for manifolds) *Let $f : M \to N$ be a smooth function and $p \in M$ a point such that $df_p : T_pM \to T_{f(p)}N$ is an isomorphism. Then, there exist two open neighborhoods $U \subseteq M$ and $V \subseteq N$ of p and $f(p)$, respectively, such that $f|_U$ is a diffeomorphism.*

Proof. First of all notice that df_p, as a linear map, can be an isomorphism between vector spaces if and only if they have the same dimension, which implies that $\dim(M) = \dim(N)$.

Select two charts (U, φ) and (V, ψ) and consider the local representation $\tilde{f}$ of f. Since the Jacobian matrix of $\tilde{f}$ is the local representation of df_p and df_p is an isomorphism, $J\tilde{f}_{\varphi(p)}$ is invertible. Thanks to this, the standard inverse function theorem can be applied to $\tilde{f}$ and so $f|_U$ is a diffeomorphism. □

Let us now pass to the implicit function theorem by first recalling its classical statement, which tells us, in a very involved way, when we can locally solve an equation as $\phi(x, y) = z_0 \in \mathbb{R}$ and express y as a function of x.

Theorem 2.4.2 (Implicit function theorem in $\mathbb{R}^n$) *Hypotheses:*

- $U \subseteq \mathbb{R}^n \times \mathbb{R}^m$*: open set;*
- $(x^1, \ldots, x^n, y^1, \ldots, y^m)$*: coordinates in* U*;*
- $\phi : U \to \mathbb{R}^m$*: differentiable function;*
- $(x_0, y_0) \in U$ *such that the matrix* $\left(\frac{\partial \phi^i}{\partial y^j}(x_0, y_0)\right)_{i,j}$ *is invertible.*

Thesis:

- $\exists$ *two open neighborhoods* $V_0 \subseteq \mathbb{R}^n$ *of* x_0 *and* $W_0 \subseteq \mathbb{R}^m$ *of* y_0*;*
- $\exists$ *a differentiable function* $F : V_0 \to W_0$*,*

such that, if $\phi(x_0, y_0) = z_0 \in \mathbb{R}$*, the level set* $\phi^{-1}(z_0) \cap (V_0 \times W_0)$ *coincides with the graph of* F*, i.e.*

$$\forall (x, y) \in V_0 \times W_0 \ : \ \phi(x, y) = z_0 \iff y = F(x).$$

Theorem 2.4.3 (Implicit function theorem for manifolds) *Hypotheses:*

- M, N*: smooth manifolds;*

- $\phi : M \times N \to N$: *smooth function;*

- $\forall p \in M$, *let*

$$\begin{array}{rccl} \phi_p : & N & \longrightarrow & N \\ & q & \longmapsto & \phi_p(q) = \phi(p, q); \end{array}$$

- $d(\phi_{p_0})_{q_0} : T_{q_0} \to T_{r_0} Y$, *where* $r_0 = \phi(p_0, q_0)$, *is an invertible linear map.*

Thesis: it exists two open neighborhoods $V_0 \subseteq M$ *of* p_0 *and* $W_0 \subseteq N$ *of* q_0 *and a smooth function* $F : V_0 \to W_0$ *such that* $\phi^{-1}(z_0) \cap (V_0 \cap W_0)$ *coincides with the graph of* F*, i.e.*

$$\forall (p, q) \in V_0 \times W_0 \ : \ \phi(p, q) = r_0 \iff q = F(p).$$

Proof. As for the inverse function theorem, by using two charts we can transport the problem to $\mathbb{R}^n$, where the standard hypotheses of the implicit function theorem hold. □

2.5 Alternative definitions of tangent vectors

In this section we complement the definition of geometric and algebraic tangent vector to a manifold at a point with other two definitions: the first is used mainly by pure mathematicians, the second mainly by physicists and engineers.

2.5.1 Tangent vectors as derivations on the algebra of germs of smooth functions

The name 'germ' is derived from 'cereal germ', which is the reproductive part of the cereal inside the seed. It is clearly used to indicate the 'heart' of a structure. It is a general concept related to topological spaces, where locality can be defined. In this section we will consider only the elements of the theory of germs that are strictly needed to give an alternative definition of tangent vectors, but the theory of germs is much more profound and not just related to differential geometry.

Def. 2.5.1 (Function element) *A smooth function element on a manifold* M *is an ordered pair* (f, U)*, where* U *is an open subset of* M *and* $f : U \to \mathbb{R}$ *is a smooth scalar function.*

Fixed any point $p \in M$, it is possible to define an equivalence relation on the set of all smooth function elements whose domains contain p as follows: given $f : U \to \mathbb{R}$ and $g : V \to \mathbb{R}$, $(f, U) \sim (g, V)$ if it exists an open neighborhood W of p such that:

$$W \subseteq U \cap V \quad \text{and} \quad f|_W = g|_W ,$$

i.e. if f and g coincide on some open neighborhood of p, however small, contained in the intersection of their domains.

Def. 2.5.2 (Germ of f at p) *The germ of* f *at* p *is the equivalence class of function elements* (f, U) *w.r.t. the equivalence relation defined above. The set of all germs of smooth functions at* p *is denoted by* $\mathscr{C}_p^\infty(M)$.

The germ of a function element (f, U) at p is denoted simply by $[f]_p$: in fact, there is no need to include the domain U in the notation because, by definition, the same germ is represented by the restriction of f to any open neighborhood of p.

$\mathscr{C}_p^\infty$ is a real vector space and an associative algebra under the point-wise defined linear operations and multiplication (of course, the sum and the multiplication are defined on the function element that has the intersection of the two functions as second entry in the couple).

We can now define the key concept of derivation on the algebra of germs.

Def. 2.5.3 (Derivation on the algebra of germs of smooth functions) *A derivation v on $\mathscr{C}_p^\infty(M)$ is a linear functional $v : \mathscr{C}_p^\infty(M) \to \mathbb{R}$ satisfying the following Leibniz rule:*

$$v([fg]_p) = (p)v([g]_p) + g(p)v([f]_p).$$

Derivations on $\mathscr{C}_p^\infty(M)$ form naturally a vector space that is denoted by $\mathscr{D}_p(M)$. Some author define the tangent space to M at p as the vector space $\mathscr{D}_p(M)$. The equivalence with the definition of tangent space in terms of derivations on $\mathscr{C}^\infty(M)$ is quite easy to prove thanks to the locality of derivations expressed by theorem 2.2.1.

Theorem 2.5.1 *The map*

$$\begin{array}{rccc} I: & \mathscr{D}_p(M) & \xrightarrow{\sim} & T_p^{\text{alg}}M \\ & v & \longmapsto & I(v), \end{array} \qquad \begin{array}{rccc} I(v): & \mathscr{C}^\infty(M) & \longrightarrow & \mathbb{R} \\ & f & \longmapsto & I(v)(f) := v([f]_p), \end{array}$$

is a natural linear isomorphism of vector spaces that allows us to identify algebraic tangent vectors to M at p with derivations on $\mathscr{C}_p^\infty(M)$.

Proof. Linearity clearly follows from the linearity of the derivation v. The injectivity of I is a consequence of the fact that, if $I(v) = 0$ (the identically null derivation on $\mathscr{C}^\infty(M)$), then, by definition, $v([f]_p) = 0$ for all $f \in \mathscr{C}^\infty(M)$, but this means that v is the null derivation on $\mathscr{C}_p^\infty(M)$, thus $\ker(I)$ is trivial.

Finally, to prove that I is surjective, we must verify that for any $w \in T_p^{alg}M$ there exists $v \in \mathscr{D}_p(M)$ such that $w = I(v)$. Thanks to theorem 2.2.1, such a $v \in \mathscr{D}_p(M)$ can simply be defined as follows:

$$v([f]_p) := w(f),$$

in fact, by definition of germ of smooth functions, if $f, g \in [f]_p$, then f and g are smooth scalar functions that coincide when restricted on an arbitrary small open neighborhood of p, so theorem 2.2.1 assures us that $w(f) = w(g)$, which guarantees that the definition of v is well-posed. Since w is a derivation on $\mathscr{C}^\infty(M)$, v also acts as a derivation on $\mathscr{C}_p^\infty(M)$, i.e. $v \in \mathscr{D}_p(M)$. □

Recall that, in order to obtain theorem 2.2.1, we had to make use of the theory of partitions of the unity and bump functions, thus, one immediate advantage of the use of germs to define tangent vectors is that we can avoid resorting to that theory and prove the same propositions with a less number of intermediate steps. We preferred to postpone until now the definition of tangent vectors via the theory of germs to avoid working with equivalence classes and to keep the notation as simple as possible.

2.5.2 Physicists' definition of tangent vectors

We introduce here **the oldest definition of tangent vector**, which is still the most widely used even today by the majority of physicists and engineers.

The construction is based on the decomposition of a tangent vector $v \in T_pM$ on the coordinates tangent vectors basis, which, as we have seen, is determined once we fix a local chart. Suppose, however, that p belongs to the intersection of two local charts (U, φ) and $(\tilde{U}, \tilde{\varphi})$, then we can decompose v w.r.t. the basis

- $(\frac{\partial}{\partial x^1}\big|_p, \ldots, \frac{\partial}{\partial x^n}\big|_p)$, where $\frac{\partial}{\partial x^i}\big|_p (f) = \frac{\partial(f\circ\varphi^{-1})}{\partial x^i}\Big|_x$

  or w.r.t. the basis

- $(\frac{\partial}{\partial \tilde{x}^1}\big|_p, \ldots, \frac{\partial}{\partial \tilde{x}^n}\big|_p)$, where $\frac{\partial}{\partial \tilde{x}^j}\big|_p (f) = \frac{\partial(f\circ\tilde{\varphi}^{-1})}{\partial \tilde{x}^j}\Big|_x$,

for all $f \in \mathscr{C}^\infty(U \cap \tilde{U})$.

Since the tangent vector v in p remains the same, we must have:

$$v = v^i \left.\frac{\partial}{\partial x^i}\right|_p = \tilde{v}^j \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p, \tag{2.28}$$

where, due to the uniqueness of the decomposition of a vector over a basis, the components v^i are uniquely associated to the coordinates on M defined by local chart (U, φ) and the components $\tilde{v}^j$ are uniquely associated to those defined by $(\tilde{U}, \tilde{\varphi})$. It is natural to ask oneself how the coefficients v^i and $\tilde{v}^j$ are related.

To answer this question, let us recall that the transition functions between these charts are, respectively:

$$\begin{array}{rccl} x^i = \varepsilon^i \circ \varphi \circ \tilde{\varphi}^{-1}: & \tilde{\varphi}(U \cap \tilde{U}) \subseteq \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & \tilde{x} = \tilde{\varphi}(p) & \longmapsto & x^i(\tilde{x}) = \varepsilon^i(\varphi(p)), \end{array}$$

and

$$\begin{array}{rccl} \tilde{x}^j = \varepsilon^j \circ \tilde{\varphi} \circ \varphi^{-1}: & \varphi(U \cap \tilde{U}) \subseteq \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & x = \varphi(p) & \longmapsto & \tilde{x}^j(x) = \varepsilon^j(\tilde{\varphi}(p)), \end{array}$$

$i, j = 1, \ldots, n$.

The tool to obtain the explicit coordinate transformations $v^i \mapsto \tilde{v}^j$ and $\tilde{v}^j \mapsto v^i$ are the following formulae:

$$\left.\frac{\partial}{\partial x^i}\right|_p = \frac{\partial \tilde{x}^j}{\partial x^i}(x) \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p, \tag{2.29}$$

and

$$\left.\frac{\partial}{\partial \tilde{x}^j}\right|_p = \frac{\partial x^i}{\partial \tilde{x}^j}(\tilde{x}) \left.\frac{\partial}{\partial x^i}\right|_p, \tag{2.30}$$

typically quoted to be the result of the application of the chain rule, without any further comment.

It is an instructive computation to verify these formulae. We will do that for the first one, the method to get the second one is identical. Consider the differential:

$$d(\tilde{\varphi} \circ \varphi^{-1})_x \left(\left.\frac{\partial}{\partial x^i}\right|_x \right) \underset{(2.23)}{=} \frac{\partial(\varepsilon^j \circ \tilde{\varphi} \circ \varphi^{-1})}{\partial x^i}(x) \left.\frac{\partial}{\partial \tilde{x}^j}\right|_{\tilde{x}} \equiv \frac{\partial \tilde{x}^j}{\partial x^i}(x) \left.\frac{\partial}{\partial \tilde{x}^j}\right|_{\tilde{x}}, \tag{2.31}$$

then, thanks to def. (2.12) of coordinate tangent vectors, we have:

$$\left.\frac{\partial}{\partial x^i}\right|_p = d(\varphi^{-1})_x \left(\left.\frac{\partial}{\partial x^i}\right|_x \right) \quad (\varphi^{-1} = \tilde{\varphi}^{-1} \circ \tilde{\varphi} \circ \varphi^{-1} \text{ and the chain rule for differential imply})$$

$$= d(\tilde{\varphi}^{-1})_{\tilde{x}} \circ d(\tilde{\varphi} \circ \varphi^{-1})_x \left(\left.\frac{\partial}{\partial x^i}\right|_x \right) \underset{(2.31)}{=} d(\tilde{\varphi}^{-1})_{\tilde{x}} \left(\frac{\partial \tilde{x}^j}{\partial x^i}(x) \left.\frac{\partial}{\partial \tilde{x}^j}\right|_{\tilde{x}} \right) \underset{(2.23)}{=} \frac{\partial \tilde{x}^j}{\partial x^i}(x) \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p,$$

which confirms eq. (2.29).

If we insert the expressions (2.29) and (2.30) in (2.28), we get

$$v^i \left.\frac{\partial}{\partial x^i}\right|_p = \tilde{v}^j \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p \iff v^i \frac{\partial \tilde{x}^j}{\partial x^i}(x) \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p = \tilde{v}^j \frac{\partial x^i}{\partial \tilde{x}^j}(\tilde{x}) \left.\frac{\partial}{\partial x^i}\right|_p.$$

By the uniqueness of the decomposition of a vector on a basis, we have that

$$v = \tilde{v}^j \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p = v^i \frac{\partial \tilde{x}^j}{\partial x^i}(x) \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p$$

implies:

$$\boxed{\tilde{v}^j = \frac{\partial \tilde{x}^j}{\partial x^i}(x) v^i} \iff \tilde{v}^j = \left. J^j_i(\tilde{x})\right|_x v^i, \qquad i,j = 1,\dots,n, \tag{2.32}$$

where $J^j_i(\tilde{x})$ is the $n \times n$ *matrix of functions* that contains the partial derivatives of the function $\tilde{x} = \tilde{\varphi} \circ \varphi^{-1} : \varphi(U \cap \tilde{U}) \subseteq \mathbb{R}^n \to \tilde{\varphi}(U \cap \tilde{U}) \subseteq \mathbb{R}^n$: each rows contains the gradient of the function $\tilde{x}^j = \varepsilon^j \circ \tilde{x}$:

$$J^j_i(\tilde{x}) = \begin{pmatrix} \nabla \tilde{x}^1 \\ \vdots \\ \nabla \tilde{x}^n \end{pmatrix} = \begin{pmatrix} \frac{\partial \tilde{x}^1}{\partial x^1} & \cdots & \frac{\partial \tilde{x}^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{x}^n}{\partial x^1} & \cdots & \frac{\partial \tilde{x}^n}{\partial x^n} \end{pmatrix},$$

once evaluated in x, this becomes a matrix of real entries that represents the Jacobian matrix $\left. J^j_i(\tilde{x})\right|_x$ of the function $\tilde{x}$ in x. Since transition functions are invertible, $\left. J^j_i(\tilde{x})\right|_x \in GL(n, \mathbb{R})$.

On the other side, the equality

$$v = v^i \left.\frac{\partial}{\partial x^i}\right|_p = \tilde{v}^j \frac{\partial x^i}{\partial \tilde{x}^j}(\tilde{x}) \left.\frac{\partial}{\partial x^i}\right|_p$$

implies:

$$\boxed{v^i = \frac{\partial x^i}{\partial \tilde{x}^j}(\tilde{x}) \tilde{v}^j} \iff v^i = {}^t\left(\left. J^j_i(\tilde{x})\right|_x \right)^{-1} \tilde{v}^j, \qquad i,j = 1,\dots,n, \tag{2.33}$$

where ${}^t\left(\left. J^j_i(\tilde{x})\right|_x \right)^{-1}$ is **the inverse and transposed** (notice the position of the indices) of the Jacobian matrix $\left. J^j_i(\tilde{x})\right|_x$. The inversion is to be expected because the transition functions $\tilde{\varphi} \circ \varphi^{-1}$ and $\varphi \circ \tilde{\varphi}^{-1}$ are one the inverse of each other.

The rule (2.32) is called **gradient or contravariant transformation** and it is an **intrinsic property of tangent vectors** (no additional properties of structures have been used to obtain (2.32) other than those related to tangent vectors).

This motivates why tangent vectors can be alternatively defined as ordered n-tuples of real scalars that undergo the contravariant transformation (2.32) under local coordinate changes.

Example: consider the polar coordinates $(x^1, x^2) = (r, \theta) \in \mathbb{R}^+ \times [0, 2\pi)$ on the plane $\mathbb{R}^2$, the point $p = (2, \pi/2)$ and the tangent vector $v \in T_p\mathbb{R}^2$ expressed by:

$$v = 3 \left.\frac{\partial}{\partial r}\right|_p - \left.\frac{\partial}{\partial \theta}\right|_p .$$

We want to find the expression of v w.r.t. Cartesian coordinates. The transition map between polar and Cartesian coordinates in an open neighborhood of $p \in \mathbb{R}^2$ is

$$\begin{cases} \tilde{x}^1 = x = r\cos\theta \\ \tilde{x}^2 = y = r\sin\theta. \end{cases}$$

The vector v can be expressed as follows:

$$v = 3 \left.\frac{\partial}{\partial r}\right|_p - \left.\frac{\partial}{\partial \theta}\right|_p = \tilde{v}^1 \left.\frac{\partial}{\partial x}\right|_p + \tilde{v}^2 \left.\frac{\partial}{\partial y}\right|_p ,$$

with $v^1 = 3$ and $v^2 = -1$. By means of eq. (2.32) we get:

$$\begin{aligned} \tilde{v}^1 &= \frac{\partial \tilde{x}^1(r,\theta)}{\partial x^1}(2,\pi/2)v^1 + \frac{\partial \tilde{x}^1(r,\theta)}{\partial x^2}(2,\pi/2)v^2 = 3\frac{\partial(r\cos\theta)}{\partial r}(2,\pi/2) - \frac{\partial(r\cos\theta)}{\partial\theta}(2,\pi/2) \\ &= (3\cos\theta)|_{(2,\pi/2)} + (r\sin\theta)|_{(2,\pi/2)} = 3\cos(\pi/2) + 2\sin(\pi/2) = 2, \end{aligned}$$

and

$$\begin{aligned} \tilde{v}^2 &= \frac{\partial \tilde{x}^2(r,\theta)}{\partial x^1}(2,\pi/2)v^1 + \frac{\partial \tilde{x}^2(r,\theta)}{\partial x^2}(2,\pi/2)v^2 = 3\frac{\partial(r\sin\theta)}{\partial r}(2,\pi/2) - \frac{\partial(r\sin\theta)}{\partial\theta}(2,\pi/2) \\ &= (3\sin\theta)|_{(2,\pi/2)} - (r\cos\theta)|_{(2,\pi/2)} = 3, \end{aligned}$$

thus:

$$v = 2 \left.\frac{\partial}{\partial x}\right|_p + 3 \left.\frac{\partial}{\partial y}\right|_p .$$

Remark: the notation $\left.\frac{\partial}{\partial x^i}\right|_p$ must not lead to think that the coordinate tangent vector $\left.\frac{\partial}{\partial x^i}\right|_p$ depends only on x^i: in fact it depends on the entire coordinate system. The geometrical reason underlying this is the fact that $\left.\frac{\partial}{\partial x^i}\right|_p$ is the derivation whose action on a smooth scalar function is defined by taking the partial derivative of the local expression of this function w.r.t. x^i, i.e. by letting x^i vary and fixing all the other local coordinates x^j, $j \neq i$. So, if we change the coordinates x^j, they are not constant anymore and this, in general, affects $\left.\frac{\partial}{\partial x^i}\right|_p$.

We illustrate this fact with the following concrete example: consider $\mathbb{R}^2$ with the standard Cartesian coordinates (x, y) and let $p = (1, 0) \in \mathbb{R}^2$, expressed w.r.t. the standard coordinates. Now, perform the coordinate change defined by

$$\begin{cases} \tilde{x} = x \\ \tilde{y} = y + x^3. \end{cases}$$

Our aim is to show that

$$\left.\frac{\partial}{\partial x}\right|_p \neq \left.\frac{\partial}{\partial \tilde{x}}\right|_p ,$$

in spite of the fact that $x = \tilde{x}$.

First of all notice that the coordinates $(\tilde{x}, \tilde{y})$ are smooth and global on $\mathbb{R}^2$ since the inverse of the coordinate change $(x, y) \mapsto (\tilde{x} = x, \tilde{y} = y + x^3)$ is $(\tilde{x}, \tilde{y}) \mapsto (x = \tilde{x}, y = \tilde{y} - \tilde{x}^3)$. Thanks to eq. (2.29), we have:

$$\left.\frac{\partial}{\partial x}\right|_p = \frac{\partial x}{\partial x}(1,0) \left.\frac{\partial}{\partial \tilde{x}}\right|_p + \frac{\partial(y+x^3)}{\partial x}(1,0) \left.\frac{\partial}{\partial \tilde{y}}\right|_p = \left.\frac{\partial}{\partial \tilde{x}}\right|_p + (3x^2)\big|_{(1,0)} \left.\frac{\partial}{\partial \tilde{y}}\right|_p = \left.\frac{\partial}{\partial \tilde{x}}\right|_p + 3 \left.\frac{\partial}{\partial \tilde{y}}\right|_p ,$$

thus $\left.\frac{\partial}{\partial x}\right|_p \neq \left.\frac{\partial}{\partial \tilde{x}}\right|_p$.

From now on, **any definition of tangent vector to a manifold at a point** (geometric, algebraic, via germs of smooth functions or the physicists' one) **will be considered as equivalent**.

2.6 Canonical identification between vector spaces and their tangent spaces and differential of linear functions

We have proven that, for every $p \in M$, T_pM is isomorphic to $\mathbb{R}^n$, which can be considered the (non canonical) prototype of any tangent space to a manifold of dimension n at a given point.

On the other side, $\mathbb{R}^n$ is also the (non canonical) prototype of another object: a real vector space V of dimension n: once we fix a basis of V, the map that links a vector of V to the vector of $\mathbb{R}^n$ given by its components w.r.t. the chosen basis is a linear isomorphism (non canonical because it depends on the basis).

Thanks to the interplay between these two non canonical isomorphisms, we can obtain a third (canonical!) one: we are going to prove that any finite-dimensional vector space V over $\mathbb{R}$ is canonically isomorphic to its tangent space at any point.

In order to prove this, we must play with the dual nature of V: it can be considered as a vector space or as a smooth manifold w.r.t. its standard differential structure defined in section 1.2.

Once we fix any vector $u \in V$, we can consider a particularly natural derivation on $\mathcal{C}^\infty(V)$: the directional derivative of a smooth scalar function $\phi \in \mathcal{C}^\infty(V)$:

- in $x \in V$, where here u is considered as a point of *the manifold* V;
- w.r.t. to the direction defined by any $v \in V$, where v is considered as a vector of *the vector space* V.

By definition, we have:

$$\begin{array}{rccl} D_v|_x : & \mathcal{C}^\infty(V) & \longrightarrow & \mathbb{R} \\ & \phi & \longmapsto & D_v|_u(\phi) = \left.\frac{d}{dt}\right|_{t=0} \phi(x+tv), \end{array}$$

where the operation $x + tv$ is guaranteed to be well-defined in V for all $t \in \mathbb{R}$ because of the vector space structure of V.

$D_v|_x$ plays a central role in the proof of the following result.

Theorem 2.6.1 *Let V, W be any two real finite-dimensional vector spaces with their standard smooth manifold structure. For each point x of V, the map*

$$\begin{array}{rccl} I_x : & V & \xrightarrow{\sim} & T_x V \\ & v & \longmapsto & I_x(v) := D_v|_x\,,\ D_v|_x\,(\phi) = \frac{d}{dt}\big|_{t=0}\,\phi(x+tv) \quad \forall \phi \in \mathscr{C}^\infty(V), \end{array}$$

is a canonical linear isomorphism such that, for any linear map $L : V \to W$, the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "I_x"] \arrow[d, "L"'] & T_x V \arrow[d, "dL_x"] \\ W \arrow[r, "I_{Lx}"'] & T_{Lx} W \end{tikzcd} \qquad \Longleftrightarrow \qquad dL_x \circ I_x = I_{Lx} \circ L.$$

So, the explicitly formula for the differential of a linear function between vector spaces is:

$$dL_x(\,D_v|_x) = D_{Lv}|_{Lx}$$

or, by identifying $D_v|_x$ with v,

$$dL_x(v) = D_{Lv}|_{Lx}\,.$$

Proof. The linearity of I_x is a direct consequence of the linearity of $D_v|_x$. Let us now prove that I_x is a bijection.

<u>Injectivity</u>: suppose $v_1, v_2 \in V$ are such that $D_{v_1}|_x = D_{v_2}|_x$, then, thanks to the fact that $D_v|_x$ is linear w.r.t. v, we have that $D_{(v_1-v_2)}\big|_x \equiv 0$, the 0 derivation. Then, for all $\phi \in \mathscr{C}^\infty(V)$:

$$D_{(v_1-v_2)}\big|_x\,(\phi) = \left.\frac{d}{dt}\right|_{t=0} \phi(x + t(v_1 - v_2)) = 0.$$

Now we note that the linear functionals $\ell : V \to \mathbb{R}$ living in the dual V^* of V are of course smooth scalar functions on V, i.e. they belong to $\mathscr{C}^\infty(V)$, so we can consider the action of $D_{(v_1-v_2)}$ on $\ell \in V^*$:

$$0 = \left.\frac{d}{dt}\right|_{t=0} \ell(x + t(v_1 - v_2)) = \left.\frac{d}{dt}\right|_{t=0} (\ell(x) + t\ell(v_1 - v_2)) = \cancelto{0}{\left.\frac{d}{dt}\right|_{t=0} \ell(x)} + \left.\frac{d}{dt}\right|_{t=0} t\ell(v_1 - v_2),$$

i.e. $\ell(v_1 - v_2) = 0$ for all $\ell \in V^*$. However, thanks to the finite-dimensional Riesz representation theorem, we know that $V \cong V^*$ and that for all $\ell \in V^*$ it exists only one vector $w_\ell \in V$ such that $\ell(v) = \langle v, w_\ell \rangle$. Thus, the equation $\ell(v_1 - v_2) = 0$ for all $\ell \in V^*$ can be reformulated as $\langle v_1 - v_2, w_\ell \rangle = 0$ for all $w_\ell \in V$, but the only vector orthogonal to all other vectors is the null vector, so $v_1 - v_2 = 0$, or $v_1 = v_2$, thus implying the injectivity of I_x.

<u>Surjectivity</u>: we can conveniently use the equivalence between $T_x^{\mathrm{geom}} V$ and $T_x^{\mathrm{alg}} V$ and prove surjectivity by considering geometric tangent vectors. The proof then simply consists in observing that any tangentially equivalent class of curves passing through x with velocity vector v clearly contains the curve $t \mapsto x + tv$.

Finally, suppose $L : V \to W$ to be a linear map, then L is of course smooth because its components w.r.t. any choice of basis (which also play the role of charts for vector spaces, as seen in chapter 1) for V and W are linear functions of the coordinates. By definition of differential and thanks to the linearity of L we get, for all $\phi \in \mathscr{C}^\infty(V)$:

$$
\begin{aligned}
dL_x(D_v|_x)(\phi) := D_v|_x(\phi \circ L) = \left.\frac{d}{dt}\right|_{t=0} \phi(L(x+tv)) = \left.\frac{d}{dt}\right|_{t=0} \phi(Lx + tLv) \\
= D_{Lv}|_{Lu}(\phi),
\end{aligned}
$$

i.e. $dL_x(D_v|_x) = D_{Lv}|_{Lx}$. □

I_x is a canonical isomorphism because it does not depend on any choice of basis of V.

An immediate, and very important, consequence of the previous theorem is that, if U is an open submanifold of a real finite-dimensional vector space V, then $T_xU \cong T_xV \cong V$, so we obtain a canonical identification of each tangent space to U with V itself. As a noticeable example, since $GL(n,\mathbb{R})$ is an open submanifold of the vector space $M(n,\mathbb{R})$, the following result holds.

Theorem 2.6.2 *For all $X \in GL(n,\mathbb{R})$ it holds that:*

$$T_X\, GL(n,\mathbb{R}) \cong M(n,\mathbb{R}),$$

i.e. the tangent space to the vector space of real invertible matrices of dimension n is the vector space of all real square matrices of dimension n.

There is another natural, and very useful, identification for tangent spaces to a product manifold, as stated in the following proposition.

Theorem 2.6.3 *Let $M_1, \ldots, M_N$ be smooth manifolds, and let $\pi_j : M_1 \times \cdots \times M_N \to M_j$, be the projection onto the j-th factor, for each $j = 1, \ldots, N$. For any point $p = (p_1, \ldots, p_N) \in M_1 \times \cdots \times M_N$ the map*

$$
\begin{array}{ccc}
T_p(M_1 \times \cdots \times M_N) & \xrightarrow{\sim} & T_{p_1}M_1 \oplus \cdots \oplus T_{p_N}M_N \\
v & \longmapsto & (d(\pi_1)_p(v), \ldots, d(\pi_N)_p(v)),
\end{array}
$$

is a canonical isomorphism.

For example, $T_{(p,q)}(M \times N)$ can be identified with $T_pM \oplus T_qN$ and T_pM and T_qN can be treated as subspaces of $T_{(p,q)}(M \times N)$.

2.7 Immersion, submersion, embedding and the problem of compatibility between differential structures

The substructures of a manifold show some subtleties that is important to underline.

First of all, let us define the rank of a smooth map in an analogous way as we did for a smooth function between Euclidean spaces.

Def. 2.7.1 *Let $f : M \to N$ be a smooth map between manifolds. The* ***rank*** *of f in $p \in M$ is the rank of the linear function $df_p : T_pM \to T_{f(p)}N$.*

Equivalently, fixed any local chart (U,φ) in p, the rank of f is the rank of the Jacobian matrix of the local expression $\tilde{f}$ of f in $x = \varphi(p)$.

If the rank of f remains constant for every point $p \in M$, then f is said to have ***constant rank****.*

Def. 2.7.2 *The smooth map $f : M \to N$ is a/an:*

- ***Immersion****: if df_p is injective for all $p \in M$;*
- ***Submersion****: if df_p is surjective for all $p \in M$;*
- ***Embedding***[3]*: if it is an immersion and $f : M \to f(M)$ is a homeomorphism.*

Examples:

1. The curve
$$\begin{array}{rccl} \alpha : & \mathbb{R} & \longrightarrow & \mathbb{R}^2 \\ & t & \longmapsto & \alpha(t) = (t^2, t^3), \end{array}$$
is injective, but $\frac{d\alpha}{dt} = (2t^2, 3t^2)$ is null for $t = 0$, so $d\alpha|_{t=0}$ is not injective;

2. The curve
$$\begin{array}{rccl} \beta : & \mathbb{R} & \longrightarrow & \mathbb{R}^2 \\ & t & \longmapsto & \alpha(t) = (t^3 - 4t, t^2 - 4), \end{array}$$
is not injective, e.g. $\beta(-2) = \beta(2) = (0, 0)$, but $\frac{d\beta}{dt} = (3t - 4, 2t)$ is never null in both coordinates, so β is an immersion, but not an embedding because it is not injective;

3. The curve
$$\begin{array}{rccl} \gamma : & (-\pi/2, 3\pi/2) & \longrightarrow & \mathbb{R}^2 \\ & t & \longmapsto & \alpha(t) = (\sin(2t), \cos(t)), \end{array}$$
γ is injective and $\frac{d\gamma}{dt} = (2\cos(2t), -\sin(t)) \neq (0, 0)$ $\forall t \in (-\pi/2, 3\pi/2)$, thus it is an immersion. However, the domain of γ is an open set in $\mathbb{R}$ and its codomain is a compact subset of $\mathbb{R}^2$, thus γ cannot be a homeomorphism between its domain and its codomain.

The curve γ, usually called **lemniscate**, or 'the 8' for its shape, shows that *even an injective immersion can fail to be an embedding.* However, the next theorem guarantees that *every immersion is*, at least, *local embedding.*

Theorem 2.7.1 *Let $f : M \to N$ be a smooth map between manifolds. If f is an immersion, then, for all $p \in M$, it exists an open neighborhood $U \subseteq M$ of p such that $f|_U : U \to f(U) \subseteq N$ is an embedding.*

The most important consequence of the previous result is that, *if $f : M \to N$ is an injective immersion, it is always possible to endow $f(M)$ with a differential structure induced by that of M.* In fact, let $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ be a smooth atlas for M such that $f|_{U_\alpha}$ is a homeomorphism with its image $f(U_\alpha) \subseteq f(M)$, then, since $\varphi : U_\alpha \subseteq M \to \varphi(U) \subseteq \mathbb{R}^n$ are homeomorphisms, we get that $\{(f(U_\alpha), \varphi_\alpha \circ f^{-1}|_{f(U_\alpha)})\}_{\alpha \in A}$ is a smooth atlas for $f(M)$.

Thus, on $f(M)$ we have two differential structures, namely, the one naturally inherited as a subset of N and the one induced by M in the way described above. It turns out that these differential structures can lack of compatibility because the underlying topologies may fail to be equivalent. This is clearly exemplified by the curve γ: the counter-image of an open

[3] An embedding is a sort of topologically coherent immersion. In French it is called *plongement.*

neighborhood of the central point of the 8, in $\mathbb{R}^2$, is the union of three open intervals in $\mathbb{R}$, while for the topology of $\mathbb{R}$ an open neighborhood is just an open interval.

In general, it can be difficult to establish if an injective immersion is an embedding, with the exception of the compact case, as stated below.

Theorem 2.7.2 *Let $f : M \to N$ be a smooth map between manifolds. If f is an injective immersion and M is compact (as topological manifold), then f is an embedding.*

The considerations above explain why we find two types of definitions for submanifolds in differential geometry.

Def. 2.7.3 (Embedded submanifold) *Let E, M be two smooth manifolds such that $E \subset M$. If the canonical inclusion $\iota : E \hookrightarrow M$ is an embedding, then E is said to be an embedded submanifold of M.*

Def. 2.7.4 (Immersed submanifold) *Let $f : M \to N$ be a smooth map between manifolds. If f is an injective immersion, then $f(M) \subset N$, endowed with the differential structure induced by M, is said to be a manifold immersed in N.*

Convention: without any further specification, a submanifold has to be intended as an embedded submanifold.

A classical example of an immersed submanifold of $\mathbb{R}^2$ that is not an embedding is the spire (coil) that envelops the torus with irrational step.

2.8 Characterization of the tangent space to a level set of a smooth function

It is possible to give a very useful characterization of the tangent space at a point to a level set of smooth functions thanks to the following result, whose proof can be found in [10], page 81 (th. 4.12).

Theorem 2.8.1 (The rank theorem) *Let M and N be smooth manifolds with dimension m and n, respectively. Let $f : M \to N$ be a smooth function with constant rank r. Then, for every $p \in M$ there exist local charts (U, φ) centered in p and (V, ψ) centered in $f(p)$, with $f(U) \subset V$, such that the local expression of f w.r.t. these charts is particularly simple, namely:*

$$\tilde{f}(x^1, \ldots, x^r, x^{r+1}, \ldots, x^m) = (x^1, \ldots, x^r, 0, \ldots, 0),$$

i.e. $\tilde{f}$ acts as the identity on the first r entries and it is identically 0 in the last $n - r$.

In particular, if f is a submersion, then $r = n$ and so

$$\tilde{f}(x^1, \ldots, x^n, x^{n+1}, \ldots, x^m) = (x^1, \ldots, x^n),$$

while, if f is an immersion, then $r = m$ and so

$$\tilde{f}(x^1, \ldots, x^m) = (x^1, \ldots, x^m, 0, \ldots, 0).$$

The rank theorem justifies the following definition.

Def. 2.8.1 *Let $S \subset M$ be a submanifold of dimension k of M. A local chart (U, φ) of M is said to be adapted to S if either $U \cap S = \varnothing$, or $\varphi(U \cap Z) = \varphi(U) \cap (\mathbb{R}^k \times \{\vec{0}\})$, where this notation means that the part of the submanifold S contained in U is mapped by φ to 0, i.e. $x^{k+1} = \ldots = x^n = 0$. An atlas of M is adapted to S is every chart of it is adapted to S.*

Theorem 2.8.2 *Embedded submanifolds always admit adapted charts.*

We can extend the definitions, given in chapter 1, of critical and regular point of a function defined between Euclidean spaces to functions between abstract manifolds.

Def. 2.8.2 *Let M and N be smooth manifolds with dimension m and n, respectively. Let $f : M \to N$ be a smooth function.*

- *$p \in M$ is a **critical point** of f if $df_p : T_pM \to T_{f(p)}N$ is not surjective. The image, via f, of a critical point of f is a **critical value** for f.*
- *A regular value of f is an element of $f(M)$ that is not a critical value.*

We denote with $\mathrm{Crit}(f) \subset M$ the set of critical points of f.

We need a last definition before stating and proving the main result of this section.

Def. 2.8.3 (Level set of a smooth function) *A level set of $f : M \to N$ is a subset of M of the type $f^{-1}(q) := \{p \in M \ : \ f(p) = q\}$, where $q \in f(M)$.*

Theorem 2.8.3 (Level set theorem for manifolds) *Let M and N be smooth manifolds with dimension $n + k$ and n, respectively, $k \geqslant 0$. Let $f : M \to N$ be a smooth function.*

1. *For all $a \in f(M)$, the set*

$$M_a = f^{-1}(a) \backslash \mathrm{Crit}(f) \qquad a - \text{level set via } f \text{ minus the critical points}$$

is an embedded submanifold of dimension k of M. In particular, if a is a regular value for f, $f^{-1}(a)$ is a k-dimensional embedded submanifold of M.

2. *If $p \in M_a$, then the tangent space T_pM_a is the kernel of $df_p : T_pM \to T_{f(p)}N$:*

$$\boxed{T_pM_a = \ker(df_a)} \ . \tag{2.34}$$

3. *If, in particular, $N = \mathbb{R}$, then $f \in \mathcal{C}^\infty(M)$ and T_pM_a is given by the derivations $D \in T_pM$ that nullify smooth scalar functions: $D(f) = 0$ for all $f \in \mathcal{C}^\infty(M)$.*

Proof.

1. By using local charts, we can reduce the problem to the local representation of f, which is a function defined on an open subset of $\mathbb{R}^{n+k}$ to $\mathbb{R}^n$. For such a function we can apply the level set theorem 1.2.1 in Euclidean spaces discussed in the first chapter.

2. Let $\iota : M_a \to M$ be the canonical inclusion of M_a in M. By theorem 2.2.3 we know that $d\iota_p : T_pM_a \to T_pM$ is a canonical linear isomorphism, thus we can identify T_pM_a with T_pM and so 2. is equivalent to $d\iota_p(T_pM_a) = \ker(df_p)$.

Since p is a regular point, $\dim(M_a) = k$, so $\dim(T_pM_a) = k$, moreover df_p is surjective, hence $\dim(\mathrm{Im}(df_p)) = n$ and the rank+nullity theorem implies

$$\dim(T_pM) = \dim(\ker(df_p)) + \dim(\mathrm{Im}(df_p)),$$

but $\dim(T_pM) = \dim(M) = n + k$, so $\dim(\ker(df_p)) = n + k - n = k$.

Thanks to the fact that $\dim(\ker(df_p)) = \dim(T_pM_a)$, to prove that T_pM_a and $\ker(df_p)$ are isomorphic it is sufficient to show that one space is included in the other. We chose arbitrarily to show that $T_pM_a \cong d\iota(T_pM_a) \subseteq \ker(df_p)$.

To do that, let us consider a derivation $v \in T_pM_a$, then $d\iota_p : T_pM_a \to T_{\iota(p)}M = T_pM$, so $d\iota_p(v) \in T_pM$ and we can apply $df_p : T_pM \to T_{f(p)}N$ to $d\iota_p(v)$, obtaining an element of $T_{f(p)}N$, i.e. $df_p(d\iota_p) \in T_{f(p)}N$. In order to understand its action, we need to apply it to a smooth scalar function $\phi \in \mathscr{C}^\infty(N)$:

$$df_p(d\iota(v))(\phi) \underset{\text{(chain rule)}}{=} d(f \circ \iota)_p(v)(\phi) := v(\phi \circ f \circ \iota),$$

but $f \circ \iota : M_a \to N$ is nothing but $f|_{M_a}$, so

$$df_p(d\iota(v))(\phi) = v(\phi \circ f|_{M_a}) = 0,$$

because $f|_{M_a}$ is, by definition of M_a, a constant function identically equal to a, and $\phi \circ f|_{M_a}$ is the constant function identically equal to $\phi(a)$, so $v(\phi \circ f|_{M_a}) = 0$ because derivations set to 0 constant functions.

This is true for all $\phi \in \mathscr{C}^\infty(N)$, so $df_p(d\iota_p(v)) = 0$, i.e. $T_pM_a \subseteq \ker(df_p)$.

3. Immediate consequence of 2. ☐

The rank theorem has also other applications. Here we discuss one which is related to manifolds with boundaries. In fact, thanks to this theorem we can build a canonical, and very important, example of manifold with boundary.

Def. 2.8.4 (Regular domain of a manifold without boundary) *Let M be a manifold without border of dimension n and $\rho \in \mathscr{C}^\infty(M)$ a smooth scalar function on M which has 0 as a regular point. The ρ-regular domain $\bar{D}_\rho$ of M is the set $\bar{D} \subset M$ defined as follows*

$$\bar{D} := \{p \in M \ : \ \rho(p) \geqslant 0\},$$

ρ is called the defining function of $\bar{D}_\rho$.

Theorem 2.8.4 *The ρ-regular domain defined above is a n-dimensional manifold with boundary with companion manifolds (without boundary) $\partial\bar{D} = \rho^{-1}(0)$ of dimension $n-1$ and* $\mathrm{int}(\bar{D}) = \{p \in M \ : \ \rho(p) > 0\}$

Proof. The generic proof is a generalization, which uses the level set theorem for manifolds, of the proof that follows which we will develop in the case when $M = \mathbb{R}^n$ to show in a simpler way how to endow $\bar{D}$ with a smooth atlas. So, let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a smooth function with 0 as regular value, so that $\rho^{-1}(0)$ is a hypersurface, i.e. a sub-manifold of dimension $n-1$, of $\mathbb{R}^n$ by the level set theorem 1.2.1.

So, let us consider the set

$$\bar{D} := \{x \in \mathbb{R}^n \ : \ \rho(x) \geqslant 0\}$$

and build a smooth atlas for it.

For $\mathrm{int}(\bar{D}) \equiv D = \{x \in \mathbb{R}^n \ : \ \rho(x) > 0\}$ it is immediate to build an atlas, which is actually a mono-chart one, in fact D is open in $\mathbb{R}^n$, so as chart domain we can take D itself, and as chart map we can take the composition of the canonical inclusion $\iota : D \hookrightarrow \mathbb{R}^n$ with any diffeomorphism between $\mathbb{R}^n$ and $\mathrm{int}(\mathbb{H}^n)$, for example the one which uses the exponential and logarithmic functions exhibited above, to get a global chart map φ:

$$\begin{tikzcd} D \arrow[r, hook, "\iota"] \arrow[rr, bend right, "\varphi"'] & \mathbb{R}^n \arrow[r, "\sim"] & \mathrm{int}(\mathbb{H}^n) \subset \mathbb{H}^n. \end{tikzcd}$$

Let us now concentrate on the n-border charts. The point $p \in \partial\bar{D} = \rho^{-1}(0)$ is not a critical point for ρ because, by hypothesis, 0 is a regular value for ρ, so it exists an open neighborhood U of p in $\mathbb{R}^n$ on which ρ has constant rank, of course equal to 1 because ρ is a scalar function.

Thanks to the rank theorem we know that, possibly under the condition of restricting U to a smaller open set, we can find a chart (U, φ) of $\mathbb{R}^n$ centered in p and a chart (V, ψ) of $\mathbb{R}$ centered in 0, such that $\psi \circ \varphi = \phi^n$ on U, where φ^n is the last, n-th, component function of $\varphi \equiv (\varphi^1, \ldots, \varphi^n)$, as depicted in the following commutative diagram:

$$\begin{tikzcd} \mathbb{R}^n \supset U \arrow[r, "\rho"] \arrow[d, "\varphi"'] & V \subset \mathbb{R} \arrow[d, "\psi"] \\ (x^1, \ldots, x^n) \in \varphi(U) \arrow[r] & x^n \in \psi(V) \end{tikzcd}$$

This means that the fact that $\rho(x) = 0$ is equivalent to the fact that $\varphi^n(x) = 0$, i.e. $x^n = 0$, which defines the border of $\mathbb{H}^n$. So, $\varphi(U \cap \partial\bar{D}) = \varphi(U) \cap \partial\mathbb{H}^n$.

Since ψ is bijective, as a local chart, it is either increasing or decreasing. Modulo a change of sign of φ^n, we can suppose that ψ is increasing, hence

$$\rho(x) > 0 \iff x^n = \varphi^n(x) = \psi(\rho(x)) > 0,$$

but then $\varphi(U \cap D) = \varphi(U) \cap \mathrm{int}(\mathbb{H}^n)$.

This implies that $(U \cap D, \varphi)$ is a border chart for $\bar{D}$. Moreover, any two border charts built in this way, i.e. as restriction of charts of $\mathbb{R}^n$, inherit their compatibility.

Having built compatible inner and border charts for $\bar{D}$, we have shown that it is a n-manifold with boundary. $\square$

2.9 Explicit calculations of tangent spaces

In this section we are going to compute some remarkable differential and apply the result to obtain the explicit characterization of tangent spaces. In order to do that, we will mix the level set theorem with the results that we have discussed about the differential.

2.9.1 The tangent space to the sphere at a point

We are going to verify that the tangent space to a sphere at a point x is the hyperplane orthogonal to the radius connecting the center to x, as intuitively expected from the depiction in fig. 2.1.

We recall that the n-sphere of radius $R > 0$ is $S^n_R = \{x \in \mathbb{R}^{n+1} \ : \ \|x\|^2 = R^2\}$, can be identified with the level set: $S^n_R = f^{-1}(R^2)$, where $f : \mathbb{R}^{n+1} \to \mathbb{R}$, $x \mapsto f(x) = \|x\|^2$.

We know that in this case the differential of f coincides with its total derivative, i.e. for all $x \in \mathbb{R}^{n+1}$, $df_x = Df(x)$, to compute it we simply observe that:

$$f(x+ty) = \|x+ty\|^2 = \langle x+ty, x+ty\rangle = \|x\|^2 + 2t\langle x, y\rangle + t^2\,\|y\|^2\,,$$

but $\|x\|^2 = f(x)$, $t^2\,\|y\|^2 = o(t)$, so $df_x(y) = Df(x)y = 2\langle x, y\rangle$ for all $y \in \mathbb{R}^{n+1}$.

By the level set theorem we get:

$$T_x S^n_R = \ker(df_x) = \{y \in \mathbb{R}^{n+1} \ : \ \langle x, y\rangle = 0\},$$

which confirms that $T_x S^n_R$ is nothing but the hyperplane in $\mathbb{R}^{n+1}$ passing through x and orthogonal to the radius of the sphere connecting x to 0.

2.9.2 The Lie group $\mathrm{O}(n)$ as an embedded submanifold of $M(n, \mathbb{R})$ and its Lie algebra $\mathfrak{o}(n)$

Here we prove that $\mathrm{O}(n) = \{A \in M(n, \mathbb{R}) \ : \ A^tA = I_n\}$, the orthogonal group, is a manifold of dimension $\frac{n(n-1)}{2}$ and we make its tangent space at any point explicit. The constraint that defines orthogonal matrices leads us naturally to consider the following function:

$$\begin{array}{rccc} f: & M(n,\mathbb{R}) & \longrightarrow & \mathrm{Sym}(n,\mathbb{R}) \\ & X & \longmapsto & f(X) = X^tX, \end{array} \tag{2.35}$$

because we can easily identify $\mathrm{O}(n)$ as the f-level set of the the identity matrix I_n, in fact:

$$f^{-1}\{I_n\} = \{X \in M(n,\mathbb{R}) \ : \ f(X) = X^tX = I_n\} \equiv \mathrm{O}(n).$$

In order to apply the level set theorem, let us compute the differential of f. Both $M(n, \mathbb{R})$ and $\mathrm{Sym}(n, \mathbb{R})$ are vector spaces, thus we can canonically identify the tangent spaces to $M(n, \mathbb{R})$ and $\mathrm{Sym}(n, \mathbb{R})$ at any point (matrix) with the vector spaces themselves. With this identification in mind, for all $X \in M(n, \mathbb{R})$, $df_X : M(n, \mathbb{R}) \to \mathrm{Sym}(n, \mathbb{R})$ and, thanks to eq. (2.22), for all $A \in M(n, \mathbb{R})$ we have:

$$\begin{aligned} df_X(A) &= \left.\frac{d}{dt}\right|_{t=0} f(X+tA) = \left.\frac{d}{dt}\right|_{t=0} \big((X+tA)^t(X+tA)\big) \\ &= \left.\frac{d}{dt}\right|_{t=0} \big(X^tX + t(X^tA + A^tX) + t^2A^tA\big) \\ &= \cancelto{0}{\left.\frac{d}{dt}\right|_{t=0} \big(X^tX\big)} + \big(X^tA + A^tX\big) + \cancelto{0}{\big(2tA^tA\big)\big|_{t=0}} \\ &= X^tA + A^tX, \end{aligned}$$

i.e. $df_X(A) = X^tA + A^tX$, which is, as it should be, a symmetric matrix.

Remark: this result could have been obtained also by identifying the differential with the total derivative and observing that:

$$f(X + tA) = X^t X + t(X^t A + A^t X) + t^2 A^t A = f(X) + Df(X)tA + o(t),$$

so that $d_X f(A) = Df(X)A = X^t A + A^t X$.

Now that the differential is explicit, let us analyze its surjectivity: for every $B \in \mathrm{Sym}(n, \mathbb{R})$ we must determine under what condition on X it exists at least one $A \in M(n, \mathbb{R})$ such that $B = df_X(A) = X^t A + A^t X$.

To obtain this result first notice that B is symmetric, so we can write:

$$B = \frac{1}{2}B + \frac{1}{2}B = \frac{1}{2}B + \frac{1}{2}B^t,$$

that must be compared to

$$B = X^t A + A^t X = X^t A + (X^t A^t)^t,$$

the two expressions are compatible if and only if $X^t A = \frac{1}{2}B$, if X is invertible, then we can solve this equation obtaining $A = \frac{1}{2}(X^t)^{-1}B$. Thus $df_X : T_X\mathrm{GL}(n, \mathbb{R}) \cong M(n, \mathbb{R}) \to \mathrm{Sym}(n, \mathbb{R})$ is surjective for all $X \in \mathrm{GL}(n, \mathbb{R})$, since every symmetric $n \times n$ real matrix B can be written as $df_X(\frac{1}{2}(X^t)^{-1}B)$, where $X \in \mathrm{GL}(n, \mathbb{R})$.

The identity I_n is symmetric, an orthogonal matrix X is invertible and $I_n = f(X)$, thus I_n is a regular value for f and the level set theorem can be applied to guarantee that $\mathrm{O}(n) = f^{-1}(I_n)$ is an embedded submanifold of $M(n, \mathbb{R})$ of dimension $\dim(\mathrm{O}(n)) = \dim(M(n, \mathbb{R})) - \dim(\mathrm{Sym}(n, \mathbb{R}))$.

The dimension of $\mathrm{Sym}(n, \mathbb{R})$ can be recovered by observing that if we want to identify a symmetric matrix of order n we must specify $\frac{n(n+1)}{2}$ real values: $n^2 - n$ is the totality of matrix elements minus those lying on the diagonal, if we divide this number by 2 we obtain the matrix element above (or below) the diagonal, to these elements we must add back the diagonal entries, thus arriving to $\frac{n^2-n}{2} + n = \frac{n(n+1)}{2}$. Hence, $\mathrm{Sym}(n, \mathbb{R})$ is isomorphic to $\mathbb{R}^{n(n+1)/2}$ and so it has dimension $\frac{n(n+1)}{2}$ as a manifold.

It follows that the dimension of $\mathrm{O}(n)$ as embedded submanifold of $M(n, \mathbb{R})$ is:

$$\dim(\mathrm{O}(n)) = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}.$$

Finally, thanks to (2.34), we can compute the tangent space to $\mathrm{O}(n)$ as follows:

$$T_X\mathrm{O}(n) = \ker(df_X) = \{A \in M(n, \mathbb{R}) \ : \ X^t A + A^t X = 0\}, \quad \forall X \in \mathrm{O}(n),$$

i.e. matrices $A \in M(n, \mathbb{R})$ such that $X^t A$ is skew-symmetric, thus, in particular, if $X = I_n$,

$$T_{I_n}\mathrm{O}(n) = \{A \in M(n, \mathbb{R}) \ : \ A + A^t = 0 \iff A^t = -A\},$$

i.e. the tangent space at the identity element of $\mathrm{O}(n)$ can be identified with the space of **skew-symmetric matrices**.

We will see in the chapter dedicated to Lie groups that $T_{I_n}\mathrm{O}(n)$ can be identified with the Lie algebra of the Lie group $\mathrm{O}(n)$, that will be denoted with the symbol $\mathfrak{o}(n)$:

$$\mathrm{O}(n) = \{A \in M(n, \mathbb{R}) \ : \ A^t = A^{-1}\}, \quad \mathfrak{o}(n) = \{A \in M(n, \mathbb{R}) \ : \ A^t = -A\}.$$

Remark: if A were a positive real number a, then we could compute the logarithm of A^{-1}, obtaining $\log A^{-1} = -\log A$, which suggests that the elements of $\mathrm{O}(n)$ could be considered as the exponential of the elements of $\mathfrak{o}(n)$. We will see that, indeed, it exists a fundamental function, called again *exponential*, that relates Lie algebras and Lie groups.

Chapter 3

Tangent, cotangent and vector bundles

Inspirational epithap wanted...

...

The simple act of taking the union of the tangent spaces to a manifold in all its points generates another manifold, the tangent bundle, with double the dimension of the original one, and with a surprisingly rich intrinsic structure that happens to be the prototype of the so-called vector bundles.

3.1 The tangent bundle over a manifold

We have seen that the tangent spaces T_pM and T_qM to a smooth manifold M of dimension n in two different points p and q are not canonically isomorphic and so they cannot be identified, in spite of the fact that they are both two copies of $\mathbb{R}^n$.

The union of the tangent spaces to M as we vary the point on M is then a *disjoint* one. The canonical symbol to denote this disjoint union is:

$$TM = \bigsqcup_{p \in M} T_pM = \{(p, v) \, : \, p \in M, \ v \in T_pM\}.$$

This space comes equipped with a natural projection:

$$\begin{array}{rccl} \pi : & TM & \longrightarrow & M \\ & (p, v) & \longmapsto & \pi(p, v) = p. \end{array}$$

Def. 3.1.1 *TM is called the* ***tangent bundle*** *of the smooth manifold M. The* ***fiber*** *over $p \in M$ is the set:*

$$\pi^{-1}(p) = \{(p, v) \, : \, v \in T_pM\} \cong T_pM.$$

The most important geometrical characteristic of the tangent bundle is its *local triviality*, i.e. the fact that, locally, it is diffeomorphic to the Cartesian product between a chart domain and $\mathbb{R}^n$, the local model of M.

Local triviality is easily understood if we consider a local chart (U, φ) in $p \in M$ and the restriction of TM to U, defined by

$$TM|_U = \bigsqcup_{p \in U} T_pM.$$

As we have seen in chapter 2, the act of fixing a local chart (U, φ) in $p \in M$ induces the non-canonical linear isomorphism $d\varphi_p : T_pM \xrightarrow{\sim} \mathbb{R}^n$ defined by $d\varphi_p(\partial_i|_p) = e_i$, where e_i is the i-th element of the canonical basis of $\mathbb{R}^n$, $i = 1, \ldots, n$, thus the extension on the whole tangent space to M at p is given by the correspondence: $T_pM \ni v = v^i \, \partial_i|_p \longleftrightarrow (v^i)_{i=1}^n \in \mathbb{R}^n$.

This holds for every point $p \in U$, so we can extend this non-canonical identification to all U as follows:

$$\begin{array}{rrcl} id_U \times d\varphi_p : & TM|_U = \bigsqcup_{p \in U} T_pM & \xrightarrow{\sim} & U \times \mathbb{R}^n \\ & (p, (v^i \, \partial_i|_p)_{i=1}^n) & \longmapsto & (p, (v^i)_{i=1}^n). \end{array}$$

Def. 3.1.2 *The map $id_U \times d\varphi_p : TM|_U \xrightarrow{\sim} U \times \mathbb{R}^n$ is called a* ***local trivialization of the tangent bundle*** *TM.*

Each chart map sends $U \subseteq M$ diffeomorphically to $\varphi(U) \subseteq \mathbb{R}^n$, thus we can further identify $U \times \mathbb{R}^n$ with an open subset $\varphi(U) \times \mathbb{R}^n$ of $\mathbb{R}^{2n}$ as follows:

$$\begin{array}{rrcl} \varphi \times id_{\mathbb{R}^n} : & U \times \mathbb{R}^n & \xrightarrow{\sim} & \varphi(U) \times \mathbb{R}^n \\ & (p, (v^i)_{i=1}^n) & \longmapsto & (x, (v^i)_{i=1}^n), \quad x = \varphi(p). \end{array}$$

By composition we obtain:

$$\begin{array}{rrcl} \Phi \equiv (\varphi \times id_{\mathbb{R}^n}) \circ (id_U \times d\varphi_p) : & TM|_U & \xrightarrow{\sim} & \varphi(U) \times \mathbb{R}^n \subseteq \mathbb{R}^{2n} \\ & (p, (v^i \, \partial_i|_p)_{i=1}^n) & \longmapsto & (x, (v^i)_{i=1}^n), \quad x = \varphi(p), \end{array}$$

which shows that the couple $(TM|_U, \Phi)$ is *a local chart for TM* with local coordinates obtained by replacing φ by its component functions $x^i \equiv (\varepsilon^i \circ \varphi)_{i=1}^n$, i.e.

$$\begin{array}{rrcl} ((x^1, \ldots, x^n) \times id_{\mathbb{R}^n}) \circ (id_U \times d\varphi_p) : & TM|_U & \xrightarrow{\sim} & \varphi(U) \times \mathbb{R}^n \subseteq \mathbb{R}^{2n} \\ & (p, (v^i \, \partial_i|_p)_{i=1}^n) & \longmapsto & (x^i(p), v^i)_{i=1}^n. \end{array}$$

Def. 3.1.3 *Given a local coordinate system $(U, \varphi \equiv (x^i))$ in $p \in M$, the coordinates defined by $(x^1(p), \ldots, x^n(p), v^1, \ldots, v^n)$, such that $v \in T_pM$ is written as $v = v^j \, \partial_j|_p$, are called the* ***natural local coordinates on the tangent bundle*** *TM.*

As we vary U in an atlas of M, we obtain a covering of TM and the charts can be proven to be compatible, so that they constitute an atlas for TM, see [10] proposition 3.18 page 66 for the technical proof. As a consequence, TM **is a $2n$-dimensional smooth manifold**.

As we will see later, the property of being diffeomorphic to the Cartesian product $U \times \mathbb{R}^n$ that the tangent bundle TM is so important to be one of the conditions included in the definition of a general vector bundle.

The next remark will have a great importance for the general theory of vector bundles: let us concentrate on the local trivialization $TM|_{U_{\alpha\beta}} \cong U_{\alpha\beta} \times \mathbb{R}^n$, where $U_{\alpha\beta} = U_\alpha \cap U_\beta$ is the intersection of two chart domains for M with chart maps φ_α and φ_β, respectively. We know that the compatibility between charts is equivalent to the request that the Jacobian matrix

$J_{\eta_{\alpha\beta}}(x)$ of $\eta_{\alpha\beta}$ evaluated in any $x \in \varphi_\beta(p)$, for all $p \in U_{\alpha\beta}$, is non singular, i.e. it belongs to $\mathrm{GL}(n, \mathbb{R})$. This means that each tangent bundle comes equipped with the following smooth functions

$$\begin{array}{rccl} \tau_{\alpha\beta}: & U_\alpha \cap U_\beta & \longrightarrow & \mathrm{GL}(n, \mathbb{R}) \\ & p & \longmapsto & \tau_{\alpha\beta}(p) = J_{\eta_{\alpha\beta}}(x), \quad x = \varphi_\beta(p), \end{array}$$

which can be easily seen to satisfy the following properties:

$$\begin{cases} \tau_{\alpha\alpha}(p) = I_n, \ \forall\, p \in U_\alpha \\ \tau_{\alpha\beta}(p) = \tau_{\beta\alpha}(p)^{-1}, \ \forall\, p \in U_\alpha \cap U_\beta \\ \tau_{\alpha\beta}(p) \circ \tau_{\beta\gamma}(p) = \tau_{\alpha\gamma}(p), \ \forall\, p \in U_\alpha \cap U_\beta \cap U_\gamma \end{cases},$$

thanks to the corresponding features of the transition functions $\eta_{\alpha\beta}$ between charts.

The functions $\tau_{\alpha\beta}$ are called **transition functions between the local trivializations** of TM given by $TM|_{U_\alpha} \cong U_\alpha \times \mathbb{R}^n$ and $TM|_{U_\beta} \cong U_\beta \times \mathbb{R}^n$.

The importance of the transition functions between the local trivializations is that they permit to construct the manifold structure of a collection of vector spaces attached to points of a manifold in a sense that will be specified more rigorously later in this chapter.

Remark: notice that, in spite of bearing the same name and of being related as described above, the transition functions $\eta_{\alpha\beta} : \varphi_\beta(U_{\alpha\beta}) \subseteq \mathbb{R}^n \to \varphi_\alpha(U_{\alpha\beta}) \subseteq \mathbb{R}^n$ between two charts of M and the transition functions $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(n, \mathbb{R})$ between two local trivializations of TM are very different objects and must not be confused.

Def. 3.1.4 (Global differential) *If $f : M \to N$ is a smooth map between smooth manifolds M and N, then the map $df : TM \to TN$ such that $df|_{T_pM} = df_p$ is called the **global differential** or global tangent to f.*

Theorem 3.1.1 *If $f : M \to N$ is a smooth map, then its global differential $df : TM \to TN$ is a smooth map.*

Proof. It is sufficient to recall eq. (2.27), which gives the local expression of the differential of f in a point $p \in M$ in coordinates as:

$$df_p \left(\left. \frac{\partial}{\partial x^i} \right|_p \right) = \frac{\partial \tilde{f}^j}{\partial x^i}(x) \left. \frac{\partial}{\partial y^j} \right|_{f(p)},$$

where $\tilde{f}^j$ are the component functions of the local expressions of f. Thus, the coordinate representation of df in terms of the natural coordinates of TM and TN is:

$$df(x^1, \ldots, x^n, v^1, \ldots, v^n) = \left(\tilde{f}^1(x), \ldots, \tilde{f}^n(x), \frac{\partial \tilde{f}^1}{\partial x^j}(x) v^j, \ldots, \frac{\partial \tilde{f}^n}{\partial x^j}(x) v^j \right),$$

$x = (x^1, \ldots, x^n)$. The smoothness of f implies that of the coordinate representation. $\square$

The properties of df listed below follow easily from those of the differential in a point.

Theorem 3.1.2 (Properties of the global differential) *Given smooth maps $f : M \to N$ and $g : N \to P$ the following properties hold.*

1. $d(id_M) = id_{TM}$.

2. ***Chain rule for the global differential:***

$$d(g \circ f) = dg \circ df.$$

3. *If f is a diffeomorphism, then $df : TM \to TN$ is a diffeomorphism and $(df)^{-1} = d(f^{-1})$.*

Thanks to 3. it is not ambiguous to write simply df^{-1} for the inverse of the global differential of a smooth function.

3.1.1 The tangent bundle as the state space of a classical mechanical system

A state of a classical mechanical system is given by specifying a configuration, i.e. the position *and* the speed of the system particles at a given time. These data are necessary and sufficient to give the initial conditions to write the system of differential equations given by Newton's second law of motion (or its equivalent Lagrangian or Hamiltonian formulations).

If the configuration space, i.e. the space that contains all the possible positions, is assumed to be a smooth manifold Q, then the state space is the tangent bundle TQ. Thanks to local triviality, if $\dim(Q) = n$, a state at the time t_0 can be locally described via these coordinates:

$$\left(q^1(t_0), \ldots, q^n(t_0), \dot{q}^1(t_0), \ldots, \dot{q}^n(t_0)\right),$$

where $q^i \equiv x^i$ in physical notation, and $q^i(t_0) \equiv \frac{dq^i(t)}{dt}(t_0)$.

3.2 Vector bundles

The tangent bundle is the prototype of a category of smooth bundles called vector bundles, to which this section is dedicated.

Before introducing the formal definition, we stress that the main idea underlying a vector bundle is to construct **a family of vector spaces E_p parameterized by points p of a manifold M** (or, as it is often said, *attached* to these points) **in such a way that these vector spaces fit together to form another manifold**, which is called a vector bundle over M. We can study with the techniques of differential geometry this new manifold, which turns out to carry a richer and more interesting structure than the original one.

The next definition contains all the information needed to 'glue together' the copies of the vector spaces attached to each point of M to form a vector bundle.

Def. 3.2.1 (Vector bundle) *A (real) vector bundle of **rank r** over a smooth manifold M of dimension $n \geqslant r$, called **base space**, is described by the triple (E, M, π), where E is a smooth manifold, called the **total space** of the bundle, and $\pi : E \to M$ is a smooth **surjective** map, such that:*

1. *for all $p \in M$, the **fiber** $E_p := \pi^{-1}(p)$ is a real vector space of dimension r;*

2. *every $p \in M$ admits an open neighborhood $U \subseteq M$ and a diffeomorphism*

$$\chi : E|_U := \pi^{-1}(U) \xrightarrow{\sim} U \times \mathbb{R}^r,$$

called ***local trivialization****, such that the following diagram commutes:*

$$\begin{tikzcd} \pi^{-1}(U) \arrow[rr, "\chi"] \arrow[dr, "\pi"'] & & U \times \mathbb{R}^r \arrow[dl, "pr_1"] \\ & U & \end{tikzcd} \qquad \Longleftrightarrow \qquad pr_1 \circ \chi = \pi.$$

3. *for all $p \in U$, the function $\chi|_p : E_p \xrightarrow{\sim} \{p\} \times \mathbb{R}^r \cong \mathbb{R}^r$ is a linear isomorphism.*

Vector bundles of rank 1 are called ***line bundles****.*

In literature, to denote (or even to define) vector bundles it is common to use either the notation (E, M, π) or $\pi : E \to M$ or simply E, depending on what has to be emphasized. We will follow this tradition.

The simplest example of vector bundle is obtained when the family of vector spaces is constant, i.e., when there is a canonical, fixed, vector space E such that $E_p = E$ for all $p \in M$: in this case there is just one copy of E for each $p \in M$ and these copies fit together to form the vector bundle $M \times E$ over M. Due to the extreme simplicity of this construction, such a vector bundle is called **trivial**.

The **tangent bundle of a manifold M of dimension n is a vector bundle of rank n**. This fundamental example shows that, in general, vector bundles are only *locally* trivial.

Any non globally trivial bundle requires more than one local trivialization, thus it is natural to ask oneself what happens in the overlap of any two local trivializations. The following result shows that, thanks to the requests 2. and 3. in the definition of vector bundle, the composition of two local trivializations on the overlap domain has a particularly simple expression.

Theorem 3.2.1 *Let $\pi : E \to M$ be a vector bundle of rank r over M and suppose that $\chi_1 : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^r$ and $\chi_2 : \pi^{-1}(U_\beta) \to U_\beta \times \mathbb{R}^r$ are two local trivializations of E with non empty intersection $U_\alpha \cap U_\beta \neq \varnothing$. Then, there exists a smooth map*

$$\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(r, \mathbb{R})$$

such that the composition $\chi_1 \circ \chi_2^{-1} : (U_\alpha \cap U_\beta) \times \mathbb{R}^r \to (U_\alpha \cap U_\beta) \times \mathbb{R}^r$ can be written as

$$\chi_1 \circ \chi_2^{-1}(p, v) = (p, \tau_{\alpha\beta}(p)v),$$

i.e. it acts as the identity on the first entry and linearly on the second entry, with the application of the non-singular matrix $\tau_{\alpha\beta}(p) \in \mathrm{GL}(r, \mathbb{R})$ on the vector $v \in \mathbb{R}^r$.

Proof. Thanks to property 2. in the definition of vector bundle, the following diagram commutes (we have not written the restriction of the local trivializations to $\pi^{-1}(U_\alpha \cap U_\beta)$ for notational simplicity).

$$\begin{tikzcd}
(U_\alpha \cap U_\beta) \times \mathbb{R}^r \arrow[d, dashed, "\sigma"'] \arrow[drr, "pr_1"] \arrow[rr, bend left, "\chi_2^{-1}"] & & \pi^{-1}(U_\alpha \cap U_\beta) \arrow[ll, "\chi_2"'] \arrow[rr, "\chi_1"] \arrow[d, "\pi"] & & (U_\alpha \cap U_\beta) \times \mathbb{R}^r \arrow[dll, "pr_1"'] \\
\mathbb{R}^r & & U_\alpha \cap U_\beta & &
\end{tikzcd}$$

This implies that $pr_1 \circ (\chi_1 \circ \chi_2^{-1}) = pr_1$, i.e. $\chi_1 \circ \chi_2^{-1}$ acts as the identity of the first entry, so that the only significant action of the composition $\chi_1 \circ \chi_2^{-1}$ is on the second entry, which belongs to $\mathbb{R}^r$, we denote this action with the smooth map $\sigma : (U_\alpha \cap U_\beta) \times \mathbb{R}^r \to \mathbb{R}^r$ so that

$$\chi_1 \circ \chi_2^{-1}(p, v) = (p, \sigma(p, v)).$$

Property 3. in the definition of vector bundle implies that, for every fixed $p \in U_\alpha \cap U_\beta$, the map $\mathbb{R}^r \ni v \mapsto \sigma(p, v) \in \mathbb{R}^r$ is a linear isomorphism, thus its action can be associated to a non-singular matrix $\tau(p) \in \mathrm{GL}(r, \mathbb{R})$ such that $\sigma(p, v) = \tau(p)v$.

The smoothness of τ is a technical matter left as an exercise. □

Def. 3.2.2 *The smooth map $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(r, \mathbb{R})$ of the previous theorem is called transition function between the local trivializations χ_1 and χ_2 of the vector bundle $\pi : E \to M$.*

As we have seen before, when $E = TM$, the transition functions map every p in $U_\alpha \cap U_\beta$ in the Jacobian matrix evaluated in $\varphi_\beta(p)$ of the transition function $\eta_{\alpha\beta}$ between two charts φ_α and φ_β of M. Moreover, as for the case of the tangent bundle, it is simple to verify that the transition functions $\tau_{\alpha\beta}$ satisfy the so-called **cocycle relations** (identical to those of the tangent bundle, with the only difference of the dimension $r \leqslant n$ for the matrix):

$$\begin{cases} \tau_{\alpha\alpha}(p) = I_r \\ \tau_{\alpha\beta}(p) = \tau_{\beta\alpha}(p)^{-1} \\ \tau_{\alpha\beta}(p)\tau_{\beta\gamma}(p) = \tau_{\alpha\gamma}(p) \end{cases},$$

for all $p \in U_\alpha$, $p \in U_\alpha \cap U_\beta$ and $p \in U_\alpha \cap U_\beta \cap U_\gamma$ respectively.

The importance of the transition functions can be fully understood by the following results, which shows how to provide a vector bundle structure to a collection of vector spaces with fixed dimension attached to the points of a manifold via the transition functions.

Theorem 3.2.2 *Suppose we are given a manifold M and a collection of real vector spaces E_p of fixed dimension r attached to each point $p \in M$. Let then:*

- $E := \bigsqcup_{p \in M} E_p$;
- $\pi : E \to M$, *such that $\pi|_{E_p}$ maps all elements of E_p to p.*

Suppose furthermore that we are given:

1. *an open cover $\{U_\alpha\}_{\alpha \in A}$ of M;*

2. *for each $\alpha \in A$, a bijective map $\chi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^r$ such that $\chi_\alpha|_{E_p}$ is a linear isomorphism between E_p and $\{p\} \times \mathbb{R}^r \cong \mathbb{R}^r$;*

3. *for each $\alpha, \beta \in A$ such that $U_\alpha \cap U_\beta \neq \varnothing$, a smooth map $\tau_{\alpha\alpha} : U_\alpha \cap U_\beta \to \mathrm{GL}(r, \mathbb{R})$ such that $\chi_\alpha \circ \chi_\beta^{-1}(p, v) = (p, \tau_{\alpha\beta}(p)v)$ for all $p \in U_\alpha \cap U_\beta$ and $v \in \mathbb{R}^r$.*

Then there exists a unique topology and smooth structure on E that make it a smooth manifold and a vector bundle of rank r over M, with projection π and smooth local trivializations $\{(U_\alpha, \chi_\alpha)\}$.

The proof is quite technical and we omit it, the interested reader can find it in [10], Lemma 10.6 page 253.

Without this results, in order to give a vector bundle structure on a collection of vector spaces attached to points of a manifold, one should have to build a manifold topology and a smooth structure on their disjoint union, then construct the local trivializations and show that they satisfy all the properties of definition 3.2.1. This is, in general, a much longer and complicated procedure than the one described in the theorem above.

3.2.1 Operations on vector bundles

The operations that can be done on vector spaces can be extended to vector bundles. The key to do that is simply to perform these operations on the fibers, which are vector spaces.

Def. 3.2.3 (Whitney (direct) sum of vector bundles) *Given a smooth manifold M and two vector bundles $\pi_1 : E_1 \to M$ and $\pi_2 : E_2 \to M$ of rank r_1 and r_2, respectively, the **Whitney sum** of E_1 and E_2 is the vector bundle over M of **rank** $r_1 + r_2$ whose fiber at each point $p \in M$ is the direct sum $(E_1)_p \oplus (E_2)_p$.*

It can be proven that, with this definition, we get indeed a vector bundle with total space

$$E_1 \oplus E_2 = \bigsqcup_{p \in M} ((E_1)_p \oplus (E_2)_p).$$

The transition functions for this bundle are $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(r_1 + r_2, \mathbb{R})$, where, for each $p \in M$, $\tau_{\alpha\beta}(p)$ is a block diagonal matrix of the form $\begin{pmatrix} (\tau_1)_{\alpha\beta}(p) & 0 \\ 0 & (\tau_2)_{\alpha\beta}(p) \end{pmatrix}$.

Def. 3.2.4 (Restriction of a vector bundle) *Given a smooth manifold M, a smooth vector bundle $\pi : E \to M$ of rank r and an immersed or embedded subset $S \subset M$, the restriction of E to S is the vector bundle with total space $E_S = \bigsqcup_{p \in S} E_p$ and projection $\pi_S = \pi|_{E_S}$.*

It can be proven that $\pi_S : E_S \to M$ is a smooth vector bundle. As a particular case, the restricted vector bundle $TM|_S$ is called the **ambient vector bundle** over M.

Def. 3.2.5 (Dual of a vector bundle) *Let E be a vector bundle of rank r over the manifold M. Then, its dual vector bundle is:*

$$E^* = \bigsqcup_{p \in M} {E_p}^*.$$

${E_p}^* = \mathrm{Hom}(E_p, \mathbb{R})$ is the dual vector space of E_p. The projection is again the map $\pi : E^* \to M$ such that its restriction to every E_p^* sends its elements to p. The rank of E^* is r. The transition functions are given by $\tau : U \to \mathrm{GL}(r, \mathbb{R})$, $\tau^*(p) = (\tau(p)^{-1})^t$ for all $p \in U$.

When we operate the dualization procedure to the tangent bundle of a manifold, we obtain a very important object, that we discuss in the next section.

3.3 The cotangent bundle over a manifold

Before formalizing the concept of cotangent bundle, let us extend to generic finite-dimensional real vector spaces what stated in Appendix B about the relationship between $\mathbb{R}^n$ equipped with its canonical basis and its dual space $(\mathbb{R}^n)^*$ equipped with the canonical dual basis.

If V is an n-dimensional real vector space, then, by convention, we call its elements $v \in V$ **vectors** and we write them in matrix form as a $n \times 1$ matrix, i.e. as **column vectors**.

Instead, the elements of its dual space V^*, i.e. linear functionals $\omega : V \to \mathbb{R}$, are called **covectors** and they are indicated in matrix form as a $1 \times n$ matrix, i.e. as **row vectors**.

We know that the dual basis $(\varepsilon^1, \dots, \varepsilon^n)$ of $(\mathbb{R}^n)^*$ is associated to the canonical basis $(e_1, \dots, e_n)$ via $\varepsilon^i(e_j) = \delta^i_j$, so that $\varepsilon^i(v^j e_j) = v^i$, the same holds for generic vector spaces and bases.

More precisely, if $(e_1, \dots, e_n)$ is *any basis* of V, the corresponding dual basis of V^*, denoted again $(\varepsilon^1, \dots, \varepsilon^n)$, is defined by:

$$\varepsilon^i(e_j) := \delta^i_j,$$

which implies that, if $v = v^j e_j$, then

$$\varepsilon^i(v^j e_j) = v^j \varepsilon^i(e_j) = v^j \delta^i_j = v^i.$$

So, also for generic vector spaces, the i-th element of the dual basis $(\varepsilon^1, \dots, \varepsilon^n)$ acts simply as the projection on the direction defined by the i-th vector e_i of a fixed basis of V.

A generic covector $\omega \in V^*$ will be written in terms of the basis $(\varepsilon^1, \dots, \varepsilon^n)$ as $\omega = \omega_i \varepsilon^i$, with the components $\omega_i \in \mathbb{R}$ satisfying

$$\omega(e_i) = \omega_j \varepsilon^j(e_i) = \omega_j \delta^j_i = \omega_i,$$

i.e. the components of ω are determined simply by applying it to all the elements of the basis $(e_1, \dots, e_n)$. As a consequence, the action of ω on a generic vector $v = v^j e_j$ is the following: $\omega(v) = (\omega_i \varepsilon^i)(v^j e_j) = \omega_i v^j \varepsilon^i(e_j) = \omega_i v^j \delta^i_j = \omega_i v^i$. The fact that

$$\omega(v) = \omega_i v^i \tag{3.1}$$

explains the convention of writing ω in matrix form as a row vector $1 \times n$ and v as a column vector $n \times 1$, so that the matrix product ωv gives rise to a scalar as in eq. (3.1).

Other useful facts that is worthwhile recalling are listed below:

- **Transpose** (or **dual**) **map**: if $A : V \to W$ is a linear operator between two finite dimensional real vector spaces V and W, the the linear map

$$\begin{array}{rccc} A^t : & W^* & \longrightarrow & V^* \\ & \omega & \longmapsto & A^t(\omega) \end{array}, \qquad \begin{array}{rccc} A^t(\omega) : & V & \longrightarrow & \mathbb{R} \\ & v & \longmapsto & A^t(\omega)(v) := \omega(Av), \end{array}$$

  is called the transpose (or dual) map of A.

- The transpose map verifies $(A \circ B)^t = B^t \circ A^t$ and $(id_V)^t = id_{V^*}$, with obvious meaning of the symbols used.

- The **bidual**, or second dual space of V is $V^{**} := (V^*)^*$. For finite-dimensional vector spaces, V and its bidual V^{**} are naturally isomorphic via the map:

$$\begin{array}{rccc} \xi : & V & \xrightarrow{\sim} & V^{**} \\ & v & \longmapsto & \xi(v) \end{array}, \qquad \begin{array}{rccc} \xi(v) : & V^* & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & \xi(v)(\omega) := \omega(v), \end{array} \tag{3.2}$$

the isomorphism being natural because only the intrinsic elements of the spaces involved have been used to define it and nothing else, in particular without the choice of a basis.

- Because of the natural identification between V and V^{**}, and since ξ is uniquely induced by v, it is custom to omit it in the expression $\xi(v)$, reducing it simply to v, which, with this omission, acquires the double role of vector of V and linear functional over V^*. Due to this, the real number $\omega(v) = \xi(v)(\omega) \equiv v(\omega)$ is often written in a more symmetric-looking way as follows:

$$\langle w, v\rangle := \omega(v) = \xi(v)(\omega) \equiv v(\omega) =: \langle v, w\rangle,$$

called **pairing** between v and ω. The pairing $\langle \varepsilon^i, e_j\rangle = \delta^i_j$ is called **canonical pairing** between bases of $V \cong V^{**}$ and V^*.

The definition of cotangent bundle over a smooth manifold M is identical to that of tangent bundle, the only difference being that the tangent spaces are replaced by their duals.

Def. 3.3.1 *$T^*_pM = \mathrm{Hom}(T_pM, \mathbb{R})$ is the dual of T_pM, called the **cotangent space to M at** p. An element $\omega \in T^*_pM$ is called **cotangent vector** to M in p, or **covector**, or **differential form**.*

Def. 3.3.2 (Cotangent bundle) *The cotangent bundle over M, denoted with T^*M is given by the following disjoint union of cotangent spaces at different $p \in M$:*

$$T^*M = \bigsqcup_{p\in M} T^*_pM = \{(p, \omega) \,:\, p \in M,\ \omega \in T^*_pM\}, \qquad \pi|_{T^*_pM}(\omega) := p.$$

Analogously to what we did for the tangent bundle, we can prove that **the cotangent bundle is manifold of dimension $2n$ and a vector bundle of rank n.**

In the case of tangent spaces, we have seen that the act of fixing a local coordinate system $(U, \varphi \equiv (x^1, \ldots, x^n))$ in $p \in M$ induces the basis $(\partial_1|_p, \ldots, \partial_n|_p)$ of T_pM. We are going to prove that the dual basis of T^*_pM can be built by taking the differential of the coordinate functions $x^i : U \subseteq M \to \mathbb{R}$, $x^i(p) = (\varepsilon^i \circ \varphi)(p) = \varepsilon^i(x^1, \ldots, x^n) = x^i$. Being scalar functions, we must apply eq. (2.17) to get

$$dx^i\big|_p\left(\left.\frac{\partial}{\partial x^j}\right|_p\right) = \left.\frac{\partial}{\partial x^j}\right|_p (x^i) \underset{(2.14)}{=} \frac{\partial(x^i \circ \varphi^{-1})}{\partial x^j}(x) = \frac{\partial(\varepsilon^i \circ \varphi \circ \varphi^{-1})}{\partial x^j}(x) = \frac{\partial \varepsilon^i}{\partial x^j}(x) = \frac{\partial x^i}{\partial x^j} = \delta^i_j,$$

so, the linear functionals of T_pM given by

$$\begin{array}{rccc} dx^i\big|_p : & T_pM & \longrightarrow & T_{x^i}\mathbb{R} \cong \mathbb{R} \\ & \partial_j|_p & \longmapsto & dx^i\big|_p(\partial_j|_p) = \delta^i_j, \end{array}$$

verify the pairing

$$\langle dx^i\big|_p, \partial_j\big|_p\rangle = \delta^i_j$$

which means that they are the dual basis of the coordinate tangent vectors $\partial_j\big|_p$. This justifies the following definition.

Def. 3.3.3 (Coordinate cotangent vectors) *The vectors* $(dx^1\big|_p, \ldots, dx^n\big|_p)$ *are called coordinate cotangent vectors and they form the* ***standard basis of*** T^*_pM *dually associated to the basis of coordinate tangent vectors* $(\partial_1\big|_p, \ldots, \partial_n\big|_p)$ *of* T_pM*.*

Once established $(dx^i\big|_p)$ as the standard basis of T^*_pM, we infer, from what recalled above for a general vector space, that:

- every cotangent vector $\omega \in T^*_pM$ can be expressed as the following linear combination:

$$\omega = \omega_i \; dx^i\big|_p, \qquad \omega_i = \omega(\partial_i\big|_p) \in \mathbb{R},$$

- the action of $dx^i\big|_p$ on the generic tangent vector $v = v^j \; \partial_j\big|_p \in T_pM$ is simply the extraction of the i-th component w.r.t. the coordinate tangent vectors of T_pM:

$$dx^i\big|_p (v^j \; \partial_j\big|_p) = v^i.$$

Analogously as for the tangent bundle, we can define the local coordinates of the cotangent bundle as follows.

Def. 3.3.4 *Given a local coordinate system* $(U, \varphi \equiv (x^i))$ *in* $p \in M$*, the coordinates defined by* $(x^1(p), \ldots, x^n(p), \omega_1, \ldots, \omega_n)$*, such that* $\omega \in T^*_pM$ *is written as* $\omega = \omega_i \; dx^i\big|_p$*, are called the* ***natural local coordinates on the cotangent bundle*** T^*M*.*

We summarize below the results that we obtained so far about the local expressions of a tangent and cotangent vectors.

- Given a local chart (U, φ) of $p \in M$ with local coordinate functions x^i,

$$\begin{array}{rrcl} x^i = \varepsilon^i \circ \varphi : & U \subseteq M & \longrightarrow & \mathbb{R} \\ & p & \longmapsto & \varepsilon^i(\varphi(p)). \end{array}$$

- The basis of T_pM induced by this chart is

$$\left(\partial_1\big|_p, \ldots, \partial_n\big|_p\right).$$

- The dual basis of T^*_pM is

$$\left(dx^1\big|_p, \ldots, dx^n\big|_p\right).$$

- They verify the following pairing: $\langle dx^i\big|_p, \partial_j\big|_p\rangle = \delta^i_j$.

- A generic tangent vector $v \in T_pM$ will be written as:

$$v = v^j \, \partial_j\big|_p \,, \qquad v^j \in \mathbb{R}, \; j = 1, \ldots, n.$$

- A generic cotangent vector (or covector, or differential form) $\omega \in T_pM$ will be written as:

$$\omega = \omega_i \, dx^i\big|_p \,, \qquad \omega_i = \omega(\,\partial_i|_p) \in \mathbb{R}, \; i = 1, \ldots, n.$$

Many times, in the physical and engineering literature, *the specification of the basis is omitted and the position of the indices is used to qualify the object*:

- **tangent vector** $(v^1, \ldots, v^n)$ **- components with indices above**
- **covector or differential form** $(\omega_1, \ldots, \omega_n)$ **- components with indices below**.

In the trivial case of $M = \mathbb{R}^n$ we have at disposal the single chart atlas $(\mathbb{R}^n, \varphi \equiv id_{\mathbb{R}^n})$ which allows us to canonically identify $\left(\partial_1|_p \,, \ldots, \, \partial_n|_p\right)$ with the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$ and $\left(dx^1\big|_p \,, \ldots, \, dx^n|_p\right)$ with the dual canonical basis $(\varepsilon^1, \ldots, \varepsilon^n)$.

3.3.1 A noticeable example of cotangent vector: the differential of a scalar function at a point

Let $\phi : M \to \mathbb{R}$ be a smooth scalar function and $p \in M$. Since $d\phi_p : T_pM \to T_{\phi(p)}\mathbb{R} \cong \mathbb{R}$ is linear, we clearly have that $d\phi_p \in T_p^*M$, i.e. $d\phi_p$ is a cotangent vector to M at p.

Fixed a local chart $(U, \varphi \equiv (x^i))$ in p such that $\varphi(p) = x$, we can of course express $d\phi_p$ as a linear combination of the coordinate cotangent vectors:

$$d\phi_p = \omega_i \, dx^i\big|_p \,,$$

and we know that:

$$\omega_i = d\phi_p \left(\frac{\partial}{\partial x^i}\bigg|_p \right) := \frac{\partial}{\partial x^i}\bigg|_p (\phi) = \frac{\partial(\phi \circ \varphi^{-1})}{\partial x^i}(\varphi(p)) = \frac{\partial \tilde{\phi}}{\partial x^i}(x),$$

where $\tilde{\phi} = \phi \circ \varphi^{-1} : \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ is the local representation of ϕ.

Thus, the explicit expression of the cotangent vector $d\phi_p$ is:

$$\boxed{d\phi_p = \frac{\partial \tilde{\phi}}{\partial x^i}(x) \, dx^i\big|_p} \,. \tag{3.3}$$

3.3.2 Transformation rule for the local components of cotangent vectors

Here we analyze how the components of a cotangent vector change when we change the local coordinates in a point. This is the analog for cotangent vectors of what we have already did in section 2.5.2 for tangent vectors and thus it can be thought as a sort of physicist definition of cotangent vectors.

Suppose that $p \in M$ belongs to the intersection of two local charts $(U, \varphi \equiv (x^i))$ and $(\tilde{U}, \tilde{\varphi} \equiv (\tilde{x}^j))$, then we can decompose $\omega \in T_p^*M$ w.r.t. the basis $(dx^1\big|_p, \ldots, dx^n\big|_p)$ or w.r.t. the basis $(d\tilde{x}^1\big|_p, \ldots, d\tilde{x}^n\big|_p)$ obtaining, respectively,

$$\omega = \omega_i \, dx^i\Big|_p = \tilde{\omega}_j \, d\tilde{x}^j\Big|_p .$$

As we have just seen, the coefficients of the cotangent vectors can be obtained by applying ω on the coordinate tangent vectors:

$$\omega_i = \omega\left(\frac{\partial}{\partial x^i}\bigg|_p\right) \quad \text{and} \quad \tilde{\omega}_j = \omega\left(\frac{\partial}{\partial \tilde{x}^j}\bigg|_p\right).$$

Recall now from eq. (2.29) that $\frac{\partial}{\partial x^i}\big|_p = \frac{\partial \tilde{x}^j}{\partial x^i}(x)\, \frac{\partial}{\partial \tilde{x}^j}\big|_p$, so we get

$$\omega_i = \omega\left(\frac{\partial \tilde{x}^j}{\partial x^i}(x)\, \frac{\partial}{\partial \tilde{x}^j}\bigg|_p\right) = \frac{\partial \tilde{x}^j}{\partial x^i}(x)\, \omega\left(\frac{\partial}{\partial \tilde{x}^j}\bigg|_p\right) = \frac{\partial \tilde{x}^j}{\partial x^i}(x)\, \tilde{\omega}_j.$$

Similarly, by using eq. (2.30) and repeating the calculations above on $\tilde{\omega}_j$ we obtain:

$$\tilde{\omega}_j = \frac{\partial x^i}{\partial \tilde{x}^j}(x)\, \omega_i.$$

As we said in section 2.5.2, in the early days of differential geometry (and still nowadays in the physicist and engineering setting), a tangent vector was interpreted as the assignment of an n-tuple of real numbers associated to each coordinate system following precise transformation rules when we change from one coordinate system to another. It is thus important to compare the transformation rules of the components of a tangent and cotangent vector:

Tangent vectors:

$$\tilde{v}^j = \frac{\partial \tilde{x}^j}{\partial x^i}(x) v^i, \quad v^i = \frac{\partial x^i}{\partial \tilde{x}^j}(\tilde{x}) \tilde{v}^j$$

Cotangent vectors:

$$\tilde{\omega}_j = \frac{\partial x^i}{\partial \tilde{x}^j}(x)\, \omega_i, \quad \omega_i = \frac{\partial \tilde{x}^j}{\partial x^i}(x)\, \tilde{\omega}_j.$$

Since eq. (2.29) for the transformation of the coordinated tangent vectors is a direct consequence of the chain rule, mathematicians considered it as a sort of ‘standard’ for the transformation under change of coordinate system and called covariant, from the Latin prefix *co-*, which means *with*, so that covariant means that an object ‘vary with’ the standard transformation rule.

It can be seen that cotangent vectors follow the standard transformation rule, eq. (2.29), while tangent vectors follow the opposite rule. For this reason, it is still customary to say that:

- **cotangent** vectors are **covariant** vectors;
- **tangent** vectors are **contravariant** vectors.

Despite the same nomenclature, this has nothing to do with covariant and contravariant functors of category theory.

3.4 Local and global sections of a vector bundle

In Physics, when we talk about a vector field we mean a vector attached to each point of a certain region in space. This concept can be made rigorous in differential geometry thanks to the definition of sections of vector bundles. Being quite simple, we will first introduce the abstract concept of section on a general vector bundle and then we will specialize it on the tangent and cotangent bundles.

Let us consider a vector bundle $\pi : E \to M$ over the smooth manifold M and a neighborhood U of a point $p \in M$. The most natural vectors associated to p are those belonging to the fiber over it, i.e. $\pi^{-1}(p) = E_p$, because each $v \in E_p$ projects on p via π. Thus, a function that associates points of U to vectors belonging the fibers over them is also a natural object. Of course, to be able to perform differential calculus over this object, we require it to be smooth, i.e. we demand that the vector assignment is smooth when we pass from one point to another.

The definition of local section gives a mathematical formalization to what just said.

Def. 3.4.1 (Local section or local vector field) *A local section (or a local vector field) of E on an open set $U \subseteq M$ is a smooth function $\sigma : U \to E$ such that $\pi \circ \sigma = \iota_U$, i.e. such that the following diagram commutes:*

$$\begin{tikzcd} & E \arrow[d, "\pi"] \\ U \arrow[ur, "\sigma"] \arrow[r, hook, "\iota_U"'] & M \end{tikzcd}$$

i.e. $\pi \circ \sigma = \iota_U$, where ι is the canonical inclusion of U in M.

Notice that the definition contains exactly the information that we wanted to formalize, in fact, thanks to the local triviality of E, $\sigma(p) = (p, v) \in U \times E_p$, so

$$(\pi \circ \sigma)(p) = \pi(p, v) = p = \iota(p).$$

In this way, we do not attach p to any vector, but to a vector v belonging to the fiber over p, which is called the *significant part* of the section σ, because it is the only information that allows us to distinguish it from another section on U.

The set of all sections of E on U is denoted with the following symbol:

$$\boxed{\Gamma(U, E) = \{\sigma : U \to E,\ \pi \circ \sigma = \iota_U\}}\ .$$

$\Gamma(U, E)$ **is an Abelian group** w.r.t. the sum of sections on U defined as follows: given two local sections $\sigma_1, \sigma_2 \in \Gamma(U, E)$, with $\sigma_1(p) = (p, v_1|_p)$ and $\sigma_2(p) = (p, v_2|_p)$, then

$$(\sigma_1 + \sigma_2)(p) := (p, v_1|_p + v_2|_p),$$

which makes perfect sense because both $v_1|_p$ and $v_2|_p$ belong to the same vector space $\pi^{-1}(p)$, so we can add them together meaningfully.

If it is possible to define σ on the entire manifold M, then we get the global sections.

Def. 3.4.2 (Global section or global vector field) *A global section (or a global vector field) of E on M is a smooth function $\sigma : M \to E$ such that $\pi \circ \sigma = id_M$, i.e. such that the following diagram commutes.*

$$\begin{tikzcd} & E \arrow[d, "\pi"] \\ M \arrow[ur, "\sigma"] \arrow[r, "id_M"'] & M \end{tikzcd}$$

Convention: without further specification, a section on a vector bundle will be considered as global.

The set of all sections of E on M is denoted with the following symbol:

$$\boxed{\Gamma(E) = \{\sigma : M \to E,\ \pi \circ \sigma = id_M\}}\ .$$

Noticeable examples of sections, or vector fields, are obtained by considering $E = TM$ and $E = T^*M$, the tangent and cotangent bundle of M, respectively.

3.4.1 Tangent vector fields

In this and in the next subsection we will omit the adjective local or global, since the definitions and results hold for both situations, with evident adjustments.

Def. 3.4.3 *A (tangent) vector field is a smooth assignment $X : M \to TM$, to each point $p \in M$, of a tangent vector to M at p, i.e. $\pi \circ X = id_M$,*

$$\begin{array}{rccl} X: & M & \longrightarrow & TM \\ & p & \longmapsto & X(p) \equiv (p, X_p), \end{array}$$

with $X_p \in T_pM$, the significant part of the (tangent) vector field X.

Due to its importance, $\Gamma(TM)$, the set of all sections of TM, is denoted with a particular symbol:

$$\boxed{\mathfrak{X}(M) = \tau(M) = \{\sigma : M \to TM,\ \pi \circ \sigma = id_M\}}\ . \tag{3.4}$$

It is custom to omit 'tangent' and write only vector field when it is clear that the vector bundle that we are considering is TM. We will use this convention.

The basic properties of $\mathfrak{X}(M)$ are listed in the following result.

Theorem 3.4.1 *The following assertions hold.*

- *$\mathfrak{X}(M)$ is a real vector space under point-wise addition and scalar multiplication, i.e.*

  $$(aX + bY)_p := aX_p + bY_p, \qquad X, Y \in \mathfrak{X}(M),\ a, b \in \mathbb{R}.$$

  The 0 element of $\mathfrak{X}(M)$ is the null vector field, that attaches to any $p \in M$ the 0 tangent vector of T_pM.

- *If $f \in \mathscr{C}^\infty(M)$ and $X \in \mathfrak{X}(M)$, then $fX : M \to TM$ defined as:*
$$(fX)_p := f(p)X_p \qquad \forall p \in M,$$
is a vector field.
- *$\mathfrak{X}(M)$ is a module over the ring $\mathscr{C}^\infty(M)$.*

Using the natural local coordinates of TM for every coordinate chart $(U, (x^i))$ we can write, for every $p \in M$,
$$X(p) = (p, X^i(p)\, \partial_i|_p),$$
where the coefficients $X^i(p) \in \mathbb{R}$, in general, vary with p. This implies the existence of n functions $X^i : U \subseteq M \to \mathbb{R}$, called **component functions** of the vector field $X \in \mathfrak{X}(M)$ in the chart $(U, (x^i))$ such that, for all $p \in M$:
$$X_p = X^i(p) \left.\frac{\partial}{\partial x^i}\right|_p,$$
which is an equation involving tangent vectors of T_pM. Using the fact that $\mathfrak{X}(M)$ is a module over the ring $\mathscr{C}^\infty(M)$, this relationship can be written also as an equation involving vector fields, i.e.
$$X = X^i \frac{\partial}{\partial x^i} \equiv X^i \partial_i,$$
where
$$\begin{array}{cccl} \frac{\partial}{\partial x^i} : & U & \longrightarrow & TM \\ & p & \longmapsto & \frac{\partial}{\partial x^i}(p) := (p, \left.\frac{\partial}{\partial x^i}\right|_p) \equiv (p, \partial_i|_p), \end{array}$$
is called the **i-th coordinate tangent vector field**.

It is clear that the restriction of a vector field to a chart domain U is smooth if and only if the component functions w.r.t. that domain are smooth.

3.4.2 1-forms or cotangent vector fields

Sections of the cotangent bundle are a fundamental object in differential geometry and its applications.

Def. 3.4.4 (1-form or cotangent vector field) *A 1-form or cotangent vector field, is a smooth assignment $\omega : M \to T^*M$, to each point $p \in M$, of a cotangent vector to M at p, i.e. $\pi \circ \omega = id_M$,*
$$\begin{array}{cccl} \omega : & M & \longrightarrow & T^*M \\ & p & \longmapsto & \omega(p) \equiv (p, \omega_p), \end{array}$$
with $\omega_p \in T^*_pM$, the significant part of the 1-form ω.

The set of all sections of T^*M is denoted with the symbols
$$\boxed{\mathfrak{X}^*(M) = \Lambda(M) = \Omega(M)}\,. \tag{3.5}$$

The easiest example of 1-form is the differential of a smooth scalar function $\phi \in \mathscr{C}^\infty(M)$:

$$\begin{array}{rccl} d\phi: & M & \longrightarrow & T^*M \\ & p & \longmapsto & d\phi(p) \equiv (p, d\phi_p). \end{array}$$

As $\mathfrak{X}(M)$, also $\Lambda(M)$ is a real vector space w.r.t. point-wise operations and a module over the ring $\mathscr{C}^\infty(M)$ w.r.t. the operation

$$(f\omega)_p := f(p)\omega_p, \qquad \forall f \in \mathscr{C}^\infty(M),\ p \in M.$$

Using the natural local coordinates of T^*M for every coordinate chart $(U, (x^i))$ we can write, for every $p \in M$,

$$\omega(p) = (p, \omega_i(p)\ dx^i\big|_p),$$

where the coefficients $\omega_i(p) \in \mathbb{R}$, in general, vary with p. Repeating the same arguments used for vector fields, we have that there are n functions $\omega_i : U \subseteq M \to \mathbb{R}$, called **component functions** of the 1-form $\omega \in \Lambda(M)$ in the chart $(U, \varphi \equiv (x^i))$ such that, for all $p \in M$:

$$\omega_p = \omega_i(p)\ dx^i\big|_p\,,$$

an equation involving cotangent vectors of T_p^*M that can be written also as an equation involving 1-forms, i.e.

$$\omega = \omega_i\ dx^i\big|_p \equiv \omega_i dx^i,$$

where

$$\begin{array}{rccl} dx^i: & U & \longrightarrow & T^*M \\ & p & \longmapsto & dx^i(p) := (p,\, dx^i\big|_p), \end{array}$$

is called the ***i*-th coordinate 1-form**.

Also in this case, it is clear that the restriction of a 1-form to a chart domain U is smooth if and only if the component functions w.r.t. that domain are smooth.

In the particular case where $\omega = d\phi$, $\phi \in \mathscr{C}^\infty(M)$, thanks to (3.3) it holds $d\phi_p = \frac{\partial\tilde{\phi}}{\partial x^i}(x)\ dx^i\big|_p$, so we can write

$$d\phi = \frac{\partial\tilde{\phi}}{\partial x^i} dx^i \tag{3.6}$$

i.e. the component functions of the differential of a scalar function are the partial derivatives of the local representation $\tilde{\phi} = \phi \circ \varphi^{-1}$ of the scalar function itself w.r.t. the chart.

If $M = \mathbb{R}^n$, we can use the single chart atlas $(\mathbb{R}^n, id_{\mathbb{R}^n})$ and $\tilde{\phi} = \phi$, thus the previous formula reduces to the well-known formula for the total derivative of ordinary differential calculus in $\mathbb{R}^n$.

Thanks to what just discussed, we get a criterion to decide weather a smooth scalar function on a manifold is constant or not, for the proof see [10] Proposition 11.22, page 282.

Theorem 3.4.2 (Criterion for constant scalar functions on a manifold) *$\phi \in \mathscr{C}^\infty(M)$ is constant on each connected components of M if and only if $d\phi = 0$.*

This result suggests that we can interpret $d\phi$ as a 'small' change of $\phi \in \mathscr{C}^\infty(M)$ generated by small changes of its variables as in ordinary calculus in $\mathbb{R}^n$.

In fact, since we are interested in small changes, we can fix any point $p \in M$ and a local chart $(U, \varphi \equiv (x^i))$ in p, with $x = \varphi(p)$, so that we can associate ϕ to its local representation $\tilde{\phi} = \phi \circ \varphi^{-1} : \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ and consider $\Delta\phi := \tilde{\phi}(x+v) - \tilde{\phi}(x)$, where the norm of $v \in \mathbb{R}^n$ is sufficiently small. By the smoothness of ϕ we can apply a Taylor expansion in x for $\tilde{\phi}$ and write:

$$\Delta\phi \approx \frac{\partial \tilde{\phi}}{\partial x^i}(x) v^i$$

but we know that the coordinate cotangent vectors $dx^i\big|_p$ act as component extractors on vectors, so $v^i = dx^i\big|_p(v)$ and thus

$$\Delta\phi \approx \frac{\partial \tilde{\phi}}{\partial x^i}(x)\; dx^i\big|_p(v) = d\phi_p(v).$$

From this computation, we infer that $d\phi$ encodes the first-order variation of $\phi \in \mathscr{C}^\infty(M)$ in an intrinsic, coordinate-free way, on every manifold M.

Chapter 4

Tensor calculus

> It is well known that around the turn of the century Riemann's theory of metrical continua, which had fallen so completely into oblivion, was revivified and deepened by Ricci and Levi-Civita; and that the work of these two decisively advanced the formulation of general relativity.
>
> ALBERT EINSTEIN, 1955

Tensor calculus, originated by problems in elasticity and formalized by G. Ricci-Curbastro and T. Levi-Civita in 1900 [19], is omnipresent in differential geometry and its applications. In this chapter we give a very basic introduction to this topic, first discussing the tensor product for vector spaces and then specializing these concepts on the fibers of a vector bundle.

4.1 Tensor product of vector spaces and vectors

Let V, W be two real vector spaces of finite dimension m and n, respectively, $V^* = \mathrm{Hom}(V, \mathbb{R})$, $W^* = \mathrm{Hom}(W, \mathbb{R})$ their dual spaces and let $\mathrm{Bil}(V \times W)$ the vector space of bilinear forms $g : V \times W \to \mathbb{R}$ on $V \times W$, i.e. linear in one variable when the other is kept fixed.

The most natural way to build a bilinear form $g : V \times W \to \mathbb{R}$ is by considering the product of two linear forms $\varphi \in V^*$ and $\psi \in W^*$, i.e. $g(v, w) = \varphi(v)\psi(w)$, in fact, by definition of bilinearity, if we fix one variable, say w, then $\psi(w)$ becomes simply a real coefficient and the linearity of φ in v guarantees the linear behavior of g w.r.t. v; of course the same holds if we exchange the role of v and w and that of φ and ψ, thus guaranteeing the bilinearity of g.

The bilinear form arising in this way is called **tensor product of φ and ψ** and denoted with $\varphi \otimes \psi$:

$$\begin{array}{rccl} \varphi \otimes \psi : & V \times W & \longrightarrow & \mathbb{R} \\ & (v, w) & \longmapsto & \boxed{\varphi \otimes \psi(v, w) := \varphi(v)\psi(w)}\ . \end{array} \tag{4.1}$$

For example, if $V = W = \mathbb{R}^2$, $\varphi = \varepsilon^1$ and $\psi = \varepsilon^2$, where ε^i is the i-th element of the canonical basis of $(\mathbb{R}^2)^*$, then, for any $v, w \in \mathbb{R}^2$ such that $v = (v^1, v^2)$ and $w = (w^1, w^2)$, we have $\varepsilon^1 \otimes \varepsilon^2(v, w) = v^1 w^2$.

The naturalness of the bilinearity of the tensor product of linear forms raises the following question: is it possible to express *all* bilinear forms on $V \times W$ as tensor product of linear forms on V and W? The answer is affirmative (for the proof see [10], proposition 12.10 page 311), in the sense that all bilinear forms can be written as tensor product of two linear forms or as a linear combination of them, which, by linearity, can of course be re-defined as a tensor product of two linear forms. Thus, if we define **the tensor product of V^* and W^*** as the vector space (w.r.t. the point-wise linear operations)

$$\boxed{V^* \otimes W^* := \{\varphi \otimes \psi \mid \varphi \in V^*,\ \psi \in W^*\}}\ ,$$

we have the canonical identification

$$\boxed{V^* \otimes W^* \cong\ \mathrm{Bil}(V \times W)}\ .$$

It is straightforward to verify the following formulae, valid for each $\varphi_1, \varphi_2 \in V^*$, $\psi_1, \psi_2 \in W^*$, $a_1, a_2, b_1, b_2 \in \mathbb{R}$:

$$(a_1\varphi_1 + a_2\varphi_2) \otimes \psi = a_1\varphi_1 \otimes \psi + a_2\varphi_2 \otimes \psi, \qquad \varphi \otimes (b_1\psi_1 + b_2\psi_2) = \varphi \otimes b_1\psi_1 + b_2\psi_2.$$

The following proposition is extremely important in Riemannian geometry.

Theorem 4.1.1 *Let $(e_1, \ldots, e_m)$ and $(f_1, \ldots, f_n)$ be a basis of V and W, respectively, with dual bases $(\varphi^1, \ldots, \varphi^m)$ and $(\psi^1, \ldots, \psi^n)$ of V^* and W^*, respectively. Then,*

$$(\varphi^i \otimes \psi^j)_{\substack{i=1,\ldots,m\\ j=1,\ldots,n}} \text{ is a basis for } V^* \otimes W^*,$$

so that every $g \in V^* \otimes W^* \cong \mathrm{Bil}(V \times W)$ can be uniquely written as

$$\boxed{g = g_{ij}\varphi^i \otimes \psi^j}\ .$$

Moreover, the real coefficients g_{ij} are determined by the action of g on the vectors of the basis of V and W, i.e.

$$\boxed{g_{ij} = g(e_i, f_j)}\,,$$

for all fixed $i = 1, \ldots, m,\ j = 1, \ldots, n$.

Proof. Let us apply a generic $g \in \mathrm{Bil}(V \times W)$ to two arbitrary vectors $v \in V$ and $w \in W$:

$$g(v, w) = g(v^i e_i, w^j f_j) \underset{\text{(bilinearity)}}{=} v^i w^j g(e_i, f_j) \equiv g_{ij} v^i w^j,$$

for all fixed $i = 1, \ldots, m,\ j = 1, \ldots, n$.

Let us now apply $g_{ij}\varphi^i \otimes \psi^j$ to the same couple of vectors:

$$g_{ij}\varphi^i \otimes \psi^j(v, w) = g_{ij}\varphi^i(v^i e_i)\psi^j(w^j f_j) = g_{ij} v^i \cancelto{1}{\varphi^i(e_i)}\ w^j \cancelto{1}{\psi^j(f_j)} = g_{ij} v^i w^j = g(v, w),$$

hence, since v and w are arbitrary, we have $g = g_{ij}\varphi^i \otimes \psi^j$ with $g_{ij} = g(e_i, f_j)$. This shows that $(\varphi^i \otimes \psi^j)_{\substack{i=1,\ldots,m\\ j=1,\ldots,n}}$ spans $V^* \otimes W^* \cong \mathrm{Bil}(V \times W)$.

The linear independence follows trivially from the consideration that if the linear combination $g_{ij}\varphi^i \otimes \psi^j$ is the null bilinear form, then $0 = g_{ij}\varphi^i \otimes \psi^j(e_i, f_j) = g_{ij}$. □

As an immediate consequence of this theorem we have that $\dim(V^* \otimes W^*) = mn$.

The results just discussed can be extended *verbatim* to a finite set of vector spaces, the only difficulty being represented by a heavier notation. We obtaining the following canonical identification:

$$\bigotimes_{i=1}^{p} V_i^* \cong \ \mathrm{Mul}(\bigtimes_{i=1}^{p} V_i),$$

where $\mathrm{Mul}(\bigtimes_{i=1}^{p} V_i)$ is the vector space of p-multilinear forms, i.e. linear in each one of the p variables separately, when all the other $p-1$ are kept fixed.

Up to now we have considered tensor products of linear forms and dual vector spaces, we can define tensor product of vectors and vector spaces by considering the natural isomorphism between a finite dimensional real vector space V and its bidual $V^{**} = \mathrm{Hom}(V^*, \mathbb{R})$:

$$\begin{array}{ccc} V & \xrightarrow{\sim} & V^{**} \\ v & \longmapsto & \alpha_v, \end{array} \quad , \qquad \begin{array}{cccc} \alpha_v : & V^* & \longrightarrow & \mathbb{R} \\ & \varphi & \longmapsto & \alpha_v(\varphi) = \varphi(v) \equiv \langle v, \varphi \rangle. \end{array}$$

By exchanging the role of V, W and V^*, W^* and thanks to the identification just recalled, we can define the **tensor product of two vectors** $v \in V$ **and** $w \in W$ as follows:

$$\begin{array}{cccc} v \otimes w : & V^* \times W^* & \longrightarrow & \mathbb{R} \\ & (\varphi, \psi) & \longmapsto & v \otimes w(\varphi, \psi) = \langle v, \varphi \rangle \langle w, \psi \rangle, \end{array}$$

i.e. the tensor product of two vectors $v, w \in V$ is a bilinear form on $V^* \times W^*$.

Thus, **both the tensor product of two linear forms and two vectors lead to bilinear forms**, what changes is just their domain.

As before, if we define **the tensor product of** V **and** W as the vector space (w.r.t. the point-wise linear operations)

$$\boxed{V \otimes W := \{v \otimes w \mid v \in V, \ w \in W\}} \ ,$$

we have the canonical identification

$$\boxed{V \otimes W \cong \ \mathrm{Bil}(V^* \times W^*)} \ .$$

Following the same reasoning as before, it can be proven that $(e_1, \ldots, e_m)$ and $(f_1, \ldots, f_n)$ are basis of V and W, respectively, then

$$(e_i \otimes f_j)_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \text{ is a basis for } V \otimes W,$$

so $\dim(V \otimes W) = mn$ and a generic element $g \in V \otimes W \cong \mathrm{Bil}(V^* \times W^*)$ can be written as

$$\boxed{g = g^{ij} e_i \otimes f_j} \ ,$$

with

$$\boxed{g^{ij} = g(\varphi^i, \psi^j)} \ ,$$

for all fixed $i = 1, \ldots, m, \ j = 1, \ldots, n$, where $(\varphi^i)_{i=1}^m$ and $(\psi^i)_{i=1}^n$ are the dual bases of $(e_i)_{i=1}^m$ and $(f_j)_{i=1}^m$.

The formulae:
$$\begin{cases} g = g^{ij} v_i \otimes w_j & v \in V,\ w \in W \\ g = g_{ij} \varphi^i \otimes \psi^j & \varphi \in V^*,\ \psi \in W^* \end{cases}$$
are vastly used and the position of the indices reveals if we are dealing with the tensor product of vectors or linear forms. The real coefficients g^{ij} and g_{ij} can be organized in a $m \times n$ matrix, for this reason the tensor product is often (erroneously) defined as a matrix.

As in the previous discussion, we can generalize these results to any finite number of finite-dimensional vector spaces by obtaining:
$$\bigotimes_{i=1}^{p} V_i \cong \ \mathrm{Mul}(\bigtimes_{i=1}^{p} V_i^*).$$
Useful canonical isomorphisms are listed below for finite-dimensional real vector spaces:
$$V \otimes W \cong W \otimes V, \qquad \text{symmetry of } \otimes$$
$$(V_1 \otimes V_2) \otimes V_3 \cong V_1 \otimes (V_2 \otimes V_3), \qquad \text{associativity of } \otimes$$
$$(V_1 \oplus V_2) \otimes W \cong (V_1 \otimes W) \oplus (V_2 \otimes W), \qquad \text{distributivity of } \otimes \text{ w.r.t. } \oplus,$$
more generally,
$$\bigoplus_{i=1}^{r} V_i \ \otimes \ \bigoplus_{i=1}^{s} W_i \ \cong \ \bigoplus_{\substack{i=1,\dots,r \\ j=1,\dots,s}} V_i \otimes W_j.$$
Finally, it holds that
$$V^* \otimes W^* \cong (V \otimes W)^*, \tag{4.2}$$
where the isomorphism is canonical.

If $(V, \langle\, ,\, \rangle_V)$ and $(W, \langle\, ,\, \rangle_W)$ are inner product spaces, then it is possible to endow the tensor product $V \otimes W$ with an inner product in a natural way as follows:
$$\boxed{\langle v_1 \otimes w_1, v_2 \otimes w_2 \rangle_{V\otimes W} := \langle v_1, v_2 \rangle_V \, \langle w_1, w_2 \rangle_W}\,.$$
From this definition, it follows easily that if $(e_i)_{i=1}^m$ is an orthonormal basis of V w.r.t. the inner product $\langle\, ,\, \rangle_V$ and $(f_j)_{j=1}^n$ is an orthonormal basis of W w.r.t. the inner product $\langle\, ,\, \rangle_W$, then $(e_i \otimes f_j)_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ is an orthonormal basis of $V \otimes W$ w.r.t. the inner product $\langle\, ,\, \rangle_{V\otimes W}$.

4.2 Tensor product of linear transformations

We can consider the tensor product of the dual of a vector space with another vector space or vice-versa. The result in this case is particularly important and it is underlined in the following result.

Theorem 4.2.1 *For any couple of finite-dimensional real vector spaces V and W the following natural identification holds:*
$$\boxed{V^* \otimes W \cong \mathrm{Hom}(V, W)}\ ,$$

where the natural isomorphism between the two spaces is defined on the generic basis element[1] $\varphi \otimes w$ *of* $V^* \otimes W$ *by:*

$$F: \begin{array}{ccc} V^* \otimes W & \longrightarrow & \mathrm{Hom}(V, W) \\ \varphi \otimes w & \longmapsto & F(\varphi \otimes w) \equiv F_{\varphi,w}, \end{array} \qquad F_{\varphi,w}: \begin{array}{ccc} V & \longrightarrow & W \\ v & \longmapsto & F_{\varphi,w}(v) = \varphi(v)w. \end{array} \tag{4.3}$$

Proof. Let $(v_1, \ldots, v_n)$ be a basis of V, $(v^1, \ldots, v^n)$ the dual basis of V^*, such that $v^i v_j = \delta^i_j$, and let $(w_1, \ldots, w_m)$ be a basis of W. These bases induce the basis $(v^i \otimes w_j)_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ of $V^* \otimes W$. The theorem will be proven if we show that F sends this basis to a basis of $\mathrm{Hom}(V, W)$.

To this aim, let us make the action of F explicit: if we apply $F(v^i \otimes w_j) \equiv F_{v^i, w_j} \equiv F^i_j \in \mathrm{Hom}(V, W)$ to an element v_k of the basis of V fixed above, then, thanks to eq. (4.3) we get

$$F^i_j(v_k) = v^i(v_k)w_j = \delta^i_k w_j. \tag{4.4}$$

Thus, we have to prove that the linear maps $(F^i_j)_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ form a basis of $\mathrm{Hom}(V, W)$.

For this, it is sufficient to consider an arbitrary $L \in \mathrm{Hom}(V, W)$ and represent it as a matrix $A = (a^j_k)$ w.r.t. the bases of V and W that we have fixed: by definition of matrix associated to a linear map, the coefficients a^j_i verify $L(v_i) = a^j_i w_j$ for every vector v_k of the basis of V. The linear combination of the maps F_{v^i, w_j} with the coefficients a^j_i, i.e. $a^j_i F^i_j$, is an element of $\mathrm{Hom}(V, W)$, let us apply this map on the generic vector v_k of the basis of V and see what we get:

$$(a^j_i F^i_j)(v_k) \underset{\text{linearity}}{=} a^j_i F^i_j(v_k) \underset{\text{eq. (4.4)}}{=} a^j_i \delta^i_k w_j = a^j_k w_j \underset{\text{def. of } L}{=} L(v_k),$$

we see that the action of the arbitrary linear map $L \in \mathrm{Hom}(V, W)$ on the arbitrary vector v_k of the basis of V is obtained by linear combination of the action of the linear maps F^i_j, hence they form a basis for $\mathrm{Hom}(V, W)$. □

Let us consider the particular case of $\mathrm{Hom}(V, V) = L(V)$ and $\mathrm{Hom}(W, W) = L(W)$, then the result just proven shows us that $V^* \otimes V \cong \mathrm{Hom}(V, V) = L(V)$ and also that $W^* \otimes W \cong \mathrm{Hom}(W, W) = L(W)$. We can perform the tensor product of the vector spaces $L(V)$ and $L(W)$, whose elements are linear transformations, this will allow us defining the tensor product of linear maps.

First of all, by using associativity and symmetry, we have:

$$L(V) \otimes L(W) \cong (V^* \otimes V) \otimes (W^* \otimes W) \cong (V^* \otimes W^*) \otimes (V \otimes W),$$

hence, thanks to eq. (4.2), we can write

$$L(V) \otimes L(W) \cong (V \otimes W)^* \otimes (V \otimes W),$$

finally, theorem 4.2.1 implies

$$\boxed{L(V) \otimes L(W) \cong L(V \otimes W)}.$$

The canonical isomorphism underlying the last identification is the map that takes the linear maps $T \in L(V)$, $S \in L(W)$ and brings them to the linear map $T \otimes S \in L(V \otimes W)$ defined as follows:

$$T \otimes S: \begin{array}{ccc} V \otimes W & \longrightarrow & V \otimes W \\ v \otimes w & \longmapsto & (T \otimes S)(v \otimes w) := Tv \otimes Sw. \end{array} \tag{4.5}$$

[1] and then, of course, extended by linearity to the whole space.

Def. 4.2.1 (Tensor product of linear transformations) *Given two finite-dimensional* $\mathbb{R}$-*vector spaces* V, W *and two linear map* $T \in L(V)$ *and* $S \in L(W)$, *their tensor product is the linear map* $T \otimes S \in L(V \otimes W)$ *defined in* (4.5).

Given the bases $e \equiv (e_1, \dots, e_m)$ and $f \equiv (f_1, \dots, f_n)$ of V and W, respectively, we can associate a $m \times m$ matrix $A^T = (a^T_{ij})^m_{i,j=1}$ to T and a $n \times n$ matrix $A^S = (a^S_{kl})^n_{k,l=1}$ to S. By direct computation, it can be verified that the matrix associated to $T \otimes S$ w.r.t. the basis $e \otimes f$ is the so-called **Kronecker product** of the matrices A^T and A^S, defined as follows:

$$A^{T\otimes S} = A^T \otimes A^S = \begin{pmatrix} a^T_{11}A^S & \cdots & a^T_{1m}A^S \\ \vdots & \ddots & \vdots \\ a^T_{m1}A^S & \cdots & a^T_{mm}A^S \end{pmatrix}.$$

It is useful to notice that the trace of $A^{T\otimes S}$ is the product of the traces of A^T and A^S, i.e.

$$\boxed{\mathrm{Tr}\left(A^{T\otimes S}\right) = \mathrm{Tr}\left(A^T\right)\,\mathrm{Tr}\left(A^S\right)}$$

in fact:

$$\mathrm{Tr}\left(A^{T\otimes S}\right) = a^T_{11}(a^S_{11}+\cdots+a^S_{nn})+\cdots+a^T_{mm}(a^S_{11}+\cdots+a^S_{nn}) = (a^T_{11}+\cdots+a^T_{mm})(a^S_{11}+\cdots+a^S_{nn}).$$

4.3 Covariant and contravariant tensors. Tensor algebra of a vector space

As we have seen in the previous section, starting from a real vector space V of finite dimension, we can build many other spaces via tensor product. These spaces are given by *multilinear functions defined on copies of* V *and* V^*.

Here we introduce a compact notation for them which is currently used in textbooks about tensor calculus.

- $T^0_0(V) = T^0(V) = T_0(V) = \mathbb{R}$
- $T^1_0(V) = T^1(V) = V$
- $T^p_0(V) = T^p(V) = V\underbrace{\otimes\cdots\otimes}_{p \text{ times}}V \quad \Longrightarrow \quad \dim(T^p_0(V)) = n^p$
- $T^0_1(V) = T_1(V) = V^*$
- $T^0_q(V) = T_q(V) = V^*\underbrace{\otimes\cdots\otimes}_{q \text{ times}}V^* \quad \Longrightarrow \quad \dim(T^0_q(V)) = n^q$
- $T^p_q(V) = T^p(V) \otimes T_q(V) = V\underbrace{\otimes\cdots\otimes}_{p \text{ times}}V \otimes V^*\underbrace{\otimes\cdots\otimes}_{q \text{ times}}V^*$
- $T^\bullet(V) = \bigoplus_{p\geqslant 0} T^p(V)$

- $T_\bullet(V) = \bigoplus_{q\geqslant 0} T_q(V)$
- $T(V) = \bigoplus_{p,q\geqslant 0} T_q^p(V)$, is called **tensor algebra** of V.

Let us fix our attention on $T_q^p(V)$.

Def. 4.3.1 *An element $t \in T_q^p(V)$ is called a* p***-contravariant and*** q***-covariant tensor on*** V*.*

t is nothing but **a multilinear form** of the type:

$$t: \quad V^* \underbrace{\times \cdots \times}_{p \text{ times}} V^* \times V \underbrace{\times \cdots \times}_{q \text{ times}} V \quad \longrightarrow \quad \mathbb{R}.$$

To understand its action, let us fix as usual a basis $(v_1, \ldots, v_n)$ of V and the dual basis $(v^1, \ldots, v^n)$ of V^*, then:

- $(v_{i_1} \otimes \cdots \otimes v_{i_p})$ is a basis of $V \underbrace{\otimes \cdots \otimes}_{p \text{ times}} V$, with independent indices $i_1, \ldots, i_p = 1, \ldots, n$;
- $(v^{j_1} \otimes \cdots \otimes v^{j_q})$ is a basis of $V^* \underbrace{\otimes \cdots \otimes}_{q \text{ times}} V^*$, with independent indices $j_1, \ldots, j_q = 1, \ldots, n$;
- $\left(v_{i_1} \otimes \ldots v_{i_p} \otimes v^{j_1} \otimes \cdots \otimes v^{j_q}\right)_{\substack{i_1,\ldots,i_p=1,\ldots,n \\ j_1,\ldots,j_q=1,\ldots,n}}$ is a basis of $T_q^p(V)$.

Since we have $p+q$ vectors in the basis of $T_q^p(V)$, each of which is parameterized by an index whose variability is between 1 and n, we have $\underbrace{n \cdots n}_{(p+q) \text{ times}} = n^{p+q}$, so

$$\boxed{\dim(T_q^p(V)) = n^{p+q}}\,.$$

The generic decomposition of the tensors of $T_q^p(V)$ on the basis previously obtained is:

$$\boxed{t = a^{i_1,\ldots,i_p}{}_{j_1,\ldots,j_q} v_{i_1} \otimes \ldots v_{i_p} \otimes v^{j_1} \otimes \ldots v^{j_q}}\,,$$

with $a^{i_1,\ldots,i_p}{}_{j_1,\ldots,j_q} \in \mathbb{R}$, of course they depend on the particular choice of the basis. Physicists omit the bases and use to write simply

$$t = \left(a^{i_1,\ldots,i_p}{}_{j_1,\ldots,j_q}\right),$$

the presence of p contravariant and q covariant indices of this sort of multi-dimensional matrix is enough to specify what type of tensor t is.

There is an obvious product between tensors... the tensor product:

$$\begin{array}{ccc} T_{q_1}^{p_1}(V) \times T_{q_2}^{p_2}(V) & \longrightarrow & T_{q_1+q_2}^{p_1+p_2}(V) \\ (t_1, t_2) & \longmapsto & t_1 \otimes t_2, \end{array}$$

It is possible to verify that, with this operation, $T(V)$ becomes an algebra.

4.4 Operations on tensors

Here we define the operations on tensors that can be found in differential geometry for different purposes. Let us start by justifying why we call $T(V)$ the tensor algebra.

4.4.1 Contraction

Contraction of type $\binom{r}{s}$: it is a linear function $C^r_s : T^p_q(V) \to T^{p-1}_{q-1}(V)$ that reduces the covariance and contravariance degree of a tensor. Moreover, it **generalizes the concept of trace to tensors**. For simplicity of notation, we can define the contraction on the basis elements of $T^p_q(V)$ (the definition is extended by linearity on the whole $T^p_q(V)$):

$$\boxed{C^r_s(v_{i_1} \otimes \cdots \otimes v_{i_p} \otimes v^{j_1} \otimes \cdots \otimes v^{j_q}) := v^{j_s}(v_{i_r}) v_{i_1} \otimes \cdots \otimes \cancel{v_{i_r}} \otimes \cdots \otimes v_{i_p} \otimes v^{j_1} \otimes \cdots \otimes \cancel{v^{j_s}} \otimes \cdots \otimes v^{j_q}}$$

explanation:

- we consider the i_r-th element of the V basis and the j_s-th element of the dual basis of V;
- we compute the real number $v^{j_s}(v_{i_r})$ in a linear way, because v^{j_s} is a linear form;
- we multiply this number to the tensor product basis of $T^p_q(V)$ taking out v_{i_r} and v^{j_s}... because they already served another purpose.

In coordinates, the contraction can be written as follows: if $t \in T^p_q(V)$, $t = \left(a^{i_1,\ldots,i_p}{}_{j_1,\ldots,j_q}\right)$,

$$C^r_s(t) = \left(a^{i_1\ldots i_{r-1} k i_{r+1}\ldots i_p}{}_{j_1\ldots j_{s-1} k j_{s+1}\ldots j_q}\right),$$

the same index k replaces the index i_r and j_s, so that a sum over k is intended!

We are now going to prove that **the operator** $C^1_1 : T^1_1(V) = V \otimes V^* \to T^0_0(V) \equiv \mathbb{R}$ **is simply the trace**. If $(v_1, \ldots, v_n)$ is a basis of V and $(v^1, \ldots, v^n)$ is the dual basis, then $t \in V \otimes V^*$, $t = a_{ij} v_i \otimes v^j$. By using the identification $V \otimes V^* \cong \mathrm{Hom}(V, V) \equiv \mathrm{End}(V)$, we can identify t with the linear function associated to the matrix $A = (a^i_j)$ w.r.t. the basis $(v_1, \ldots, v_n)$. By definition, the action of C^1_1 is as follows:

$$C^1_1(t) = C^1_1(a^i_j v_i \otimes v^j) \underset{\text{linearity}}{=} a^i_j C^1_1(v_i \otimes v^j) := a^i_j v^j(v_i) \cancel{v_i} \otimes \cancel{v^j} = a^i_j \delta^j_i = a^i_i = \mathrm{Tr}(A).$$

4.4.2 Symmetrization and antisymmetrization

We know that some particular multilinear forms are associated with important geometric concepts. For example, *symmetric bilinear forms* define real-valued scalar product, which can be used to define the *angle* between vectors and the concept of *orthogonality*; *alternating forms* define determinants, which are involved in the measure of *areas* and *volumes*.

Since tensors are multilinear forms, it makes sense to analyze the extension of these properties to tensors, this will be essential to build important objects such as the p-forms.

We will develop our analysis on $T^p(V)$, the one on $T_q(V)$ can be reproduced analogously. $t \in T^p(V)$ is such that:

$$\begin{aligned} t: \quad \underbrace{V^* \times \cdots \times V^*}_{p \text{ times}} &\longrightarrow \mathbb{R} \\ (\alpha^1, \ldots, \alpha^p) &\longmapsto t(\alpha^1, \ldots, \alpha^p) = a^{i_1 \ldots i_p} \alpha^1(v_{i_1}) \cdots \alpha^p(v_{i_p}). \end{aligned}$$

We want to single out those multilinear forms t which are *symmetric*, i.e.

$$t(\alpha^{\sigma(1)}, \ldots, \alpha^{\sigma(p)}) = t(\alpha^1, \ldots, \alpha^p)$$

for every permutation of the set of indices $\{1, \ldots, p\}$, and those which are *alternating*, i.e.

$$t(\alpha^{\sigma(1)}, \ldots, \alpha^{\sigma(p)}) = \operatorname{sign}(\sigma) t(\alpha^1, \ldots, \alpha^p),$$

where $\operatorname{sign}(\sigma) = (-1)^{N(\sigma)} \in \{-1, 1\}$, where $N(\sigma)$ is the number of *inversions* performed by σ, where an inversion is a switch of ordinal position between two indices after the application of σ. This means that:

$$\operatorname{sign}(\sigma) = \begin{cases} +1 & \text{if } \sigma \text{ performs an even number of inversions} \\ -1 & \text{if } \sigma \text{ performs an odd number of inversions.} \end{cases}$$

Some examples for $T^2(V)$: if $v, w \in V$, then:

- $t_0 = v \otimes w$ is, in general, not symmetric, nor alternating;
- $t_1 = v \otimes w + w \otimes v$ is **symmetric**, in fact the change $v \leftrightarrow w$ leaves t_1 unaffected;
- $t_2 = v \otimes w - w \otimes v$ is **alternating** (or **antisymmetric**), in fact the change $v \leftrightarrow w$ transforms t_2 to $-t_2$.

Notation:

$$\boxed{S^p(V)} \quad : \quad \text{subspace of } T^p(V) \text{ of symmetric tensors on } V$$

$$\boxed{\Lambda^p(V) = \mathbb{A}^p(V)} \quad : \quad \text{subspace of } T^p(V) \text{ of alternating tensors on } V$$

It can be proven that, if $\dim(V) = n$, then

$$\dim(T^p(V)) = n^p, \quad \dim(S^p(V)) = \binom{n+p-1}{p}, \quad \dim(\Lambda^p(V)) = \begin{cases} \binom{n+p-1}{p} & 0 \leqslant p \leqslant n \\ 0 & p > n \end{cases}.$$

The case of bilinear forms, i.e. $p = 2$ is special, let us see why:

$$\dim(T^2(V)) = n^2, \quad \dim(S^2(V)) = \frac{(n+1)n}{2}, \quad \dim(\Lambda^2(V)) = \frac{n(n-1)}{2},$$

so that $\dim(T^2(V)) = \dim(S^2(V)) + \dim(\Lambda^2(V))$, this is a consequence of the fact that every tensor $t \in T^2(V)$ can be written as the sum of a symmetric and an alternating tensor in a unique way as follows:

$$v \otimes w = \frac{v \otimes w + w \otimes v}{2} + \frac{v \otimes w - w \otimes v}{2} \iff t_0 = t_1 + t_2,$$

thus:

$$\boxed{T^2(V) = S^2(V) \oplus \Lambda^2(V)}\,.$$

For $n > 2$ *this is no longer true* because of a dimensional argument: $n^p \neq \binom{n+p-1}{p} + \binom{n}{p}$.

The operations that transform a generic tensor to a symmetric and and alternating one are called:

· **Symmetrization**: defined on the basis of $T^p(V)$ as follows

$$\begin{array}{rccl} S: & T^p(V) & \longrightarrow & S^p(V) \\ & v_1 \otimes \cdots \otimes v_p & \mapsto & S(v_1 \otimes \cdots \otimes v_p) = \frac{1}{p!}\sum\limits_{\sigma} v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(p)}, \end{array}$$

and extended by linearity to the whole space;

· **Antisymmetrization**: defined on the basis of $T^p(V)$ as follows

$$\begin{array}{rccl} A: & T^p(V) & \longrightarrow & \Lambda^p(V) \\ & v_1 \otimes \cdots \otimes v_p & \mapsto & A(v_1 \otimes \cdots \otimes v_p) = \frac{1}{p!}\sum\limits_{\sigma} \mathrm{sign}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(p)}, \end{array}$$

and extended by linearity to the whole space.

The normalization factor $1/p!$ comes from the fact that $p!$ is the number of distinct permutations of a set of p elements and it is introduced so that S and A reduce to the identity operator if they act, respectively, on symmetric and alternating tensors. For example: if $t = t_1 \in S^2(V)$, then

$$S(t_1) = S(v\otimes w + w\otimes v) \underset{\text{linearity}}{=} S(v\otimes w) + S(w\otimes v) := \frac{1}{2}(v\otimes w + w\otimes v) + \frac{1}{2}(w\otimes v + v\otimes w) = t_1,$$

where, for each term, we have applied the only two permutations on a set of two elements: the identity and the switch $v \leftrightarrow w$.

4.4.3 Symmetric product and external product

The symmetric product of tensors brings a couple of symmetric tensors to another symmetric tensor, and the external product of tensors brings a couple of alternating tensors to another alternating tensor. Let us start with the **symmetric product**:

$$\begin{array}{rccl} \odot: & S^p(V) \times S^q(V) & \longrightarrow & S^{p+q}(V) \\ & (t_1, t_2) & \mapsto & t_1 \odot t_2 := \frac{(p+q)!}{p!q!} S(t_1 \otimes t_2), \end{array}$$

by construction, it holds $t_1 \odot t_2 = t_2 \odot t_1$, i.e. $\odot$ is a commutative operation.

If we want $\odot$ to be an internal operation, we have to make p and q 'disappear', which can be done by taking the direct sum:

$$S(V) = \bigoplus_{p \geqslant 0} S^p(V),$$

$(S(V), \odot)$ is called **symmetrical algebra of** V.

Example: let $v, w \in S^1(V) = V$, then

$$v \otimes w \in T^2(V),\ S(v \otimes w) = \frac{1}{2!}(v \otimes w + w \otimes v)$$

$$v \odot w = \frac{2!}{1!1!}\frac{1}{2!}(v \otimes w + w \otimes v) = v \otimes w + w \otimes v,$$

which shows the usefulness of the normalization coefficients.

Analogously, if $v_1, \ldots, v_r \in V$, then

$$v_1 \odot \cdots \odot v_r = \sum_{\sigma} v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(r)}.$$

Let us now define the **external product**:

$$\begin{array}{cccc} \wedge : & \Lambda^p \times \Lambda^q(V) & \longrightarrow & \Lambda^{p+q}(V) \\ & (t_1, t_2) & \mapsto & t_1 \wedge t_2 := \frac{(p+q)!}{p!q!} A(t_1 \otimes t_2), \end{array}$$

and

$$\Lambda(V) = \bigoplus_{p=0}^{n} \Lambda^p(V),$$

we stop at $n = \dim(V)$ because, for $p > n$, $\Lambda^p(V) = \{0\}$.

$(\Lambda(V), \wedge)$ is the **external algebra of** V.

Example: let $v, w \in \Lambda^1(V) = V$, then

$$v \otimes w \in T^2(V),\ A(v \otimes w) = \frac{1}{2!}(v \otimes w - w \otimes v)$$

$$v \wedge w = \frac{2!}{1!1!}\frac{1}{2!}(v \otimes w - w \otimes v) = v \otimes w - w \otimes v.$$

Analogously, if $v_1, \ldots, v_r \in V$, then

$$v_1 \wedge \cdots \wedge v_r = \sum_{\sigma} \operatorname{sign}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(r)}.$$

4.5 Tensor bundles and tensor fields

All the previous constructions and operations on tensors have been defined for an arbitrary real vector space V of finite dimension. Thus, they can be applied in the case of vector bundles, where the fiber over each point of the base manifold is, by definition, a vector space.

If $\pi : E \to M$ is a vector bundle, then we have already seen that the dual bundle is constructed by taking the union of the dual spaces E_p^* of the fibers E_p, as p varies in M. This permits to build in a natural way the tensor bundle.

Def. 4.5.1 (Tensor bundle) *The tensor bundle $T_q^p(E)$ is the vector bundle whose fibers over $p \in M$ are given by*

$$T_q^p(E_p) = T^p(E_p) \otimes T_q(E_p) = E_p \underbrace{\otimes \cdots \otimes}_{p \text{ times}} E_p \otimes E_p^* \underbrace{\otimes \cdots \otimes}_{q \text{ times}} E_p^*.$$

Def. 4.5.2 (Tensor field) *A (local or global) p-contravariant and q-covariant tensor field is a (local or global) section of $T_q^p(E)$.*

Analogously, we can define the algebras $T(E), S(E), \Lambda(E)$.

The most important example is given by the tangent and cotangent bundle $E = TM$, $E = T^*M$ of a manifold M and particularly important is the external algebra of the cotangent bundle to a manifold M:

$$\Lambda(T^*M) = \bigoplus_{p=0}^{n} \Lambda^p(T^*M),$$

called **external algebra of** M (omitting TM).

Def. 4.5.3 (k-form) *A k-form on a manifold M is a section of $\Lambda^k(T^*M)$, i.e. a smooth assignment of an alternating tensor on T^*M. The set of all k-forms on M is a vector space w.r.t. the point-wise linear operations that is denoted either $\mathbb{A}^k(M)$ or $\Omega^k(M)$.*

As always, let us look at these objects in the local coordinates of a point $p \in M$ induced by a chart $(U, \varphi = (x^1, \ldots, x^n))$: we know that $(\partial_1|_p, \ldots, \partial_n|_p)$ is a basis for T_pM and this holds for every $p \in U$, thus it is possible to define the sections of TM given by

$$\begin{array}{rccl} \partial_i : & U & \longrightarrow & TM \\ & p & \longmapsto & \partial_i(p) := \partial_i|_p, \ \text{ with } \pi(\partial_i|_p) = p, \end{array}$$

Def. 4.5.4 (Local frame for TM) *The set $(\partial_1, \ldots, \partial_n)$ is called a local frame of TM on U.*

Similarly, by considering the dual basis $(dx^1|_p, \ldots, dx^n|_p)$ of T^*M, we can define the sections of T^*M given by

$$\begin{array}{rccl} dx^i : & U & \longrightarrow & T^*M \\ & p & \longmapsto & dx^i(p) = dx^i|_p, \ \text{ with } \pi(dx^i|_p) = p, \end{array}$$

every dx^i is a 1-form.

Def. 4.5.5 (Local frame for T^*M) *The set $(dx^1, \dots, dx^n)$ is called a local frame of T^*M on U.*

It is possible to verify that

$$(dx^{i_1} \wedge \cdots \wedge dx^{i_k}), \quad 1 \leqslant i_1 < \cdots < i_k \leqslant n,$$

is a **local frame of $\Omega^k(T^*M)$**.

Notice that the condition $1 \leqslant i_1 < \cdots < i_k \leqslant n$ is imposed to guarantee that *the indices $i_1, \dots, i_k$ are different*, otherwise the external product would be zero because of its anti-symmetry, which can be easily show by taking just two external factors $dx^h \wedge dx^h = -dx^h \wedge dx^h$ which implies $dx^h = 0$. Of course the name of the indices can always be permuted to fulfill the ordering written above.

Every k-form ω can be written, locally, as follows:

$$\boxed{\omega = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} a_{i_1 \cdots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}} \,,$$

where $a_{i_1 \cdots i_k} : U \to \mathbb{R}$ are scalar functions on the local chart domain U.

Def. 4.5.6 (Closed and exact forms, potentials) *A k-form ω is closed if $d\omega = 0$, it is exact if it exists a $(k-1)$-form η, called potential form, such that $\omega = d\eta$.*

Thus, an exact form is in the image of d, and a closed form belongs to the kernel of d.

For example, a **2-forms** can be written as $\omega = \omega_{ij} dx^i \wedge dx^j$ and the matrix ω_{ij} containing its coefficients its *anti-symmetric*.

Chapter 5

All about vector fields: flux, Lie derivative and bracket, distributions and foliations

The scope of this chapter is to discuss some fundamental objects of differential geometry that are related with vector fields. We will formalize the relationship between vector fields and differential equations via the flux theorem, which will allow us to introduce the Lie bracket and derivative. Then we will introduce the concept of distribution (totally unrelated to the distributions of the analytical domain…) and foliation.

5.1 Vector fields and derivations of the commutative algebra $\mathscr{C}^\infty(M)$

In (3.4), we have defined $\mathfrak{X}(M)$, the space of tangent vector fields on a smooth manifold M as the set of sections on the tangent bundle of M, i.e. $\mathfrak{X}(M) = \{\sigma : M \to TM,\ \pi \circ \sigma = id_M\}$.

We are now going to see an algebraic characterization of this space that is useful in many situations, e.g. for the definition of the Lie bracket in section 5.3.

Recall that we have defined T_pM, the space of tangent vectors on p to M, as $\mathrm{Der}_p(M)$, the space of derivations on M in p, i.e. $\mathbb{R}$-linear Leibniz-like functionals defined on the vector space $\mathscr{C}^\infty(M)$ of smooth scalar functions on M. Such functionals have the property to nullify constant functions.

If we want to extend the connection between derivations in a point p and tangent vectors to p to vector fields, we must get rid of the dependence of the derivation on the point p and give a more general definition.

Def. 5.1.1 (Derivation of an algebra) *Given a commutative algebra $\mathcal{A}$ on a field $\mathbb{K}$, we call derivation on $\mathcal{A}$ any linear function[1] $D : \mathcal{A} \to \mathcal{A}$ that satisfies the Leibniz rule, i.e.*

$$D(ab) = D(a)b + aD(b) \qquad \forall a, b \in \mathcal{A},$$

the juxtaposition of symbols means that we are multiplying by using the product of $\mathcal{A}$.

[1] Notice that, in this definition, $D \in \mathrm{End}(\mathcal{A})$, so D is not a functional but an endomorphism of $\mathcal{A}$.

The set of all derivations on $\mathcal{A}$ is written as $\mathrm{Der}(\mathcal{A})$ and it is a vector space w.r.t. linear operations defined point-wise.

In differential geometry, we have at disposal a commutative algebra: $\mathscr{C}^\infty(M)$, thus the vector space $\mathrm{Der}(\mathscr{C}^\infty(M))$ is perfectly defined and:

$$D(\phi_1\phi_2) = D(\phi_1)\phi_2 + \phi_1 D(\phi_2),$$

$\forall \phi_1, \phi_2 \in \mathscr{C}^\infty(M)$.

Remark 5.1.1 It is important to stress the difference between $\mathrm{Der}(\mathscr{C}^\infty(M))$ and $\mathrm{Der}_p(M)$:

- the derivations belonging to $\mathrm{Der}(\mathscr{C}^\infty(M))$ are endomorphisms D of $\mathscr{C}^\infty(M)$ which act globally on smooth scalar functions on M: $D : \mathscr{C}^\infty(M) \to \mathscr{C}^\infty(M)$

- those belonging to $\mathrm{Der}_p(M)$ are the tangent vectors to M at p, so they are linear functionals v_p acting locally, in an open neighborhood of p: $v_p : \mathscr{C}^\infty(U) \to \mathbb{R}$.

In spite of being different objects, there is a clear correspondence between them: we can define tangent vectors independently of a specified point by considering a section of the tangent bundle TM, i.e. a vector field on M, as formalized in the following result.

Theorem 5.1.1 *The vector space of vector fields on M and of derivations on $\mathscr{C}^\infty(M)$ are canonically isomorphic:*

$$\boxed{\mathfrak{X}(M) \cong \mathrm{Der}(\mathscr{C}^\infty(M))}\ .$$

Some part of the proof are quite technical, what it is really important to retain is how to build the isomorphism: consider the vector field

$$\begin{array}{rccl} X: & M & \longrightarrow & TM \\ & p & \longmapsto & X(p) = (p, X_p), \end{array}$$

such that $X_p \in T_pM \equiv \mathrm{Der}_p(M)$, i.e. X_p is a derivation at p, a Leibniz-like linear functional on $\mathscr{C}^\infty(M)$. Then, define

$$\begin{array}{rccl} D_X: & \mathscr{C}^\infty(M) & \longrightarrow & \mathscr{C}^\infty(M) \\ & \phi & \longmapsto & D_X(\phi), \end{array}$$

where the smooth scalar valued function $D_X(\phi)$ is defines as follows:

$$\begin{array}{rccl} D_X(\phi): & M & \longrightarrow & \mathbb{R} \\ & p & \longmapsto & D_X(\phi)(p) := X_p(\phi), \end{array}$$

thanks to the Leibniz-like behavior of X_p, D_X is a derivation of the algebra $\mathscr{C}^\infty(M)$.

Vice-versa, starting from the derivation $D : \mathscr{C}^\infty(M) \to \mathscr{C}^\infty(M)$, we can univocally define the vector field $X : M \to TM$, $p \mapsto (p, X_p)$, where the local expression of $X_p \in T_pM$ in the local coordinate system $(x^1, \dots, x^n)$ of $p \in M$ is:

$$X_p = D(x^j)(p)\, \partial_j\big|_p\,,$$

perfectly well-defined because $x^j : M \to \mathbb{R}$ are smooth functions and so $D(x^j) \in \mathscr{C}^\infty(M)$, hence $D(x^j)(p) \in \mathbb{R}$. The Leibniz-like behavior is then clear.

5.2 Integral curves and flux of a vector field

In this section we point out the relationship between vector fields and differential equations. In order to accomplish this task, we first need to recall a classical result of the theory of differential equations in $\mathbb{R}^n$.

Theorem 5.2.1 ($\exists$! of the solution of a system of ordinary differential equations in $\mathbb{R}^n$) *Let $U \subseteq R^n$ an open set and let $(x^1, \ldots, x^n) : U \to \mathbb{R}$ be smooth functions. Then:*

- $\boxed{\exists}$ *: for all $t_0 \in \mathbb{R}$ and $x_0 \in U$ there exists $\delta > 0$ and an open subset $U_0 \subset U$, with $x_0 \in U_0$, such that, for all $x \in U_0$, there exists a curve $\gamma_x : (t_0 - \delta, t_0 + \delta) \to U$ which solves the following Cauchy problem:*

$$\begin{cases} \frac{d\gamma_x^j(t)}{dt} = x^j(\gamma_x(t)), \qquad j = 1, \ldots, n \\ \gamma_x(t_0) = x_0. \end{cases} \tag{5.1}$$

- $\boxed{\textit{Smooth dependence on initial data}}$ *: the function*

$$\begin{array}{rccc} \Theta : & (t_0 - \delta, t_0 + \delta) \times U_0 & \longrightarrow & U \\ & (t, x) & \longmapsto & \Theta(t, x) = \gamma_x(t), \end{array}$$

  is smooth, i.e. $\gamma_x(t)$ is smooth w.r.t. $t \in (t_0 - \delta, t_0 + \delta)$ and w.r.t. $x \in U_0$.

- $\boxed{!}$ *: two solutions of the Cauchy problem always coincide in the intersection of their domains.*

Since this result holds locally, we can imagine that it is possible to generalize it to manifolds. This is indeed the case and to prove it we must introduce a suitable terminology.

Def. 5.2.1 (Integral curve of a vector field) *Given a smooth manifold M, let us consider:*

- $X \in \mathfrak{X}(M)$
- $p \in M$
- *$I \subset \mathbb{R}$ open and such that $0 \in I$*
- *$\gamma : I \to M$ smooth.*

Then γ is the integral curve of the vector field X passing through p if:

$$\begin{cases} \gamma'(t) = X(\gamma(t)), \quad \forall t \in I \\ \gamma(0) = p. \end{cases}$$

Geometrically, the fact that γ is the integral curve of X means that *the tangent vector $\gamma'(t)$ to M at each element of its support $\{\gamma(t),\ t \in I\} \subset M$ coincides with the tangent vector assigned by the vector field X in the point $\gamma(t)$.*

We can transform locally the search for integral curves of a vector field in the situation considered in theorem 5.2.1, as formalized in the following result.

Theorem 5.2.2 ($\exists$! of the integral curves of a vector field on a manifold) *Let $X \in \mathfrak{X}(M)$, $p \in M$ and $(U, \varphi = (x^j)_{j=1,\dots,n})$ a local chart in p. Then the assertions of theorem 5.2.1 holds if we replace the Cauchy problem (5.1) with the following:*

$$\begin{cases} \frac{d\tilde{\gamma}^j(t)}{dt} = X^j(\tilde{\gamma}(t)), \qquad j = 1, \dots, n \\ \tilde{\gamma}(t_0) = \varphi(p) \in \mathbb{R}^n, \end{cases} \tag{5.2}$$

where $\tilde{\gamma} : \varphi \circ \gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ and $X = X^j \partial_j$ is the decomposition of X induced by the local coordinates $(x^1, \dots, x^n)$.

Proof. The proof consists in composing γ with φ to get a curve $\tilde{\gamma} : \varphi \circ \gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ with ε small enough so that $\gamma(-\varepsilon, \varepsilon) \subset U$. If we write its components as $(\tilde{\gamma}^1, \dots, \tilde{\gamma}^n)$, then

$$\tilde{\gamma}'(t) = (\tilde{\gamma}^j)'(t)\, \partial_j|_{\tilde{\gamma}(t)}\,,$$

where $\partial_j|_{\tilde{\gamma}(t)} \in T_{\tilde{\gamma}(t)}\mathbb{R}^n \cong \mathbb{R}^n$.

It is now clear that γ is an integral curve of X if and only if $\tilde{\gamma}$ is a solution of the Cauchy problem in $\mathbb{R}^n$ written in (5.2) and so the theorem 5.2.1 can be applied. $\square$

Let us now use this result to prove a very powerful theorem, fundamental for differential geometry of smooth manifolds.

Theorem 5.2.3 (The flux theorem) *Let M be a manifold. For every $X \in \mathfrak{X}(M)$ there exists a unique open neighborhood $\mathcal{U}$ of $\{0\} \times M$ in $\mathbb{R} \times M$ and a unique smooth function $\Theta : \mathcal{U} \to M$ such that the following assertions hold.*

1. *For all* <u>*fixed*</u> $p \in M$*, the set*
$$\mathcal{U}^p := \{t \in \mathbb{R} \ : \ (t,p) \in \mathcal{U}\} \subseteq \mathbb{R}$$
is an open interval containing 0*.*

2. *For all* <u>*fixed*</u> $p \in M$*, the curve*
$$\begin{array}{rccl} \vartheta^p : & \mathcal{U}^p \subseteq \mathbb{R} & \longrightarrow & M \\ & t & \longmapsto & \vartheta^p(t) = \Theta(t,p) \end{array}$$
is the only maximal integral curve of X passing through p (i.e. it cannot be extended to a larger domain remaining an integral curve of X).

3. *For all* <u>*fixed*</u> $t \in \mathbb{R}$*, the set*
$$\mathcal{U}_t := \{p \in M \ : \ (t,p) \in \mathcal{U}\} \subseteq M$$
is an open subset of M.

4. *For all* <u>*fixed*</u> $t \in \mathbb{R}$*, the map*
$$\begin{array}{rccl} \vartheta_t : & \mathcal{U}_t \subseteq M & \longrightarrow & M \\ & p & \longmapsto & \vartheta_t(t) = \Theta(t,p) \end{array}$$
is a diffeomorphism such that $\boxed{\vartheta_t^{-1} = \vartheta_{-t}}$ and $\boxed{\vartheta_0 = id_{\mathcal{U}_t}}$. Moreover, if $p \in \mathcal{U}_t$, then, fixed $s \in \mathbb{R}$, $p \in \mathcal{U}_{t+s}$ if and only if $\Theta(t,p) \in \mathcal{U}_s$ and in this case it holds that
$$\boxed{\vartheta_s(\vartheta_t(p)) = \vartheta_{s+t}(p)} \, .$$

5. *For all $\phi \in \mathscr{C}^\infty(M)$ and all $p \in M$:*
$$\boxed{\frac{d}{dt}(\phi \circ \vartheta^p)\Big|_{t=0} = X(\phi)(p)} \, .$$

6. *For all $(t,p) \in \mathcal{U}$:*
$$\boxed{d(\vartheta_t)_p(X_p) = X_{\vartheta_t(p)}} \, .$$

Before going through the details of the proof, let us remark that, **when we write $\vartheta_s(\vartheta_t(p))$ we are considering two different integral curves of X**: from p, we first follow the integral curve of X passing through p for a time t and we stop when we arrive at the point $\vartheta_t(p) \in M$. From here, we continue by following the integral curve of X passing through $\vartheta_t(p)$ for a time s and we stop when we arrive at the point $\vartheta_s(\vartheta_t(p)) \in M$.

This is the reason why the equality $\vartheta_s(\vartheta_t(p)) = \vartheta_{s+t}(p)$ is so strong: it says that, if we perform the procedure just described, we arrive exactly to the same point as if we followed just the integral curve of X passing from p for a time $s+t$.

Proof.

1. Theorem 5.2.2 implies that, for all $p \in M$, there exists always an integral curve of X passing through it and that two integral curves of X passing through p coincide in the intersection of their domains. This allows us to simply define $\mathcal{U}^p$ as **the union of all the open intervals $I \subseteq \mathbb{R}$ containing 0 on which an integral curve $\gamma : I \to M$ of X passing through p is defined**. Being the union of open sets, $\mathcal{U}^p$ is open.

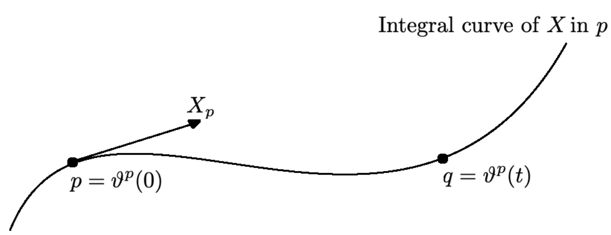
2. The previous argument implies that it exists $\vartheta^p : \mathcal{U}^p \to M$, integral curve of X passing through p and defined on the whole $\mathcal{U}^p$. This is the maximal integral curve and it is unique as a consequence of the unicity of the solution of the Cauchy problem (5.2). Moreover, this, together with the fact that we know how to construct $\mathcal{U}^p$, allow us to build $\mathcal{U}$ and Θ:

$$\mathcal{U} := \{(t,p) \in \mathbb{R} \times M \;:\; t \in \mathcal{U}^p\}, \quad \Theta : \mathcal{U} \to M,\ \Theta(t,p) := \vartheta^p(t).$$

Notice that this definition of $\mathcal{U}$ does not imply immediately that it is open, we will prove it later.

3. To prove that $\mathcal{U}$ is open and that Θ is smooth a quite technical use of theorem 5.2.2 must be performed, together with the result in 4. We skip these details and pass directly to the more interesting proof of 4.

4. The proof of 3. will be simpler if we first deal with the point 4. By definition $\mathcal{U}_0 = M$ and $\vartheta_0 = id_M$. Let $p \in M$ and $t \in \mathcal{U}^p$, we write $q = \vartheta^p(t)$, as represented in the figure below.



It is useful to perform a re-parameterization of ϑ^p as follows: let us define

$$\mathcal{U}^p - t := \{s \in \mathbb{R} \;:\; s + t \in \mathcal{U}^p\}$$

then, the curve

$$\begin{array}{rccc} \sigma: & \mathcal{U}^p - t & \longrightarrow & M \\ & s & \longmapsto & \sigma(s) = \vartheta^p(s+t) \end{array}$$

is still an integral curve of X because

$$\sigma'(s) = \frac{d\vartheta^p}{ds}(s+t) = X(\vartheta^p(s+t)) =: X(\sigma(s))$$

and σ passes through q, in fact $\sigma(0) = \vartheta^p(t) = q$.

We observe now that, by unicity of the integral curve, $\sigma(s) = \vartheta^q(s)$, so it must be:

$$\vartheta^{\vartheta^p(t)}(s) = \vartheta^q(s) = \sigma(s) =: \vartheta^p(s+t),$$

i.e. $\Theta(s, \Theta(t,p)) = \Theta(s+t, p)$ or $\vartheta_{s+t}(p) = \vartheta_s(\vartheta_t(p))$. Moreover, $\mathcal{U}^p - t \subseteq \mathcal{U}^q$.

Since $0 \in \mathcal{U}^p$, $0 - t = -t \in \mathcal{U}^q$, but $\mathcal{U}^q$ is the domain of ϑ^q, so the fact that $-t \in \mathcal{U}^q$ means that $\vartheta^q(-t) = p$. This formalizes the fact that, if we are placed in q, then we can turn back to p by 'reversing the time' of the quantity t.

If we interchange the couple $(-t, q)$ with (t, p) we get that $\mathcal{U}^q + t \subseteq \mathcal{U}^q$, thus it holds that $\mathcal{U}^q - t = \mathcal{U}^q = \mathcal{U}^{\Theta(t,p)}$. But then $q = \Theta(t,p) \in \mathcal{U}_s$ if and only if $p \in \mathcal{U}_{s+t}$, which concludes the proof of 4.

5. Since $\vartheta^p(0) = p$ and $(\vartheta^p)'(0) = X_p$, thanks to the definition of differential we have:

$$X(\phi)(p) = d\phi_p(X_p) = \left.\frac{d}{dt}(\phi \circ \vartheta^p(t))\right|_{t=0}.$$

6. Let $(t_0, p_0) \in \mathcal{U}$ and $\phi \in \mathscr{C}^\infty(M)$, then:

$$\begin{aligned}
d(\vartheta_{t_0})_{p_0}(X_{p_0})(\phi) &= X_{p_0}(\phi \circ \vartheta_{t_0}) \qquad \text{(def. of differential)} \\
&= \left.\frac{d}{dt}(\phi \circ \vartheta_{t_0} \circ \vartheta^{p_0}(t))\right|_{t=0} \qquad \text{(by using 5.)} \\
&= \left.\frac{d}{dt}\phi(\vartheta_{t_0+t})\right|_{t=0} = \left.\frac{d}{dt}\phi(\vartheta^{p_0}(t_0+t))\right|_{t=0} \\
&= X_{\vartheta^{p_0}(t_0)}(\phi) = X_{\vartheta_{t_0}(p_0)}(\phi),
\end{aligned}$$

since the result hold for all $\phi \in \mathscr{C}^\infty(M)$, we have $d(\vartheta_t)_p(X_p) = X_{\vartheta_t(p)}$. □

The function Θ contains the information about all the integral curves of X passing through all the points of M. For this reason it deserves a special name and it characterizes certain special vector fields.

Def. 5.2.2 (Flux) *For every $X \in \mathfrak{X}(M)$, the function $\Theta : \mathcal{U} \subseteq \mathbb{R} \times M \to M$ is called the local flux of the vector field X.*

Def. 5.2.3 (Complete vector field) *$X \in \mathfrak{X}(M)$ is called complete if $\mathcal{U} = \mathbb{R} \times M \to M$, i.e. if all the integral curves of X are defined for all $t \in \mathbb{R}$.*

Def. 5.2.4 (X-invariant vector field) *Let $X, Y \in \mathfrak{X}(M)$, and Θ the local flux of X. Y is said to be X-invariant if, for all (t,p) belonging to the domain of the local flux of X, we have:*

$$d(\vartheta_t)_p(Y_p) = Y_{\vartheta_t(p)}, \qquad \vartheta_t(p) = \Theta(t,p).$$

Let us interpret this last definition: for all $p \in M$ we can evaluate Y in p, obtaining Y_p, a tangent vector to M. We then move along the integral curve of X for a time t, until arriving to the point $q = \vartheta_t(p)$. We can compare the tangent vector $Y_{\vartheta_t(p)}$ with the one that we obtain by applying the differential map to $\vartheta_t(p)$ calculated in Y_p, i.e. $d(\vartheta_t)_p(Y_p)$, which is a tangent vector to M at q, so it belongs to the same tangent space as $Y_{\vartheta_t(p)}$ and so the comparison is meaningful. If it happens that these two tangent vectors are the same, then Y is X-invariant[2].

Thanks to the property 6. of the flux theorem, X **is** X**-invariant**.

[2]more synthetically: computing the tangent vector Y at $\vartheta_t(p)$ is the same as sending the tangent vector Y_p to $Y_{\vartheta_t(p)}$ via differential along the integral curve of X passing through p.

5.3 The Lie bracket

As previously said, the concept of Lie bracket shows the usefulness of interpreting vector fields as derivations on the ring of smooth scalar functions. In fact, if $X, Y \in \mathfrak{X}(M)$ are interpreted as derivations, i.e. $X, Y : \mathscr{C}^\infty(M) \to \mathscr{C}^\infty(M)$ are linear Leibniz-like operators, then they can be composed to get two new linear operators on $\mathfrak{X}(M)$, namely $X \circ Y$ and $Y \circ X$, this is a privilege that we do not have if we interpret X, Y as sections of TM. Linearity is obviously preserved by composition, however the Leibniz-like behavior is not, in fact, by using first the Leibniz behavior of Y and then of X we get:

$$(X \circ Y)(fg) = X(Y(fg)) = X(fY(g)+Y(f)g) = fX(Y(g))+X(f)Y(g)+X(g)Y(f)+X(Y(f))g,$$

this is different than $fX(Y(g)) + X(Y(F))g$, which is what we would expect from an hypothetical Leibniz-like behavior of $X \circ Y$. In fact, if we consider the geometrical meaning of the two intermediate terms of $X \circ Y$, we see that they act as a second-order differential operators, thus making, globally, $X \circ Y$ a second-order differential operator, instead of a first-order one, as it should be, a vector field is associated to the first order Cauchy problem (5.2).

Nonetheless, the intermediate terms of $X \circ Y$ are symmetrical w.r.t. the exchange of X with Y, thus, if we compute $Y \circ X$ and we subtract it from $X \circ Y$, we erase these spurious terms and we remain with a derivation. These considerations justify the following definition.

Def. 5.3.1 (Lie bracket) *Given $X, Y \in \mathfrak{X}(M)$, their Lie bracket is the vector field $[X, Y] \in \mathfrak{X}(M)$ defined by:*

$$\boxed{[X, Y] := X \circ Y - Y \circ X}\ .$$

*X, Y are said to **commute** if $[X, Y] = 0$, the null vector field.*

The properties of the Lie bracket of vector fields are listed below.

Theorem 5.3.1 *Let $X, Y, Z \in \mathfrak{X}(M)$, $f, g \in \mathscr{C}^\infty(M)$ and $a, b \in \mathbb{R}$, then the following properties hold.*

1. $[Y, X] = -[X, Y]$.

2. $[aX + bY, Z] = a[X, Z] + b[Y, Z]$ *and* $[Z, aX + bY] = a[Z, X] + b[Z, Y]$.

3. $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$.

4. $[fX, gY] = fg[X, Y] + fX(g)Y - gY(f)X$, *in particular, if $f \equiv 1$, $[X, gY] = g[X, Y] + X(g)Y$, i.e.*
$$\begin{array}{rccc} [X, \cdot] : & \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(M) \\ & Y & \longmapsto & [X, Y], \end{array}$$
is a derivation on $\mathfrak{X}(M)$, as commutative algebra over $\mathscr{C}^\infty(M)$.

5. *If $X = X^h \partial_h$ and $Y = Y^k \partial_k$ are the representations of X and Y in a local coordinate system, then the local coordinate expression for the Lie bracket $[X, Y]$ is the following:*
$$\boxed{[X, Y] = (X^h \partial_h Y^k - Y^h \partial_h X^k)\partial_k}.$$
In particular, $[\partial_h, \partial_k] = 0$, as a consequence of Schwarz's theorem.

Proof.

1. and 2. Direct computation.

3. We have:

$$[X,[Y,Z]] = [X, YZ - ZY] = XYZ - XZY - YZX + ZYX,$$
$$[Y,[Z,X]] = [Y, ZX - XZ] = YZX - YXZ - ZXY + XZY,$$
$$[Z,[X,Y]] = [Z, XY - YX] = ZXY - ZYX - XYZ + YXZ,$$

summing the left hand sides and the rightmost hand sides we get 0.

4. We have:

$$\begin{aligned}[fX, gY] &= fX(gY) - gY(fX) = fX(g)Y + fgXY - gY(f)X - gfYX \\ &= fg(XY - YX) + fX(g)Y - gY(f)X = fg[X,Y] + fX(g)Y - gY(f)X.\end{aligned}$$

5. We have:

$$\begin{aligned}[X,Y] &= [X^h\partial_h, Y^k\partial_k] \\ &\quad \text{(by linearity)} \\ &= X^h\partial_h(Y^k\partial_k) - Y^k\partial_k(X^h\partial_h) \\ &\quad \text{(applying the Leibniz rule for the action of the partial derivatives } \partial_h \text{ and } \partial_k\text{)} \\ &= X^h(\partial_h Y^k)\partial_k + X^hY^k\partial_h\partial_k - Y^k(\partial_k X^h)\partial_h - Y^kX^h\partial_k\partial_h \\ &\quad \text{(by Schwarz's theorem for second order partial derivatives)} \\ &= X^h(\partial_h Y^k)\partial_k + \cancel{X^hY^k\partial^2_{hk}} - \cancel{Y^kX^h\partial^2_{kh}} - Y^k(\partial_k X^h)\partial_h \\ &= X^h(\partial_h Y^k)\partial_k - Y^k(\partial_k X^h)\partial_h \\ &\quad \text{(by exchanging } h \text{ with } k \text{ in the second term)} \\ &= X^h(\partial_h Y^k)\partial_k - Y^h(\partial_h X^k)\partial_k \\ &= (X^h\partial_h Y^k - Y^h\partial_h X^k)\partial_k.\end{aligned}$$

□

Thanks to the properties just proven, the vector space of all vector fields on M endowed with the Lie bracket, i.e. $(\mathfrak{X}(M), [\ ,\ ])$ **is a Lie algebra**, as it is clear from the definition that follows.

Def. 5.3.2 (Lie algebra) *A vector space $\mathfrak{a}$ over a field $\mathbb{K}$ is a Lie algebra*[3] *if there exists a binary operation $[\ ,\ ] : \mathfrak{a} \to \mathfrak{a}$, called Lie bracket, that satisfies the following properties for all $a, b \in \mathbb{K}$ and all $x, y, z \in \mathfrak{a}$:*

1. *Anti-symmetry:* $[y,x] = -[x,y]$
2. *Bilinearity:* $[ax + by, z] = a[x,z] + b[y,z]$ *and* $[z, ax + by] = a[z,x] + b[z,y]$
3. *Jacobi identify:* $[x,[y,z]] + [z,[x,y]] + [y,[z,x]] = 0.$

By anti-symmetry it follows that $[x,x] = 0$ for all $x \in \mathfrak{a}$.

[3] A lower case fraktur letter is usually used to denote a Lie algebra.

5.4 The Lie derivative

The Lie derivative allows us to define the concept of **derivative w.r.t. a vector field** on a manifold. As we will see in chapter 8, this is *not exactly* the perfect generalization of the concept of directional derivative in $\mathbb{R}^n$ to abstract manifolds.

As always, let us first analyze the trivial case of $M = \mathbb{R}^n$. In this situation the tangent spaces to each point of M are canonically identified with $\mathbb{R}^n$, so a vector field $X \in \mathfrak{X}(\mathbb{R}^n)$ is simply a section of $T\mathbb{R}^n \cong \mathbb{R}^{2n}$, i.e. a smooth map $X : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$, where U is an open neighborhood of $p \in \mathbb{R}^n$ and, as usual, we have made use of the identification $T_p\mathbb{R}^n \cong \mathbb{R}^n$. Thus the derivative of another vector field $Y \in \mathfrak{X}(\mathbb{R}^n)$ along an integral curve of X in p can be simply reduced to the directional derivative of Y in p along the direction given by the tangent vector $X_p \in \mathbb{R}^n$.

However, it is immediate to understand that these considerations do not work anymore in a non-trivial manifold M. Consider, in fact, the situation depicted in the figure 5.1.

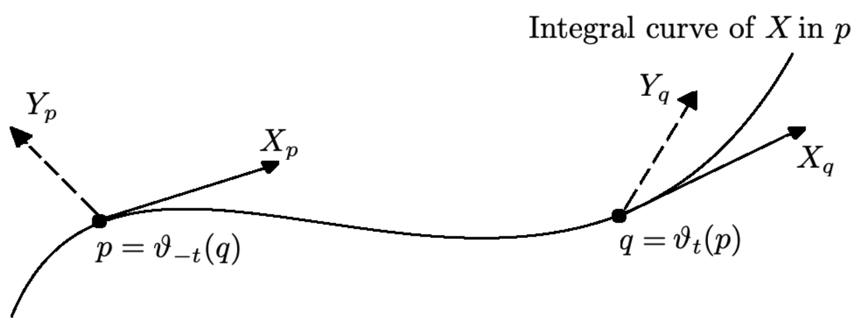


Figure 5.1: Comparing tangent vectors at different points of an integral curve.

If we want to estimate the rate of change of the vector field Y when we pass from $p = \vartheta_0(p)$ to $q = \vartheta_t(p)$, where $\vartheta_t(p) = \Theta(t, p)$, Θ being the local flux of X, then we should compute the quantity:

$$\lim_{t\to 0} \frac{Y_{q=\vartheta_t(p)} - Y_{p=\vartheta_0(p)}}{t},$$

but $Y_q \in T_qM$ and $Y_p \in T_pM$, thus the comparison $Y_q - Y_p$ is ill-posed because the two vectors live in different vector spaces!

The solution to this problem is to take back Y_q to the vector space T_pM along the integral curve of X. Notice that

$$\begin{array}{rccl} \vartheta_{-t} : & M & \longrightarrow & M \\ & q & \longmapsto & \vartheta_{-t}(q) = p, \end{array}$$

so, we clearly have to apply the differential to ϑ_{-t} to move the tangent vectors to the integral curve of X at q to bring them back to tangent vectors to the integral curve of X at p, i.e.

$$\begin{array}{rccl} d(\vartheta_{-t})_q : & T_qM & \longrightarrow & T_{\vartheta_{-t}(q)=p}M \\ & Y_q & \longmapsto & d(\vartheta_{-t})_q(Y_q), \end{array}$$

since $Y \circ \Theta$ is smooth, the function $t \mapsto d(\vartheta_{-t})_{\vartheta_t(p)}(Y_{\vartheta_t(p)})$ is a smooth curve in T_pM that depends smoothly on p. Notice that, in general, $d(\vartheta_{-t})_q(Y_q)$ will be different than Y_p, as depicted in figure 5.2, thus the difference between these two tangent vectors will be different than the null vector of T_pM. We can now formalize the concept of Lie derivative as follows.

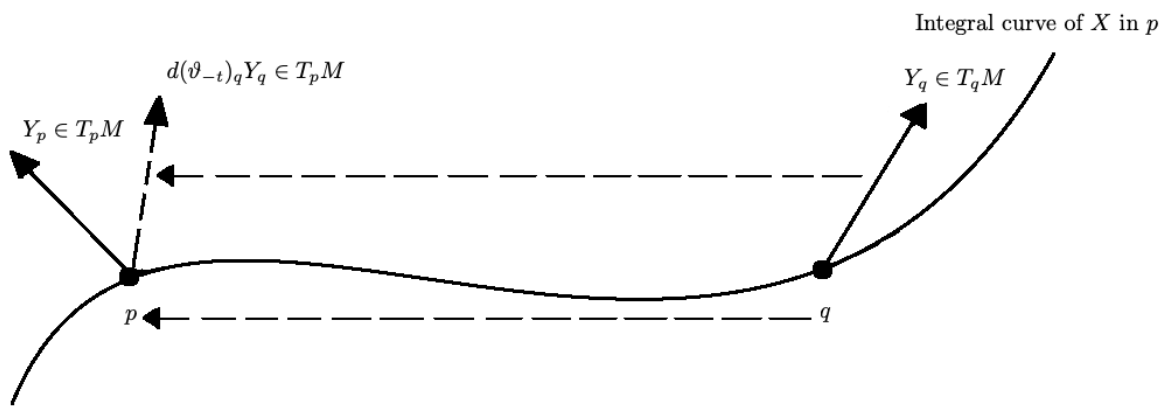


Figure 5.2: Construction of the Lie derivative of a vector field.

Def. 5.4.1 (Lie derivative of a vector field) *The Lie derivative of the vector field $Y \in \mathfrak{X}(M)$ along the vector field $X \in \mathfrak{X}(M)$ is the linear operator:*

$$\begin{array}{rccc} \pounds_X : & \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(M) \\ & Y & \longmapsto & \pounds_X Y, \end{array}$$

where

$$\pounds_X Y(p) := \lim_{t \to 0} \frac{d(\vartheta_{-t})_{\vartheta_t(p)}(Y_{\vartheta_t(p)}) - Y_p}{t} \equiv \frac{d}{dt}\left(d(\vartheta_{-t})_{\vartheta_t(p)}(Y_{\vartheta_t(p)})\right)\Big|_{t=0}. \tag{5.3}$$

which is called the Lie derivative of Y along X in the point $p \in M$.

It is clear that, **if Y is a X-invariant vector field, then** $\pounds_X Y = 0$, the null vector field.

Formula (5.3) is clearly not easy to handle, which is why mathematicians searched for a simpler expression, the result is surprising: thanks to the properties of the flux of vector fields, it can be proven that the Lie derivative is simply the Lie bracket!

Theorem 5.4.1 *For all $X, Y \in \mathfrak{X}(M)$, it holds that*

$$\boxed{\pounds_X Y = [X, Y]}\,.$$

The link between the Lie derivative and bracket shows that this latter hides a geometrical meaning that we investigate in the following subsection.

It is possible to generalize the concept of Lie derivative also to arbitrary tensor fields.

Def. 5.4.2 (Lie derivatives of scalar functions) *Given $X \in \mathrm{Der}(\mathscr{C}^\infty) \cong \mathfrak{X}(M)$, the Lie derivative of a rank-0 tensors, i.e. a scalar fields $\phi \in \mathscr{C}^\infty(M) \equiv T_0^0(M)$, along X is the linear operator:*

$$\begin{array}{rccc} \pounds_X : & \mathscr{C}^\infty(M) & \longrightarrow & \mathscr{C}^\infty(M) \\ & \phi & \longmapsto & \boxed{\pounds_X \phi = X(\phi)}. \end{array}$$

Let us now pass to 1-forms $\omega \in \Omega^1(M) \equiv T^0_1(M)$: given a vector field $Y \in \mathfrak{X}(M) \equiv T^1_0(M)$, we can build a scalar field simply by applying ω to Y, in fact

$$\begin{array}{rccl} \omega(Y): & M & \longrightarrow & \mathbb{R} \\ & p & \longmapsto & \omega(Y)(p) := \omega_p(Y_p) \end{array}$$

is perfectly defined because $\omega_p \in T^*_pM$ and $Y_p \in T_pM$. In local coordinates, if $\omega = \omega_i dx^i$ and $Y = Y^j\partial_j$, with $\omega_i : M \to \mathbb{R}$ and $Y^j : M \to \mathbb{R}$ smooth coefficient functions, then

$$\omega(Y) = \omega_i Y^i,$$

in fact $\omega(Y) = \omega_i dx^i(Y^j\partial_j) = \omega_i Y^j dx^i(\partial_j) = \omega_i Y^j \delta^i_j = \omega_i Y^i$.

Let us impose that $\pounds_X(Y)$ verifies the Leibniz rule:

$$\pounds_X(\omega(Y)) = (\pounds_X\omega)Y + \omega(\pounds_X Y),$$

so $(\pounds_X\omega)Y = \pounds_X(\omega(Y)) - \omega(\pounds_X Y)$, but we already know how the Lie derivative is defined for scalar and vector fields, i.e. $\pounds_X(\omega(Y)) = X(\omega(Y))$ and $\pounds_X Y = [X, Y]$, respectively, thus we get:

$$(\pounds_X\omega)Y = X(\omega(Y)) - \omega([X, Y]).$$

This simple computation explains the definition of the Lie derivative of a 1-form as follows.

Def. 5.4.3 (Lie derivatives of a 1-form) *The Lie derivative of a 1-form $\omega \in \Omega^1(M) \equiv T^1_0(M)$ along X is the linear operator:*

$$\begin{array}{rccl} \pounds_X: & \Omega^1(M) & \longrightarrow & \Omega^1(M) \\ & \omega & \longmapsto & \pounds_X\omega, \end{array}$$

$$\begin{array}{rccl} \pounds_X\omega: & TM & \longrightarrow & \mathbb{R} \\ & Y & \longmapsto & \boxed{\pounds_X\omega(Y) := X(\omega(Y)) - \omega([X, Y])}\,. \end{array}$$

The general case is given by a tensor field $t \in T^p_q(M)$: if $Y_1, \ldots, Y_q \in T^1_0(M) = TM$ and $\omega_1, \ldots, \omega_p \in T^0_1(M) = \Omega^1(M)$, then $t(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_p) \in \mathscr{C}^\infty(M)$ and so:

$$\pounds_X(t(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_p)) = X(t(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_p)),$$

thus, to define the Lie derivative of t, we must impose, as before, the Leibniz behavior and solve for $\pounds_X t$:

$$\begin{aligned} \pounds_X(t(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_p)) = {} & (\pounds_X t)(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_p) + \\ & + t([X, Y_1], \ldots, Y_q, \omega_1, \ldots, \omega_p) + \ldots \\ & + t(Y_1, \ldots, Y_{q-1}, [X, Y_q], \omega_1, \ldots, \omega_p) \\ & + t(Y_1, \ldots, Y_q, \pounds_X\omega_1, \omega_2, \ldots, \omega_p) + \ldots \\ & + t(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_{p-1}, \pounds_X\omega_p), \end{aligned}$$

i.e. the Lie derivative of the tensor field $t \in T^p_q(M)$ along $X \in \mathfrak{X}(M)$ is:

$$\begin{aligned} (\pounds_X t)(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_p) := {} & X(t(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_p)) \\ & - t([X, Y_1], \ldots, Y_q, \omega_1, \ldots, \omega_p) + \ldots \\ & - t(Y_1, \ldots, Y_{q-1}, [X, Y_q], \omega_1, \ldots, \omega_p) \\ & - t(Y_1, \ldots, Y_q, \pounds_X\omega_1, \omega_2, \ldots, \omega_p) + \ldots \\ & - t(Y_1, \ldots, Y_q, \omega_1, \ldots, \omega_{p-1}, \pounds_X\omega_p). \end{aligned}$$

5.4.1 Geometrical features of the Lie bracket

Given two vector fields $X, Y \in \mathfrak{X}(M)$, figure 5.3 depicts the following path:

- we start from $p \in M$ and we follow the integral curve of X passing through p for an amount of 'time' measured by the value h of the parameter t, arriving in q;
- we restart from q, but now we follow the integral curve of Y passing through q for the same amount of time h, arriving in r;
- we restart from r, following the integral curve of X passing through r for an amount of time $-h$, arriving in s;
- finally, from s, we follow the integral curve of Y passing through s for an amount of time $-h$, arriving in the point T that we indicate as $\gamma(h)$.

The curve $h \mapsto \gamma(h)$ is smooth and such that $\gamma(0) = p$.

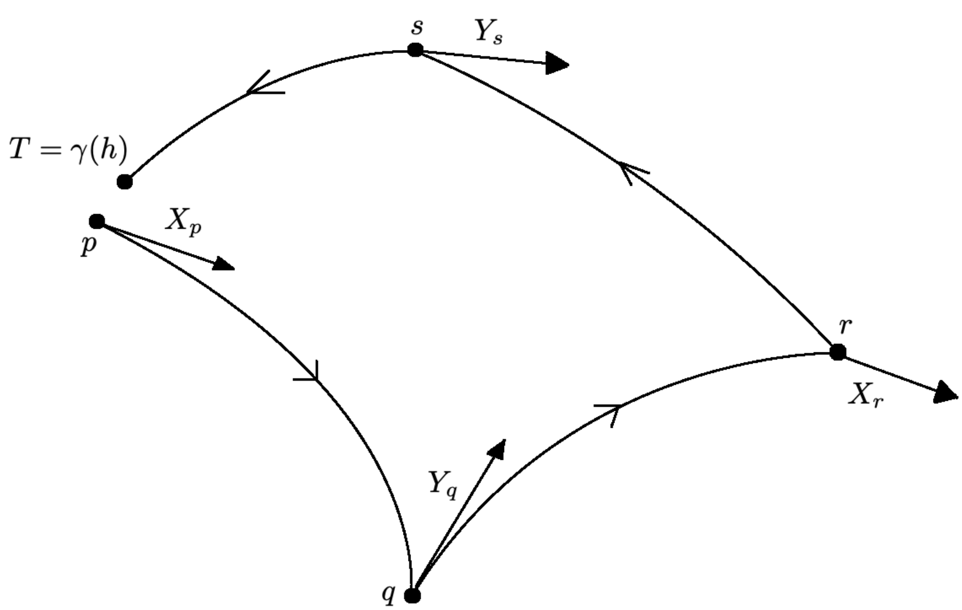


Figure 5.3: Geometrical interpretation of the Lie bracket.

In [22] we can find the proof of the following result.

Theorem 5.4.2 *With the notations above, it holds that:*

1. $\gamma'(0) = 0$, *i.e. at first order, the quadrilateral depicted in figure 5.3 is closed;*
2. $\gamma''(0) = 2[X, Y]$, *i.e. at second order, the obstruction to the closure of the quadrilateral depicted in figure 5.3 is measured by the Lie bracket between X and Y.*

5.4.2 Pushforward of a vector field by a diffeomorphism

The concept of Lie bracket (and so of Lie derivative) can be associated to an operation that relates vector fields between manifolds: the pushforward.

To introduce this operation, consider a smooth map $f : M \to N$ and a vector field X on M, then, for each $p \in M$, $df_p : T_pM \to T_{f(p)}N$, but $X_p \in T_pM$, thus $df_p(X_p)$ is a tangent vector to N at $q = f(p)$.

However, depending on the properties of f, this technique may fail to defines a vector field Y on N. In fact, if f is not surjective, there is no rule to assign a tangent vector to N at the points $q \in N \backslash f(M)$. On the other hand, if f is not injective, then there are at least two distinct points $p_1, p_2 \in M$ such that $f(p_1) = f(p_2) = q \in N$, in this case $Y^1_{p_1} := df_{p_1}(X_{p_1})$ and $Y^2_{p_2} := df_{p_2}(X_{p_2})$ would be two possibly different tangent vectors to N at the same point q, thus creating an ambiguity in the assignment for the hypothetical vector field on N that we would like to create via f and X.

These considerations motivate why we can push a vector field forward to another manifold by means of a map if and only if the map is a diffeomorphism.

Def. 5.4.4 (Pushforward of a vector field) *Let M, N be to manifolds, $X \in \mathfrak{X}(M)$ and $f : M \to N$ a diffeomorphism. We call pushforward induced by f the linear map $f_* \equiv df$ defined as follows:*

$$\begin{array}{rccl} f_* \equiv df : & \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(N) \\ & X & \longmapsto & f_*(X) \equiv df, \end{array}$$

where, for all $q \in N$, $f_*(X)(q) \equiv f_*(X)_q$, or $df(X)(q) \equiv df(X)_q$, is defined as follows:

$$f_*(X)_q = df_{f^{-1}(q)}(X_{f^{-1}(q)}) \qquad or \qquad df(X)_q = df_{f^{-1}(q)}(X_{f^{-1}(q)}).$$

Again, we underline that the need of a diffeomorphism is clear from the definition: thanks to this, we can bring back any point $q \in N$ to a point $p = f^{-1}(q) \in M$ and use the vector field $X \in \mathfrak{X}(M)$.

Let us discuss a simple example: we consider the manifold $M := \{(x, y) \in \mathbb{R}^2 : x > 0,\ y > 0\}$ and the map

$$\begin{array}{rccl} f : & M & \longrightarrow & M \\ & (x, y) & \longmapsto & f(x, y) = (u, v) := (xy, \frac{y}{x}). \end{array}$$

We want to verify that f is a diffeomorphism and then to compute explicitly the pushforward $f_*(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y})$.

f is clearly smooth on M and,

$$\begin{cases} u = xy \\ v = \frac{y}{x} \end{cases} \implies \begin{cases} uv = y^2 \\ \frac{u}{v} = x^2 \end{cases} \implies \begin{cases} x = \sqrt{\frac{u}{v}} \\ y = \sqrt{uv} \end{cases},$$

so f is invertible with

$$\begin{array}{rccl} f^{-1} : & M & \longrightarrow & M \\ & (u, v) & \longmapsto & f^{-1}(u, v) = \left(\sqrt{\frac{u}{v}}, \sqrt{uv}\right), \end{array}$$

again smooth on M. So, f is a diffeomorphism.

Let us now compute the pushforward: the most important fact to keep in mind when computing the pushforward via a diffeomorphism is that this one provides a coordinate transformation, in this case f provides the formulae to express x and y in terms of u and v and vice-versa. The vector field that we have to pushforward is written in terms of the first coordinate system and if we take a look back to the definition of pushforward, and we try to understand the essence of it, without considering for a moment the technical details, we understand that **the pushforward of a vector field is exactly the same vector field, but written in the novel coordinate system prompted by the diffeomorphism!** Loosely speaking, computing the pushforward of a vector field is nothing but an exercise of coordinate change.

So, we rewrite x and y as the precise functions of u and v that we have found above, and $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ as linear combinations of $\frac{\partial}{\partial u}$ and $\frac{\partial}{\partial v}$ (the new basis of the tangent vector w.r.t. the new variables!). Thanks to the theorem of differentiation of composed functions of multiple variables we have:

$$\frac{\partial}{\partial x} = \left(\frac{\partial u}{\partial x}\right)\frac{\partial}{\partial u} + \left(\frac{\partial v}{\partial x}\right)\frac{\partial}{\partial v} = y\frac{\partial}{\partial u} + \frac{y}{x^2}\frac{\partial}{\partial v} = \sqrt{uv}\frac{\partial}{\partial u} - v\sqrt{\frac{v}{u}}\frac{\partial}{\partial v},$$

analogously we obtain

$$\frac{\partial}{\partial y} = \sqrt{\frac{u}{v}}\frac{\partial}{\partial u} + \sqrt{\frac{v}{u}}\frac{\partial}{\partial v}.$$

It follows that

$$f_*\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right) = \frac{u}{v}\left(\sqrt{uv}\frac{\partial}{\partial u} - v\sqrt{\frac{v}{u}}\frac{\partial}{\partial v}\right) + \sqrt{uv}\left(\sqrt{\frac{u}{v}}\frac{\partial}{\partial u} + \sqrt{\frac{u}{v}}\frac{\partial}{\partial v}\right) = 2u\frac{\partial}{\partial u}.$$

Again, it is important to remark that the vector fields

$$x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} \quad \text{and} \quad 2u\frac{\partial}{\partial u}$$

are exactly the same intrinsic object, but written w.r.t. two different coordinate systems.

If $f : M \to N$ is only a smooth map and not a diffeomorphism, then it is impossible to define the pushforward of vector fields, however, it is still possible to correlate them, in the sense formalized below.

Def. 5.4.5 (f-related vector fields) *Let $f : M \to N$ be a smooth function between manifolds, $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$. X and Y are f-related if, for all $p \in M$*

$$Y_{f(p)} = df_p(X_p),$$

i.e. if the tangent vectors determined by Y in the points of N belonging to the range of f coincide with the tangent vectors determined by the differential map of f applied to the tangent vectors determined by X in the points of M.

If f is a diffeomorphism, it is easy to see that $Y = f_*(X)$ is the only vector field on N f-related to X, see [10] Proposition 8.19 page 183.

The properties of f-related vector fields are listed in the following result, for the proof see [10] chapter 8.

Theorem 5.4.3 *Let $f : M \to N$ be a smooth function between manifolds, $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$.*

1. *Y is f-related to X if and only if, for every $\phi \in \mathscr{C}^\infty(N)$, it holds that*

$$X(\phi \circ f) = Y(\phi) \circ f.$$

2. *If Y_1 is f-related to X_1 and Y_2 is f-correlated to X_2, then $[Y_1, Y_2]$ is f-correlated to $[X_1, X_2]$. In other words, the Lie bracket is compatible with the f-correlation.*

3. *If f is a diffeomorphism, then*

$$[f_*(X_1), f_*(X_2)] = f_*([X_1, X_2]),$$

i.e. the pushforward $f_* : \mathfrak{X}(M) \to \mathfrak{X}(N)$ is compatible with the Lie bracket.

These properties are used to solve the following problem: suppose to have a local frame $X_1, \ldots, X_n$ for TM, $\dim(M) = n$, is there a set of conditions to guarantee that it exists a local chart (U, φ) of M such that $X_j = \partial_j$, $j = 1, \ldots, n$ on U?

A necessary condition can be found very easily: since $[\partial_i, \partial_j] = 0$ $\forall i, j = 1, \ldots, n$, it is necessary that $[X_i, X_j] = 0$ $\forall i, j = 1, \ldots, n$. In theorem 5.4.6 we will see that this condition is also sufficient, but to formulate it properly we have to define a new concept and to introduce intermediate results.

Def. 5.4.6 (Regular and singular points of a vector field) *Let $X \in \mathfrak{X}(M)$, a point $p \in M$ is said to be a regular point of the vector field X if $X_p \neq 0$, i.e. if the tangent vector to M assigned by X in p is not null, otherwise, if $X_p = 0$, p is called a singular point for X.*

Theorem 5.4.4 *Let $X \in \mathfrak{X}(M)$ and $p \in M$ a regular point for X. Then, it exists a local chart (U, φ) centered in p, i.e. $\varphi(p) = 0 \in \mathbb{R}^n$, such that:*

$$X|_U = \partial_1,$$

i.e. in an open neighborhood of p, the tangent vectors assigned by X are all parallel.

The following theorem says that if $[X, Y] = 0$, then the quadrilateral depicted in figure 5.3 is closed, not only at the second order, but at every order, i.e. the obstruction to its closure is totally contained in the Lie bracket.

Theorem 5.4.5 *Let $X, Y \in \mathfrak{X}(M)$ with flux $\Theta : \mathcal{U} :\to M$ and $\Psi : \mathcal{V} :\to M$, respectively. Then, the following assertions are equivalent:*

1. *$[X, Y] = 0$*

2. *Y is X-invariant*

3. *X is Y-invariant*

4. *$\psi_s \circ \vartheta_t = \vartheta_t \circ \psi_s$ as long as one of the two is defined, i.e. the fluxes of X and Y commute.*

We can now see that the condition of commuting is necessary and sufficient for linearly independent vector fields X_j in $\mathfrak{X}(M)$ to be written locally as ∂_j.

Theorem 5.4.6 *Let $X_1, \dots, X_k \in \mathfrak{X}(M)$ linearly independent vector fields in every point of M, thus $k \leqslant n = dim(M)$ (if $k = 1$ then $X_p \neq 0$ $\forall p \in M$). Then, the following properties are equivalent:*

1. *for all $p \in M$ it exists a local chart (U, φ) centered in p such that: $X_j|_U = \partial_j$, $\forall j = 1, \dots, k$.*

2. *$[X_i, Y_j] = 0$ $\forall i, j = 1, \dots, k$.*

5.5 Foliation of a manifold: distributions and the Frobenius theorem

Let $X \in \mathfrak{X}(M)$ be a *nowhere vanishing* vector field. Thanks to the existence and uniqueness of integral curves of X, it is possible to decompose M in the disjoint union of integral curves of M: in fact the integral curve of X passes through p and $T_pM \neq T_qM$ if $p \neq q$.

If we forget the curve parameterization and pay attention just to its trajectory in M we have, for all $p \in M$, the smooth assignment of a 1D subspace of T_pM, which is the straight line generated by X_p, a tangent vector to M at p, different from 0 by hypothesis.

The Frobenius theorem generalizes this fact to k-dimensional subspaces, instead of 1-dimensional ones, of T_pM by using the concept of distribution.

Def. 5.5.1 (k-dimensional distribution) *A k-dimensional distribution on a manifold M is a subset $D \subset TM$ of the tangent bundle to M such that $D_p := D \cap T_pM$ is a vector subspace of dimension k of T_pM, for all $p \in M$.*

D is smooth if, for every $p \in M$, it exist an open subset $U \subset M$ of p and k local vector fields $Y_1, \dots, Y_k \in \mathfrak{X}(U)$ such that

$$D_p = span(Y_1(p), \dots, Y_k(p)), \quad \forall p \in U.$$

$Y_1, \dots, Y_k$ is said to be a local frame or a local basis for D on U.

D_p is precisely what generalizes the integral curves discussed in the previous sections.

Def. 5.5.2 (Local section of a distribution) *A local section of a smooth distribution D on an open set $U \subset M$ is a vector field $X \in \mathfrak{X}(U)$ such that $X_p \in D_p$ for all $p \in U$.*

The space of all local sections of D on U will be denoted as $\mathfrak{X}_D(U)$.

Def. 5.5.3 (Involutive distribution) *If, for all open subset $U \subset M$ and $X, Y \in \mathfrak{X}_D(U)$ we have $[X, Y] \in \mathfrak{X}_D(U)$, then D is said involutive.*

So, involutive is used to denote the stability of local sections of a distribution w.r.t. the Lie bracket. This is surely true if X and Y commute, because in this case their bracket is the null vector field which takes values in the null vector, that is contained in any vector subspace! It may however happen that X and Y do not commute, but their Lie bracket belongs to $\mathfrak{X}_D(U)$.

Def. 5.5.4 (Integral submanifold of a distribution) *An integral submanifold S of a distribution is an immersed submanifold of M such that*

$$\forall p \in S, \; T_pS = D_p.$$

This definition clearly shows that the concept of integral submanifold of a distribution is the k-dimensional generalization of the concept of integral curve of one vector field. When the vector fields are k, the integral curve becomes an integral submanifold of dimension k.

Def. 5.5.5 (Integrable distribution) *The distribution D is said to be integrable if all $p \in M$ is contained in an integral submanifold of D.*

As we will define more precisely later on, we can already understand from the previous definition that distributions satisfying a the request of integrability give rise to partitions of the manifold into smaller submanifolds, which is called a 'foliation'.

The first result about distributions that we mention is the fact that:

$$\boxed{\text{Integrability of } D \implies \text{Involutivity of } D}.$$

Theorem 5.5.1 *Every smooth integrable distribution is also involutive.*

Let us now introduce a particular type of chart for a distribution.

Def. 5.5.6 (Flat chart of a distribution) *Let D be a k-dimensional smooth distribution of a n-dimensional manifold M. A local chart (U, φ) of M is said to be flat for D if*

1. *φ identifies its domain U with an open subset of $\mathbb{R}^n$ which can be decomposed as the Cartesian product of two open subsets, i.e. there exist open sets $V' \in R^k$ and $V'' \in \mathbb{R}^{n-k}$ such that $\varphi(U) = V' \times V''$;*

2. *$(\partial_1, \ldots, \partial_k)$ is a local frame for D on U.*

Def. 5.5.7 (Leaves of a distribution) *If (U, φ) is flat for D, then fixed the $n-k$ real values $c^{k+1}, \ldots, c^n \in R$, the subsets of M given by*

$$\{p \in U \,:\, x^{k+1} = c^{k+1}, \ldots, x^n = c^n\},$$

i.e. those for which the last $n-k$ local coordinates are fixed, are called leaves of U for the distribution D.

Figure 5.4 give a graphical representation of the concept of leaves.

Def. 5.5.8 (Completely integrable distribution) *If, for all $p \in M$, there exists a local chart (U, φ) of M centered in p which is flat, then the distribution D is said to be completely integrable.*

Lemma 5.5.1 *Every completely integrable smooth distribution D is also integrable.*

Proof. If D is completely integrable, then the leaves of a local flat chart (U, φ) for D are, by construction, integral submanifolds of M, hence D is integrable. □

So, up to now we have this chain of implications:

$$\boxed{\text{Complete integrability of } D \implies \text{Integrability of } D \implies \text{Involutivity of } D}.$$

The (local) Frobenius theorem states that the implications can be reversed and, in fact, there three properties are totally equivalent.

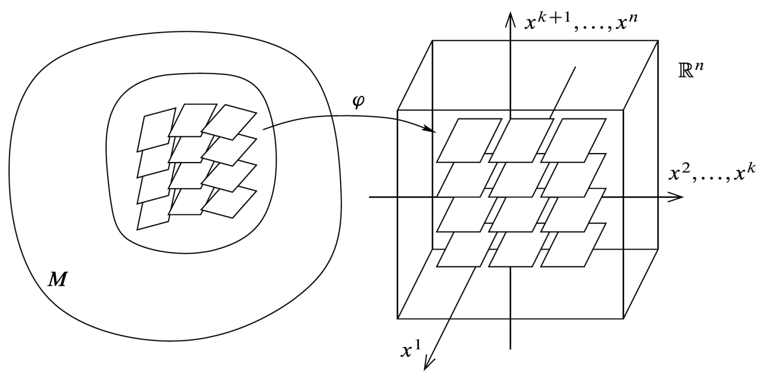


Figure 5.4: Given an open subset $U \subset M$, a k-distribution is associated to a family of k vector subspaces, depicted as plains inside U. Then (U, φ) is a flat chart on U for D if these subspaces are transformed by φ to vector subspaces (again depicted as plains) of $\mathbb{R}^n$ which are parallel to the first k coordinate axes. Source: [10], chapter 19.

Theorem 5.5.2 (Frobenius) *Every smooth involutive k-distribution D is completely integrable, and so also integrable.*

Hence:

$$\boxed{\text{Complete integrability of } D \iff \text{Integrability of } D \iff \text{Involutivity of } D},$$

this is actually a highly non-intuitive because involutivity seems a much weaker condition than being completely integrable.

Proof. The proof is very important because it is constructive. Thanks to theorem 5.4.6, to prove that the smooth involutive k-distribution D is completely integrable it is sufficient to prove that, for every $p \in M$ it can be found an open neighborhood of p such that there exists a local frame in U for D given by *commuting vector fields.*

Applied to our case, theorem 5.4.6 implies that the subspace generated by these commuting vector fields, which is the subspace generated by D, can be identified with the first k coordinate tangent vectors ∂_j, $j = 1, \dots, k$. Hence this subspace, when mapped via the local chart map φ in $\mathbb{R}^n$ will be parallel to the first k coordinate axes, which is equivalent to the existence of a local flat chart.

Keeping this in mind, let us consider a local chart (U, φ) centered in p such that it exists a local reference frame $(X_1, \dots X_k)$ for D on U given by non-necessarily commuting vector fields. The local frame can of course be completed to a basis of T_pM and, modulo a permutation of the indices, this basis can be written as $(X_1|_p, \dots, X_k|_p, \partial_{k+1}|_p, \dots, \partial_{k+1}|_p)$. To make the notation more uniform, let us redefine $X_j|_p := \partial_j|_p$, for $j = k+1, \dots, n$.

We have of course another local frame for TM, namely $\partial_1, \dots, \partial_n$, hence a change of basis matrix composed by functions $a_i^j \in \mathscr{C}^\infty(U)$ such that

$$X_i = a_i^j \partial_j, \quad i = 1, \dots, n.$$

Since $(X_1|_p, \dots, X_1|_p)$ is a basis of T_pM, the numerical matrix $A(p) := (a_i^j(p))_{i,j=1}^n$ is an invertible matrix, i.e. $\det(A(p)) \neq 0$. Since the function a_i^j are smooth, they are also continuous, and thus it surely exists a suitably small open neighborhood, that we still denote with U, such that $A(p)$ is invertible for all $p \in U$. Let then $B(p) = A(p)^{-1}$ for all $p \in U$, where the numerical values in $B(p)$ come from the evaluation of smooth functions $b_i^j \in \mathscr{C}^\infty(U)$ in $p \in U$.

Hence, we can also express ∂_j in terms of X_i with the functional coefficients b_j^i, that we can decompose as follows:

$$\partial_j = \sum_{i=1}^{n} b_j^i X_i = \sum_{i=1}^{k} b_j^i X_i + \sum_{i=k+1}^{n} b_j^i X_i, \quad j = 1, \dots, n, \tag{5.4}$$

recalling that $X_j := \partial_j$, for $j = k+1, \dots, n$.

Let us now concentrate only on the vector field appearing in the first sum: since $(X_1, \dots, X_k)$ is a local frame on U for the distribution D, it follows that also

$$Y_j := \sum_{i=1}^{k} b_j^i X_i \in \mathfrak{X}_D(U), \quad j = 1, \dots, k,$$

i.e. each Y_j is a local section of D on U.

The core of the proof of the Frobenius theorem is based on the fact that $Y_1, \dots, Y_k$ constitute a local vector field on U for D given commuting vector fields. Thanks to the argument recalled at the beginning of the proof, once we manage to prove the commutativity of the fields Y_i, the theorem will be demonstrated.

To this aim, let us consider the function $f : U \to \mathbb{R}^k$ defined as $f = \pi \circ \phi$, where $\pi : \mathbb{R}^n \to \mathbb{R}^k$ is just the natural projection on the first k coordinates. Explicitly:

$$M \ni p \underset{\varphi}{\longmapsto} \varphi(p) = (x^1(p), \dots, x^n(p)) \underset{f}{\longmapsto} (x^1(p), \dots, x^k(p)) \in \mathbb{R}^k.$$

We see that f acts as the identity on the first k components of each point of U, hence, if we take the differential of f in any other point $q \in U$ and we apply it to $\partial_j|_q$ but only for all $j = 1, \dots, k$, we map it again to ∂_j, but this time acting on the tangent space $T_{f(q)}M$, so:

$$df_q(\partial_j|_q) = \partial_j|_{f(q)}, \quad j = 1, \dots, k.$$

On the other side, by linearity and recalling from eq. (5.4) that $\partial_j = Y_j + \sum_{i=k+1}^{n} b_j^i \partial_i$, we also have:

$$df_q(\partial_j|_q) = df_q(Y_j|_q) + \sum_{i=k+1}^{n} b_j^i(q) \cancelto{0_{T_{f(q)}}}{df_q(\partial_i|_q)} = df_q(Y_j|_q),$$

where the differential gives rise to the null vector because f acts only on the first k components, so the last $n-k$ tangent vectors, ranging from $k+1$ to n, nullify the map f!

So, for all $q \in U$, we obtain the equality

$$df_q(Y_j|_q) = \partial_j|_{f(q)}, \quad j = 1, \dots, k,$$

but we know that the coordinate tangent vectors $\partial_1|_{f(q)}, \dots, \partial_k|_{f(q)}$ are linearly independent in every point of U, hence also the tangent vectors $Y_1|_q, \dots, Y_k|_q$ are linearly independent, because a linear function as df_q cannot send linearly-dependent vectors to linearly independent ones! This proves that the k set of linearly independent vector fields $(Y_1, \ldots, Y_k)$ form a local frame for the k-distribution D on U,

This also shows that, in spite of the fact that the differential $d_q : T_qM \to \mathbb{R}^k$ is not injective because $k < n$, its restriction $d_q|_{D_q}$ is indeed injective for all $q \in U$ because it sends a basis of its domain into a basis of its range, as just shown.

Moreover, importantly, for all $q \in U$

$$df_q([Y_i|_q, Y_j|_q]) = [\partial_i, \partial_j]|_{f(q)} = 0, \quad i, j = 1, \ldots, k$$

because the coordinate tangent vectors commute.

Thanks to item 2 of theorem 5.4.3, it follows that $df_q([Y_i|_q, Y_j|_q]) = 0$ only implies that $[Y_i|_q, Y_j|_q] \in \ker df_q$, not that $[Y_i, Y_j]|_q = 0$, because we have underlined above that df_q is not one-to-one, so its kernel is not reduced to the null vector.

To prove that this is indeed the case, i.e. that $[Y_i, Y_j]|_q = 0$ we must use the initial hypothesis, i.e. that D is involutive, that we haven't already used! This is actually the only step of the proof where the hypothesis helps. The involutivity of D implies that $[Y_i, Y_j]|_q \in D_q$ for all $q \in U$, but, having previously noticed that $df_q|_{D_q}$ is an injective map, its kernel is indeed reduced to the null vector, so $[Y_i, Y_j]|_q = 0$ for all $q \in U$. This shows that the vector fields $Y_1, \ldots, Y_k$ commute. □

To resume, the constructive points of the Frobenius proof are the following:

- we start with a local reference frame $(X_1, \ldots, X_k)$ for the distribution and we build the change of basis matrix w.r.t. the local coordinate fields $(\partial_1, \ldots, \partial_n$., which is a problem of linear algebra;

- we then invert this matrix, again a problem of linear algebra;

- finally we use the coefficients of the inverse matrix to build a local frame for the distribution with commuting fields via the linear combination of these last coefficients with the vector fields of the initial reference frame.

Def. 5.5.9 (Foliation of a manifold) *We say that a k-dimensional smooth involutive, or integrable, or completely integrable, distribution induces a foliation of an n-dimensional manifold M given by the union of its leaves.*

Chapter 6

Integration of differential forms on oriented manifolds

Inspirational epithap wanted...

...

The integration of k-forms on manifolds has many important applications both in pure mathematics and in physics and engineering. In this chapter, we are going to show that, in order to properly define the integral of a k-form, we must introduce a linear map, called *pull-back*, that is the analogue of the push-forward for vector fields and also to restrict the set of manifolds to those whose can be *oriented*.

Let us recall that, given a smooth manifold M of dimension n, a k-form is a section of the fiber bundle given by the k-th external power of the cotangent bundle to M, i.e. of $\Lambda^k T^*M$. The symbol $\mathbb{A}^k(M)$ denotes the real vector space of k-forms on M, where the letter $\mathbb{A}$ reminds that differential forms are alternating, i.e. anti-symmetric tensors.

Finally, remember that, in local coordinates w.r.t. the local chart $(U, \varphi = (x^1, \ldots, x^n))$ of M, every k-form $\omega \in \mathbb{A}^k(M)$ can be written as follows:

$$\omega|_U = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \omega_{i_1,\ldots,i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$

where the coefficients $\omega_{i_1,\ldots,i_k}$ are smooth scalar functions on U: $\omega_{i_1,\ldots,i_k} \in \mathscr{C}^\infty(U)$.

6.1 The pull-back of a k-form

We have also seen that a diffeomorphism between manifolds induces a map between vector fields, the push-forward. We will see that, in order to integrate a k-form, the analogue of the push-forward for k-forms, called pull-back, will be of fundamental importance. Contrarily to the case of vector fields, any smooth map between manifolds can be used to define the pull-back, without requiring it to be invertible.

Def. 6.1.1 (Pull-back of a k-form) *Let $f : M \to N$ be a smooth map between manifolds and let $\omega \in \mathbb{A}^k(N)$. The pull-back of ω, denoted with $f^*\omega$ allows us to 'bring back' ω to a k-form on M: $f^*\omega \in \mathbb{A}^k(M)$. For all $p \in M$ and $v_1, \ldots, v_k \in T_pM$, it is defined as follows:*

$$(f^*\omega)_p(v_1, \ldots, v_k) := \omega_{f(p)}(df_p(v_1), \ldots, df_p(v_k)), \tag{6.1}$$

The definition is well-posed because $df_p(v_j) \in T_{f(p)}(M)$ for all $j = 1, \ldots, k$, and there is absolutely no need for f to be invertible.

If M is a submanifold of N and f is the canonical inclusion $\iota : M \hookrightarrow N$, then it is custom to write $\iota^*\omega = \omega|_M$.

The **most important properties of the pull-back** are listed in the following theorem.

Theorem 6.1.1 *With the notations above the assertions below hold.*

1. $f^* : \mathbb{A}^k(N) \to \mathbb{A}^k(M)$ *is a linear map for all* $k \geqslant 0$*, hence it also defines a linear map* $f^* : \mathbb{A}^\bullet(N) = \bigoplus_{k \geqslant 0} \mathbb{A}^k(V)(N) \to \mathbb{A}^\bullet(M) = \bigoplus_{k \geqslant 0} \mathbb{A}^k(V)(M)$.

2. *The pull-back is compatible with the external product, i.e.* $f^*(\eta \wedge \omega) = (f^*\eta) \wedge (f^*\omega)$, *for all* $\eta, \omega \in A^\bullet(N)$.

3. *If the expression of* ω *in a local chart* $(V, \psi = (y^1, \ldots, y^n))$ *of* N *is*
$$\omega|_V = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \omega_{i_1, \ldots, i_k} dy^{i_1} \wedge \cdots \wedge dy^{i_k}$$
then the expression of its pull-back $f^*\omega$ *in the local chart* $(f^{-1}(V), (f^1 = y^1 \circ f, \ldots, y^n \circ f))$ *of* M *is*
$$f^*\omega|_{f^{-1}(V)} = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \omega_{i_1, \ldots, i_k} df^{i_1} \wedge \cdots \wedge df^{i_k}.$$

4. *If* $g_1, \ldots, g_k \in \mathscr{C}^\infty(N)$*, then*
$$f^*(dg_1 \wedge \cdots \wedge dg_k) = d(g_1 \circ f) \wedge \cdots \wedge d(g_k \circ f).$$

5. *If* $\dim M = \dim N = n$ *and* $(U, \varphi = (x^1, \ldots, x^n))$, $(V, \psi = (y^1, \ldots, y^n))$, $f(U) \subseteq V$ *and* $\tilde{f} = \psi \circ f \circ \varphi^{-1}$ *is, as usual, the local representation of* f *w.r.t. the previous local charts, then, for all* $g \in \mathscr{C}^\infty(V)$ *it holds that*
$$f^*(g\, dy^1 \wedge \cdots \wedge dy^n) = (g \circ f) \det(d\tilde{f})\, dx^1 \wedge \cdots \wedge dx^n, \tag{6.2}$$
where $d\tilde{f}$ *coincides with* $J\tilde{f}$*, the Jacobian matrix of the function* $\tilde{f}$.

6. *If* $h : N \to S$ *is another smooth function between manifolds, then*
$$(h \circ f)^* = f^* \circ h^*.$$
In particular, if f *is a diffeomorphism, then* f^* *is an isomorphism of algebras and it holds that*
$$(f^*)^{-1} = (f^{-1})^*.$$

Proof.

1\. & 2. : they follow directly from the definition of pull-back.

3. Let us suppose that $p \in f^{-1}(V)$ and $v_j \in T_pM$, then, by definition of pull-back

$$
\begin{aligned}
(f^*\omega)_p(v_1, \dots, v_k) &= \omega_{f(p)}(df_p(v_1), \dots, df_p(v_k)) \\
&\underset{\text{loc. expr.of } \omega_{f(p)}}{=} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \omega_{i_1,\dots,i_k}(f(p)) dy^{i_1} \wedge \cdots \wedge dy^{i_k}(df_p(v_1), \dots, df_p(v_k)) \\
&\underset{\text{def. of wedge product}}{=} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \omega_{i_1,\dots,i_k}(f(p)) df^{i_1} \wedge \cdots \wedge df^{i_k}(v_1, \dots, v_k),
\end{aligned}
$$

having also used the fact that $df^i = d(y^i \circ df) = dy^i \circ df$.

4. It follows from the fact that $f^*(dg) = d(g \circ f)$ for all $g \in \mathscr{C}^\infty(N)$:

$$f^*(dg_1 \wedge \dots dg_k) \underset{2.}{=} (f^* dg_1) \wedge \cdots \wedge (f^* dg_k) = d(g_1 \circ f) \wedge \cdots \wedge d(g_k \circ f).$$

5. By 3. we have

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

but

$$df^h = \frac{\partial f^h}{dx^k} dx^k,$$

so

$$f^*(g\, dy^1 \wedge \cdots \wedge dy^n) = (g \circ f)\, \frac{\partial f^1}{dx^{k_1}} dx^{k_1} \wedge \cdots \wedge \frac{\partial f^n}{dx^{k_n}} dx^{k_n} =$$

$$
\begin{aligned}
f^*(g\, dy^1 \wedge \cdots \wedge dy^n) &= (g \circ f)\, \frac{\partial f^1}{dx^{k_1}} dx^{k_1} \wedge \cdots \wedge \frac{\partial f^n}{dx^{k_n}} dx^{k_n} \\
&= (g \circ f) \left(\sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \frac{\partial f^1}{dx^{\sigma(1)}} \cdots \frac{\partial f^n}{dx^{\sigma(n)}} \right) dx^1 \wedge \cdots \wedge dx^n,
\end{aligned}
$$

where, to pass from the right-hand side of the first row to the formula written in the second row we have used these three facts:

- in the right-hand side of the first row there are n implicit *non-ordered* sums over the indices $k_1, \dots, k_n$;
- $\wedge$ is a *multilinear* function, so we can factor the partial derivative out of the wedge products;
- $\wedge$ is *alternating*, so we can order the wedge products as $dx^1 \wedge \cdots \wedge dx^n$ but paying the price of introducing the sign of the permutation that we have used in the ordering come into play because.

It can be recognized that what is written inside the parenthesis is nothing but the determinant of the differential of $\tilde{f}$, which implies the assertion of item 5.

6. It is a simple consequence of the chain rule $d(h \circ f)_p = dh_{f(p)} \circ df_p$ and of the fact that $id^* = id$.

$\square$

6.2 The exterior derivative

The exterior derivative, introduced by Élie Cartan in 1899, is a very important function which allows us to extend the definition of differential of a smooth function. The calculus that uses the exterior derivative instead of the canonical differential is called **exterior calculus** and it allows for an intrinsic, coordinate-independent, generalization of vector calculus to smooth manifolds more general than $\mathbb{R}^n$, including, noticeably, the theorems of Gauss, Green and Stokes.

As we will see shortly, the exterior derivative and the pull-back have a special property when considered together: they commute when applied to differential forms.

In general, we have given an intrinsic, coordinate-free, definition of an operator and then we have searched for its local representation, this time we reverse the scheme by first providing the local definition of the exterior derivative because of its extreme simplicity.

To better understand this definition, recall that we have set $\mathbb{A}^0(M) \equiv \mathscr{C}^\infty(M)$, so we define the exterior derivative of a 0-form, which is a smooth function $f \in \mathscr{C}^\infty(M)$ as the classical differential, i.e. the following function that is applied to 0-form and gives back 1-forms:

$$\begin{array}{rccl} d: & \mathbb{A}^0(M) & \longrightarrow & \mathbb{A}^1(M) \\ & f & \longmapsto & df = \sum\limits_{j=1}^n \frac{\partial f}{\partial x^j} dx^j. \end{array}$$

Using this fact it is quite natural to define the exterior derivative of a generic k-form in coordinates by adding the wedge product with an extra differential form given by the local coefficients of the k-form itself!

Def. 6.2.1 (Exterior derivative) *The exterior derivative*

$$d : \mathbb{A}^k(M) \longrightarrow \mathbb{A}^{k+1}(M)$$

transforms the k-form with local expression

$$\omega = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \omega_{i_1,\ldots,i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \quad \omega_{i_1,\ldots,i_k} \in \mathscr{C}^\infty(M)$$

in the $(k+1)$-form with local expression

$$d\omega = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \underbrace{\left(\sum_{j=1}^n \frac{\partial \omega_{i_1,\ldots,i_k}}{\partial x^j} dx^j \right)}_{d\omega_{i_1,\ldots,i_k}} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} \tag{6.3}$$

Theorem 6.2.1 (Identifying properties of the external derivative) *For all $k \in \mathbb{N}$, the external derivative $d : \mathbb{A}^k(M) \longrightarrow \mathbb{A}^{k+1}(M)$ is the only map such that:*

1. *d is $\mathbb{R}$-linear and, if $\omega \in \mathbb{A}^k(M)$ and $\eta \in \mathbb{A}^\ell(M)$, then*

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta,$$

which is an almost Leibniz-like behavior w.r.t. the wedge product of differential forms, where of course the 'almost' refers to the presence of the extra factor $(-1)^k$;

2. $d \circ d : \mathbb{A}^k(M) \longrightarrow \mathbb{A}^{k+2}(M)$ *is the identically zero map, i.e.* $d \circ d \equiv 0$*;*

3. $d : \mathbb{A}^0(M) \longrightarrow \mathbb{A}^1(M)$ *is the differential of a smooth function.*

The proof is a straightforward application of the definition in local coordinates.

While the properties in item 1. confirm a derivative-like behavior of d, property 3. is not expected and it happens to have a fundamental importance. First of all, the fact that $d \circ d \equiv 0$ implies that the infinite sequence of functions defined by

$$\mathbb{A}^0(M) \underset{d}{\longrightarrow} \mathbb{A}^1(M) \underset{d}{\longrightarrow} \mathbb{A}^2(M) \underset{d}{\longrightarrow} \cdots \underset{d}{\longrightarrow} \mathbb{A}^k(M) \underset{d}{\longrightarrow} \mathbb{A}^{k+1}(M) \underset{d}{\longrightarrow} \mathbb{A}^{k+2}(M) \underset{d}{\longrightarrow} \cdots$$

is a **complex of vector spaces**, i.e. the composition of two consecutive maps between these vector spaces is identically null.

We can single out two special classes of k-forms via d.

Def. 6.2.2 (Closed and exact k-forms in terms of the exterior derivative) *A k-form is said to be closed if it belongs to the vector space*

$$Z^k(M) := \ker\{d : \mathbb{A}^k(M) \to \mathbb{A}^{k+1}(M)\},$$

while it is said to be exact if it belongs to the vector space

$$B^k(M) := \mathrm{Im}\{d : \mathbb{A}^{k-1}(M) \to \mathbb{A}^k(M)\}.$$

It is clear that

$$B^k(M) \subseteq Z^k(M),$$

so **exact k-forms are also closed**, but the other way round is not guaranteed. In fact, in $B^k(M)$ we have k-forms that are obtained via the external derivative of $k-1$ forms, called potentials, so, if we apply to them another time d we surely obtain the identically 0 map because $d \circ d \equiv 0$, hence the elements of $B^k(M)$ are necessarily in the kernel of d, i.e. they constitute a subset of $Z^k(M)$.

Def. 6.2.3 (de Rham coomology groups) *For all $k \in \mathbb{N}$, the quotient spaces*

$$H^k_{dR}(M) := Z^k(M)/B^k(M),$$

are called the de-Rham coomology groups of the manifold M.

$H^k_{dR}(M)$ measures how much closed and exact k-forms on M differ, in fact $H^k_{dR}(M)$ is trivial if $B^k(M) = Z^k(M)$, while it grows as $B^k(M)$ becomes more and more different than $Z^k(M)$.

In spite of the fact that the de-Rham coomology groups of the manifold M are defined using heavily the differential structure of M, their analysis provides many interesting properties on the topology of M.

As announced before, d and the pull-back are interchangeable, as the following result states.

Theorem 6.2.2 *Let $f : M \to N$ be a smooth function between manifolds, then for all $k \in \mathbb{N}$ and all $\omega \in \mathbb{A}^k(N)$ we have:*

$$f^*(d\omega) = d(f^*\omega),$$

i.e. the external derivative and the pull-back commute.

Let us finally give the intrinsic, coordinate-free, definition of external derivative.

Def. 6.2.4 (Intrinsic definition of exterior derivative) *We have:*

1. *If $\omega \in \mathbb{A}^1(M)$ and $X, Y \in \mathfrak{X}(M)$, then*

$$d\omega(X,Y) := X(\omega(Y)) - Y(\omega(X)) - \omega([X,Y]).$$

2. *More generally, if $\omega \in \mathbb{A}^k(M)$ and $X_1, X_2, \ldots, X_{k+1} \in \mathfrak{X}(M)$, then*

$$\begin{aligned} d\omega(X_1, X_2, \ldots, X_{k+1}) := &\sum_{1 \leqslant i \leqslant k+1} (-1)^{i-1} X_i(\omega(X_1, \ldots, \cancel{X_i}, \ldots, X_{k+1})) \\ &+ \sum_{1 \leqslant i < j \leqslant k+1} (-1)^{i+j} \omega([X_i, X_j], X_1, \ldots, \cancel{X_i}, \ldots, \cancel{X_j}, \ldots, X_{k+1}). \end{aligned}$$

The formula appearing in item 2. of the previous definition is the intrinsic version of the locally defined, eq. (6.3).

Let us explain in a more detailed form the previous definitions, which may seem quite complicated at first glance.

If $\omega \in \mathbb{A}^1(M)$, then $d\omega \in \mathbb{A}^2(M)$, so it must be applied to two vector fields. Moreover, ω applied to a vector field gives rise to a smooth scalar function, and the vector field is a derivation of $\mathscr{C}^\infty(M)$, hence the definition in item 1. is perfectly well-posed.

The definition in item 2. is simply the generalization of these ideas, in some books, instead of appearing canceled, the vector fields which are removed appear with a 'hat symbol' above them. The coefficients $(-1)^{i-1}$ and $(-1)^{i+j}$ depend on how the range of the indices and may vary from book to book. In any case, the definition in item 2. is the intrinsic writing of the local formula (6.3).

Having introduced the pull-back and the exterior derivative, we can now pass to treat the problem of how to integrate k-forms. We will start with the easiest case, that of the integration of n-forms in open sets of $\mathbb{R}^n$, and then we will deal with cases of increasing difficulty.

6.3 Integration of n-forms in open sets of $\mathbb{R}^n$

Let $V \subset \mathbb{R}^n$ an open set and let $\eta = f(x)\, dx^1 \wedge \cdots \wedge dx^n$ a n-form with compact support, which simply means that the smooth coefficient function f is null outside a compact set $K \subset V$.

Def. 6.3.1 *With the notations above, the integral of η is defined as*

$$\int_V \eta = \int_V f(x)\, dx^1 \wedge \cdots \wedge dx^n := \int_V f(x) dm(x),$$

where $dm(x)$ is the Lebesgue measure on V.

Let us now try to extend this simple case to a smooth manifold M. Consider a local chart (U, φ) and an n-form ω on M with compact support $K \subset U$. Using the property that $\phi : U \to \phi(U) \equiv V \subset \mathbb{R}^n$ is a homeomorphism, hence, in particular, an invertible function, we have $\varphi^{-1} : V \subset \mathbb{R}^n \to U \subset M$. This shows that φ^{-1} has the correct domain and range to pull-back ω to an open subset of $\mathbb{R}^n$:

$$\eta := (\varphi^{-1})^*\omega.$$

Since we know how to integrate η, the easiest way to extend the definition of integral of an n-form to a smooth manifold seems to define it as follows:

$$\int_U \omega := \int_V \eta = \int_{\varphi(U)} (\varphi^{-1})^*\omega.$$

However, this definition is intrinsic, i.e. independent of the local chart chosen, if and only if on the intersection of two local charts the integral is the same, otherwise it would depend on the local coordinates chosen to express the n-form.

Let us check if this is satisfied or not: consider two local charts $(U, \varphi \equiv (x^1, \ldots, x^n))$ and $(\tilde{U}, \tilde{\varphi} \equiv (y^1, \ldots, y^n))$ such that the support of ω is contained in $U \cap \tilde{U}$. The situation can be visualized as follows:

$$\begin{tikzcd}
 & U \cap \tilde{U} \arrow[dl, "\varphi"'] \arrow[dr, "\tilde{\varphi}"] & \\
(x^1, \ldots, x^n) \in \varphi(U \cap \tilde{U}) \subset \mathbb{R}^n \arrow[rr, dashed, bend right, "F"'] & & (y^1, \ldots, y^n) \in \tilde{\varphi}(U \cap \tilde{U}) \subset \mathbb{R}^n
\end{tikzcd}$$

where F is the transition functions between the coordinates $(x^1, \ldots, x^n)$ which live in the subset $\varphi(U \cap \tilde{U})$ of a copy of $\mathbb{R}^n$ to the the coordinates $(y^1, \ldots, y^n)$ which live in the subset $\tilde{\varphi}(U \cap \tilde{U})$ of a different copy of $\mathbb{R}^n$, so $F(x^1, \ldots, x^n) = (y^1, \ldots, y^n)$.

We can operate the pull-back of η to $\varphi(U \cap \tilde{U})$ via φ^{-1} and the pull-back of η to $\tilde{\varphi}(U \cap \tilde{U})$ via $\tilde{\varphi}^{-1}$ obtaining two local expressions

$$\eta := (\varphi^{-1})^*\omega = f(x)\, dx^1 \wedge \cdots \wedge dx^n \quad \text{and} \quad \tilde{\eta} := (\tilde{\varphi}^{-1})^*\omega = \tilde{f}(y)\, dy^1 \wedge \cdots \wedge dy^n,$$

with $f, \tilde{f} \in \mathscr{C}^\infty$ on their domains.

η and $\tilde{\eta}$ represent the same differential form ω, but w.r.t. different coordinate systems. Since the coordinate transformation is expressed by the transition function F, we have that $\eta = F^*\tilde{\eta}$. If the integral of ω did not depend on the coordinate system chosen, then we should have:

$$\int_{U \cap \tilde{U}} \omega = \int_{\varphi(U \cap \tilde{U})} \eta = \int_{\tilde{\varphi}(U \cap \tilde{U})} \tilde{\eta}.$$

To verify if this is true or not, we can use the fact that we know how to explicitly express in coordinates the pull back of a differential form, so:

$$\eta = F^*(\tilde{\eta}) = F^*(\tilde{f}(y)\, dy^1 \wedge \cdots \wedge dy^n) \underset{(6.2)}{=} (\tilde{f} \circ F(x)) \det\left(\frac{\partial F^i}{\partial x^j}\right) dx^1 \wedge \cdots \wedge dx^n.$$

We thus find:

$$\int_{\varphi(U\cap\tilde{U})} \eta = \int_{\varphi(U\cap\tilde{U})} \tilde{f}(F(x)) \det\left(\frac{\partial F^i}{\partial x^j}\right) dm(x), \tag{6.4}$$

which differs from the expression that we would get by using the well-known change of variable formula in multiple integrals within the Lebesgue integration theory, i.e.

$$\int_{\tilde{\varphi}(U\cap\tilde{U})} \tilde{\eta} = \int_{\tilde{\varphi}(U\cap\tilde{U})} \tilde{f}(y)dm(y) = \int_{\varphi(U\cap\tilde{U})} \tilde{f}(F(x)) \left|\det\left(\frac{\partial F^i}{\partial x^j}\right)\right| dm(x). \tag{6.5}$$

The lesson that we learn by comparing the two last formulae is that, in general, it is not possible to define the integral of a differential form over a generic manifold, unless the determinant of the transition function between two different local charts is always positive. These manifolds are precisely those called *oriented manifold*, that we define in the following section.

6.4 Oriented manifolds and volume forms

The definition of oriented manifolds is more easily understood if we first discuss the orientation of generic *real* vector spaces.

Let V be a real vector space and $(v_1, \dots, v_n)$ a basis, hence an *ordered* set of linearly independent generators of V. Let also $(v'_1, \dots, v'_n)$ be another basis of V, then we know that the relationship between them is expressed by the following formula:

$$v'_i = v_j a^j_i,$$

where $A = (a^j_i)^n_{i,j=1} \in \mathrm{GL}(n, \mathbb{R})$ is the change of basis matrix.

Def. 6.4.1 (Equioriented bases) *If* $\det(A) > 0$*, the two bases* $(v_1, \dots, v_n)$ *and* $(v'_1, \dots, v'_n)$ *are said be equioriented*[1]*.*

It can be proven quite easily that the relation of equiorientation is an equivalence one and that there are only two equivalence classes w.r.t. it because, being different from 0, $\det(A)$ can only be strictly positive or strictly negative.

Def. 6.4.2 (Orientation of a real vector space) *An orientation of a real vector space is an equivalence class of bases w.r.t. the equiorientation relation.*

So, there exist only two orientations of a real vector space and we can say that in the positive orientation we put all the bases for which $\det(A) > 0$ and in the negative orientation we put all the bases for which $\det(A) < 0$.

To have an example at hand, consider the simplest non trivial example, that of a vector space V of dimension 2, then the two bases (v_1, v_2) and (v_2, v_1) belong to different orientations because:

$$(v_2, v_1) = (v_1, v_2)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the determinant of the change of basis matrix is -1.

[1] of course, the need of a *real* vector space is that, if V where complex, $\det(A)$ would still be different than 0, but, being a complex number, it would not be meaningful to order it w.r.t. 0.

Let us now pass to a manifold M of dimension n: it is clear that the role of the generic vector space V must now be replaced by the natural vector spaces that appear when analyzing M, i.e. the tangent spaces T_pM for each point $p \in M$. Since we need a basis, we choose a local chart $(U, \varphi \equiv (x^1, \ldots, x^n))$ in p and we consider the coordinate basis vectors of T_pM given by $(\partial_1|_p, \ldots, \partial_n|_p)$.

Def. 6.4.3 *It is custom, though completely arbitrary, to define the basis* $(\partial_1|_p, \ldots, \partial_n|_p)$ *as positively oriented and so to define the positive orientation of* T_pM *as the equivalence class in which* $(\partial_1|_p, \ldots, \partial_n|_p)$ *belongs.*

The vector fields $p \mapsto \partial_j|_p$ depend smoothly on $p \in U$, thus the local chart (U, φ) induces an orientation on T_pM which depends smoothly on $p \in U$.

This remark allows us to extend the definition of equiorientation to manifolds in a natural way.

Def. 6.4.4 *Two local charts* $(U_\alpha, \varphi_\alpha)$ *and* (U_β, φ_β) *of* M *are equioriented if the determinant of the differential of the transition functions* $\eta_{\alpha\beta} := \varphi_\alpha \circ \varphi_\beta^{-1}$ *is positive when evaluated in every point of* $\varphi_\beta(U_\alpha \cap U_\beta)$.

By extension we get the definition of oriented atlas.

Def. 6.4.5 (Oriented atlas) *An atlas* $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}_\alpha$ *of the* M *is oriented if every couple of charts of* $\mathcal{A}$ *is equioriented.*

Two different oriented atlases are said to be equioriented if their union is still an oriented atlas.

The last definition means, in practice, that the orientation defined by the local charts of the first atlas is the same as the one defined by those of the second atlas.

We can finally give the natural definition of oriented manifold as follows.

Def. 6.4.6 (Oriented manifold) *A manifold* M *is oriented if it admits an oriented atlas. An orientation of* M *is an equivalence class of oriented atlases of* M.

A simple *technique to invert the orientation induced by a chart* is the following: let us consider a local chart $(U, \varphi \equiv (x^1, \ldots, x^{n-1}, x^n))$ of a manifold M, then the chart $\overline{\varphi} \equiv (x^1, \ldots, x^n, x^{n-1})$ induces on U the opposite orientation because, by direct computation, it can be verified that $\det(d(\overline{\varphi} \circ \varphi^{-1})) < 0$. In the following, whenever we want to change the orientation induced by a chart φ, we will use the symbol $\overline{\varphi}$. It is canonical to denote the manifold with opposite orientation w.r.t. M as $-M$.

Orientation of spheres. Spheres constitute one of the easiest example of manifolds, thus it is natural to ask ourselves if they are oriented or not. We will show that the answer is positive by analyzing the more general situation of a manifold M with an atlas $\mathcal{A}$ given by only two local charts (U_0, φ_0) and (U_1, φ_1), with $U_0 \cap U_1$ connected. In this case there exists only one transition function, i.e. $\varphi_1 \circ \varphi_0^{-1}$ and its inverse $\varphi_0 \circ \varphi_1^{-1}$. We know that $\det(d(\varphi_1 \circ \varphi_0^{-1})) = \det(d(\varphi_0 \circ \varphi_1^{-1}))^{-1} \neq 0$ on $U_0 \cap U_1$, so it cannot change sign on this set, hence it is always either positive or negative. If $\det(d(\varphi_1 \circ \varphi_0^{-1})) = \det(d(\varphi_0 \circ \varphi_1^{-1}))^{-1} > 0$, then $\mathcal{A}$ is oriented because the chart are equioriented, if not, it means that the two charts define opposite orientations, in this case we can simply swap one of them, say φ_1 with $\overline{\varphi}_1$, obtaining an oriented atlas.

In the case of the spheres S^n, with $n \geqslant 2$, we can build an atlas with the two stereographic projections from the north and the south pole, (U_N, φ_N) and (U_S, φ_S), with $U_N \cap U_S$ being the entire sphere without the poles, which is a connected set for all $n \geqslant 2$. It follows that all the spheres S^n with $n \geqslant 2$ are oriented manifolds. By a simple direct computation, it can be proven that S^1 is an oriented manifold too even if $U_N \cap U_S$ is not connected in this case.

Let us now come to the link between orientation and integration of differential forms that we have anticipated in the previous section. To better understand this link, let us consider a manifold of dimension n and the vector bundle $\Lambda^n T^*M$, i.e. the n-th external power of the cotangent bundle of M, which we know has dimension 1, i.e. it is a line bundle.

Imagine now that it exists a n-form $\nu \in \mathbb{A}^n(M)$ which is never null when computed in every point of M. Then, for every $p \in M$, ν_p is a basis of $\Lambda^n T_p^*M$, which implies that every other n-form $\omega \in \mathbb{A}^n(M)$ is written as $\omega = f\nu$, where $f \in \mathscr{C}^\infty(M)$ is the suitable smooth scalar function which connects ω with ν. For reasons that will be clear soon, it is convenient to give a name to ν and the most appropriate one is contained in the following definition.

Def. 6.4.7 (Volume or orientation form) *A n-form $\nu \in \mathbb{A}^n(M)$ on an n-dimensional manifold M which is never null when computed in every point of M is called a volume form, or a volume element, or again an orientation form, on M.*

The next important result justifies why we can interchange volume and orientation in the previous definition.

Theorem 6.4.1 (Equivalence between orientation and existence of a volume form) *Let M be an n-dimensional manifold, then M is oriented if and only if it admits a volume form $\nu \in \mathbb{A}^n(M)$.*

Proof. The proof is constructive for both implications.

$\boxed{\Longrightarrow}$: suppose that M admits a volume form $\nu \in \mathbb{A}^n(M)$, then we must exhibit the existence of an oriented atlas $\mathcal{A}$ for M. We denote with $\mathcal{A}$ the set of all local charts $(U_\alpha, \varphi_\alpha)$ of M that verify the following requests:

1. they are compatible with the original differential structure of M, i.e. the transition functions between these charts and those of the atlas that provides M with a differential structure are smooth;

2. if we evaluate the volume form ν on the coordinate tangent vectors defined by the chart $(U_\alpha, \varphi_\alpha)$, which, for every $p \in U_\alpha$, we denote with $\partial_{1,\alpha}|_p, \ldots, \partial_{n,\alpha}|_p$, we obtain a positive value, i.e.
$$\nu_p(\partial_{1,\alpha}|_p, \ldots, \partial_{n,\alpha}|_p) > 0, \quad \forall p \in U_\alpha.$$

We want to prove that the collection $\mathcal{A}$ of local charts built with these two requests is an oriented atlas.

Let us first verify that $\mathcal{A}$ is indeed an atlas for M. Provided that we suitably restrict a local chart domain U in a point p, we can always consider it connected. In this situation, the scalar function
$$U \ni p \longmapsto \nu_p(\partial_1|_p, \ldots, \partial_n|_p)$$
is smooth and never null, because when have supposed that ν is a volume form, hence it has constant sign on the connected set U. If this sign is positive, then $(U, \varphi) \in \mathcal{A}$, if it is negative,

then the chart $(U, \overline{\varphi}) \in \mathcal{A}$ thanks to the fact that ν_p is alternating, so if we switch two tangent vectors the result changes sign! This shows that $\mathcal{A}$ is an atlas for M.

Now let us verify that $\mathcal{A}$ is oriented. If $(U_\alpha, \varphi_\alpha \equiv (x^1_\alpha, \ldots, x^n_\alpha))$ and $(U_\beta, \varphi_\beta \equiv (x^1_\beta, \ldots, x^n_\beta))$ are two local charts defining different local coordinate systems, then we know that, in $U_\alpha \cap U_\beta$ it holds that

$$\nu(\partial_{1,\alpha}, \ldots, \partial_{n,\alpha}) = \det\left(\left(\frac{\partial x^h_\alpha}{\partial x^k_\beta}\right)^n_{h,k=1}\right)\nu(\partial_{1,\beta}, \ldots, \partial_{n,\beta}),$$

where of course it appears the determinant of the Jacobian matrix of the change of coordinates.

Thanks to the fact that ν is supposed to be a volume form, both $\nu(\partial_{1,\alpha}, \ldots, \partial_{n,\alpha})$ and $\nu(\partial_{1,\alpha}, \ldots, \partial_{n,\alpha})$ are positive-valued, hence also the determinant appearing in the right-hand side of the previous formula must be positive, which implies that the two generic charts are equioriented. So, $\mathcal{A}$ is oriented.

$\boxed{\Longleftarrow}$: suppose now that it exists an oriented atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha) \equiv (x^1_\alpha, \ldots, x^n_\alpha)\}$ of M, we have to prove that we can exhibit the existence of a volume form on M. Thanks to the local coordinate system defined by each chart indexed with α, we can define a collection of n-forms

$$\nu_\alpha = dx^1_\alpha \wedge \cdots \wedge dx^n_\alpha$$

defined on the open subsets U_α of M. As always, in order to extend a locally-defined object to the entire manifold, we use a partition of the unity $\{\rho_\alpha\}_\alpha$ subordinated to the covering $\{U_\alpha\}_\alpha$. Thanks to the properties of the partitions of the unity the sum

$$\nu = \sum_\alpha \rho_\alpha \, dx^1_\alpha \wedge \cdots \wedge dx^n_\alpha,$$

with ρ_α with range in $(0, 1]$, defines a n-form on the whole M, i.e. $\nu \in \mathbb{A}^n(M)$. If we prove that ν is never null, then the proof will be completed.

First of all we notice that, thanks to the hypothesis that $\mathcal{A}$ is oriented, all the locally defined n-forms ν_α are never null in U_α. We must just check what happens in the intersection $U_\alpha \cap U_\beta$ of two generic charts, but we know that in this case it holds that

$$dx^1_\alpha \wedge \cdots \wedge dx^n_\alpha = \det\left(\left(\frac{\partial x^h_\alpha}{\partial x^k_\beta}\right)^n_{h,k=1}\right) dx^1_\beta \wedge \cdots \wedge dx^n_\beta,$$

hence ν_α and ν_β differ only for a strictly positive multiplicative factor, because the charts labeled by α and β belong to an oriented atlas!

As a consequence, for every point of M it exists an open neighborhood in which the n-form ν is the sum of terms which are never null and strictly positive multiples of them, hence ν can never be null itself. □

Thanks to the previous result, the following definition is well-posed.

Def. 6.4.8 *Let M be an oriented manifold of dimension n and let $\nu \in \mathbb{A}^n(M)$ be a volume form on it.*

- *A basis $(v_1, \ldots, v_n)$ of T_pM is said to be positively oriented if, for every $p \in M$, $\nu_p(v_1, \ldots, v_n) > 0$, otherwise we say that the basis is negatively oriented.*

- *A local chart (U, φ) is oriented if $(\partial_1, \ldots, \partial_n)$ is a positively oriented basis of T_pM.*
- *The atlas $\mathcal{A}$ given by all the oriented local charts is called the atlas associated to the volume form ν.*
- *If N is another oriented manifold and $F : M \to N$ is a local diffeomorphism, then we say that F preserves the orientation if, for all $p \in M$, the image of a positively-oriented basis of T_pM via dF_p is a a positively-oriented basis of $T_{F(p)}M$. Otherwise, if F reverses the orientation of all positively-oriented bases of T_pM we say that F inverts the orientation.*

Regarding the last item of the previous set of definitions, it is very easy to characterize orientation preserving (or inverting) local diffeomorphisms between oriented manifolds. In fact, given two generic F-related local charts (U, φ) of M and (V, ψ) of N, i.e. such that $F(U) \subseteq V$, the change of basis is represented by the Jacobian matrix of the local representation $\tilde{F} = \psi \circ F \circ \varphi^{-1}$ of F, then F preserves (risp. inverts) the orientation if and only if $\det(J\tilde{F}) > 0$ (risp. $\det(J\tilde{F}) < 0$).

6.5 Integration of n-forms on n-dimensional oriented manifolds

We are now ready to connect to define the integral of n-forms on oriented manifolds.

Theorem 6.5.1 *Let M be a n-dimensional oriented manifold and let $\omega \in \mathbb{A}^n(M)$ be a n-form with compact support K. Let us also suppose that (U, φ) and $(\tilde{U}, \tilde{\varphi})$ are two local chart such $K \subset U \cap \tilde{U}$. Then*

$$\int_{\varphi(U)} (\varphi^{-1})^* \omega = \int_{\tilde{\varphi}(U)} (\tilde{\varphi}^{-1})^* \omega.$$

Proof. In the hypotheses of the theorem, the two formulae (6.4) and (6.5) coincide because the determinant of the Jacobian matrix of the change of variable function is positive, so the two integrals are exactly the same. $\square$

Thanks to this result we can finally define the integral of a n-form with compact support.

Def. 6.5.1 *Let M be a n-dimensional oriented manifold and let $\omega \in \mathbb{A}^n(M)$ be a n-form with compact support K. Given any local chart (U, φ) of M such that $K \subset U$, the integral of ω on M is*

$$\int_M \omega := \int_{\varphi(U)} (\varphi^{-1})^* \omega.$$

The definition is of course well-posed because the choice of the local chart is totally arbitrary. This allows the freedom of choosing the local chart for which the computation of the integral is the easiest. We remark again that that the integral written at the right-hand side of the previous formula is an integral in $\mathbb{R}^n$ in the Lebegue sense.

Once solved this problem, we remain with the one of *extending the definition of integral of n-forms with a compact support which is not necessarily contained in the domain of one single chart.* It comes with no surprise that this can be achieved thanks to the partition of the unity.

Theorem 6.5.2 *Let M be a n-dimensional oriented manifold, $\omega \in \mathbb{A}^n(M)$ any n-form on M with compact support, $\mathcal{A} = \{U_\alpha, \varphi_\alpha\}_\alpha$ an oriented atlas of M and $\{\rho_\alpha\}_\alpha$ a partition of the unity subordinated to the covering $\{U_\alpha\}_\alpha$. Then, the number*

$$\sum_\alpha \int_M \rho_\alpha \omega$$

does not depend on the choice of $\mathcal{A}$ nor on $\{\rho_\alpha\}_\alpha$.

Proof. Since ω is supposed to be compactly supported, the supports of the functions ρ_α form a locally finite covering, thence the sum $\sum_\alpha \rho_\alpha \omega$ contains only a finite number of terms and so it is well-defined.

To show the independence of the choice of the atlas and the partition of the unity, let us consider another oriented atlas $\tilde{\mathcal{A}} = \{\tilde{U}_\beta, \tilde{\varphi}_\beta\}_\beta$ and another partition of the unity $\{\tilde{\rho}_\beta\}_\beta$ subordinated to the covering $\{\tilde{U}_\beta\}_\beta$.

By definition, $\sum_\beta \tilde{\rho}_\beta$ is equal to the identically constant function $\mathbf{1}$, so we can write

$$\int_M \rho_\alpha \omega = \int_M \left(\sum_\beta \tilde{\rho}_\beta \right) \rho_\alpha \omega = \sum_\beta \int_M \tilde{\rho}_\beta \rho_\alpha \omega,$$

where now $\tilde{\rho}_\beta \rho_\alpha \omega$ has support in $U_\alpha \cap \tilde{U}_\beta$ because the partitions are null outside. As a consequence, if we now add up all the terms $\int_M \rho_\alpha \omega$ in order to extend the integral to the whole manifold M, we obtain that

$$\sum_\alpha \int_M \rho_\alpha \omega = \sum_\alpha \sum_\beta \int_M \tilde{\rho}_\beta \rho_\alpha \omega,$$

does not depend on the local charts (and so on the atlases), because, thanks to theorem 6.5.1, $\int_M \tilde{\rho}_\beta \rho_\alpha \omega$ is an intrinsic object.

Analogously, if we started with the atlas $\tilde{\mathcal{A}}$ we would get

$$\sum_\beta \int_M \tilde{\rho}_\beta \omega = \sum_\alpha \sum_\beta \int_M \rho_\alpha \tilde{\rho}_\beta \omega,$$

so

$$\sum_\alpha \int_M \rho_\alpha \omega = \sum_\beta \int_M \tilde{\rho}_\beta \omega.$$

□

This last result allows us to give extend the validity of the definition of integral of a n-form with compact support K without requiring that K is entirely contained in a single chart domain.

Def. 6.5.2 (Integral of a compactly supported n-form on an oriented manifold) *Let M be a n-dimensional oriented manifold and let $\omega \in \mathbb{A}^n(M)$ be a n-form with compact support. Given any oriented atlas $\mathcal{A} = \{U_\alpha, \varphi_\alpha\}_\alpha$ of M and $\{\rho_\alpha\}_\alpha$ any partition of the unity subordinated to the covering $\{U_\alpha\}_\alpha$, the integral of ω on M is defined as follows:*

$$\int_M \omega := \sum_\alpha \int_M \rho_\alpha \omega = \sum_\alpha \int_{\varphi(U_\alpha)} \rho_\alpha (\varphi_\alpha^{-1})^* \omega.$$

We can also define some useful integrals correlated with the one above.

Def. 6.5.3 (Integral of a compactly supported smooth scalar function on M) *Let $\nu \in \mathbb{A}^n$ be a volume form which induces an orientation on the n-dimensional manifold M and $f \in \mathscr{C}_c^\infty(M)$, then*

$$\int_M f := \int_M f\nu.$$

It is important to stress that the integral of f just defined depends on the volume form ν chosen. In particular, if M is a *compact manifold*, by taking the identically constant function $\mathbf{1}$ we can define the volume of M as follows.

Def. 6.5.4 (Volume of a compact oriented manifold) *Let $\nu \in \mathbb{A}^n$ be a volume form which induces an orientation on the n-dimensional manifold M, then*

$$\text{vol}_\nu(M) := \int_M \nu.$$

This definition explains why the n-forms with the properties of ν are called volume forms!

Remark: we have managed to define the integral of a compactly supported n-form on a n-dimensional manifold by restricting the definition to oriented manifolds. It is possible to define the integral on manifolds without demanding that they are oriented by modifying the objects of integration: in this case, instead of n-forms, we can integrate so-called *densities*, which are objects that transform under changes of coordinates with a factor that is the absolute value of the Jacobian determinant. The price to pay for this request is that densities are not tensors anymore.

If $-M$ is the manifold with opposed orientation w.r.t. M, then, for all $\omega \in \mathbb{A}^n$ with compact support we have

$$\int_{-M} \omega = -\int_M \omega,$$

in fact, if (U, φ) is a positively-oriented chart of M and $(U, \overline{\varphi})$ is the corresponding negatively-oriented chart, then the change of variable formula for multiple integrals implies that

$$\int_{\varphi(U)} (\varphi^{-1})^*\omega = -\int_{\overline{\varphi}(U)} (\overline{\varphi}^{-1})^*\omega,$$

from which the formula $\int_{-M} \omega = -\int_M \omega$ descends immediately.

6.5.1 Integration of k-forms on k-dimensional oriented sub-manifolds

It is possible to extend the concept of integral of a compactly supported n-form on a n-dimensional oriented manifold to k-forms on k-dimensional oriented sub-manifolds, with $k < n$, where the submanifold can be both immersed or embedded in M.

It is important to stress that, in this case, the manifold M is not required to be oriented, the only two requests that must be satisfied are that the sub-manifold is indeed oriented and that its dimension coincides with the degree of the differential form.

So, let us consider a non-necessarily oriented manifold M and let S be an immersed sub-manifold with the additional technical request that the immersion $F : S \to M$ is a *proper* *function*, i.e. that the inverse image of any compact subset of M is compact in S. We demand this because, in this case, if $\omega \in \mathbb{A}^k(M)$ has compact support in M, the k-form $F^*\omega$ has compact support in S.

Under these hypotheses, we can define the integral of ω on S, that we will denote with the symbol $\int_F \omega$ by defining

$$\int_F \omega := \int_S F^*\omega.$$

In particular, if $F = \iota : S \hookrightarrow M$ is an oriented embedded sub-manifold of M, then $\iota^*\omega$ is simply the restriction of ω to S. In this case we simply write $\int_S \omega$ or $\int_S \omega|_S$ instead of $\int_\iota \omega$.

6.6 The Stokes theorem

The Stokes theorem is the most important result about integration of n-forms and it compares an integral performed over a manifold with one over its boundary.

We have developed the theory of manifolds with border in chapter 1, now we introduce an orientation on them, in order to be able to perform integrals in a meaningful way.

6.6.1 Orientation of manifolds with boundary

The definition of oriented manifold with boundary is exactly the same as the one given for manifold without boundary, i.e. definition 6.4.6.

Importantly, the boundary ∂M of an oriented manifold with boundary M inherits automatically an orientation itself by restricting the charts of the atlas of M to the boundary ones, as assured by the following theorem, whose easy proof can be found e.g. in [10].

Theorem 6.6.1 *Let $U_0, U_1 \subseteq \mathbb{H}^n$ two open sets of the n-dimensional manifold with boundary M and let $\tilde{U}_0, \tilde{U}_1$ such that their intersections with the boundary of $\mathbb{H}^n$, i.e. $\tilde{U}_j = U_j \cap \partial\mathbb{H}^n$, $j = 1, 2$, are not empty. Let $F : U_0 \to U_1$ be a diffeomorphism with positive Jacobian determinant in every point of U_0.*

Then, the Jacobian determinant of the restriction $\tilde{F} := F|_{\tilde{U}_0} : \tilde{U}_0 \to \tilde{U}_1$, interpreted as a diffeomorphism between to open sets of $\mathbb{R}^{n-1}$, is positive in every point of $\tilde{U}_0$.

In order to have a simple statement of the Stokes theorem, it is convenient to define the orientation induced on ∂M as follows.

Def. 6.6.1 (Orientation of the border) *Let M be a n-dimensional manifold with boundary ∂M and let $\mathcal{A}$ and $\partial\mathcal{A}$ be the oriented atlases of M and ∂M, the latter being the one induced by $\mathcal{A}$ restricting the charts to the border ones.*

We define the orientation of ∂M as the one defined by ∂A if n is even and the opposed one if n is odd.

We will see that, if we do not change the orientation when n is odd, in the Stokes theorem there would appear a coefficient $(-1)^n$, which we can avoid thanks to the convention adopted in the previous definition.

6.6.2 Extension of tangent, cotangent and tensor bundles to manifolds with boundary

Since in the Stokes theorem we must deal with differential forms, we must extend them on manifolds with boundary. The procedure to extend the tangent and cotangent spaces and bundles to manifolds with boundary is quite easy and the technical details can be found again in [10]. In what follows, we simply recap the main results that we need in order to arrive as quick as possible to the Stokes theorem.

- The canonical inclusion $\iota : \partial M \hookrightarrow M$ is a smooth immersion.
- Thanks to the previous item, we can identify $T_p(\partial M)$ with its image via $d\iota_p$.
- Explicitly, $T_p(\partial M)$ is the vector subspace of T_pM given by the tangent vectors in $\mathrm{Der}_p(M)$ that take null values on the constant functions of $\mathscr{C}^\infty(M)$.
- One fixed a n-border local chart $(U, \varphi \equiv (x^1, \ldots, x^n))$, a basis for $T_p(\partial M)$ is given by the derivations $(\partial_1|_p, \ldots, \partial_{n-1}|_p)$, with $p \in \partial M$, hence $\mathrm{dim} T_p(\partial M) = n-1$ as expected.
- Having defined $T_p(\partial M)$ for all $p \in \partial M$ and having already at disposal $T_p(\mathrm{int}(M))$, since $\mathrm{int}(M)$ is a manifold without boundary, we known how to define T_pM for all $p \in M$.
- The disjoint union of the tangent spaces and their duals result in the tangent and cotangent bundles (respectively) to the manifold with boundary M and so also all the tensor bundles $T^p_q(M)$ can be defined.

So, there is no problem to define the concept of n-form on manifolds with boundaries.

If $\omega \in \mathbb{A}^k(M)$, we can restrict is on the border ∂M in the usual way by taking the pull-back via the canonical inclusion $\iota : \partial M \hookrightarrow M$:

$$\omega|_{\partial M} = \iota^* \omega.$$

Some care must be taken when considering the restriction, in fact it may happen that $\omega|_{\partial M}$ is null even if ω is not. The easiest example of this behavior is provided by a n-form ω of a n-dimensional manifold with boundary: in this case we obtain a n-form $\omega|_{\partial M}$ defined on the $(n-1)$-dimensional manifold ∂M, however there are no non-trivial n-forms on a $(n-1)$-dimensional manifold, hence $\omega|_{\partial M}$ must be null.

The same can happen for k-forms with $k < n$. For example, in a local chart where the last coordinate x^n is constant on ∂M, every non-null form that can be written as $\omega = \eta \wedge dx^n$ is nullified when restricted to the border because x^n is constant on ∂M, hence $dx^n = 0$ on ∂M.

6.6.3 Integration on manifolds with border

If M is a n-dimensional oriented manifold with boundary and $\eta \in \mathbb{A}^n(M)$ is a compactly supported n-form on M, where the support can intersect non-trivially ∂M, then we can define in the same way as done for manifolds without boundary the integral

$$\int_M \eta.$$

Analogously, if ω is a compactly supported $(n-1)$-form in M, its restriction $\omega|_{\partial M}$ is a $(n-1)$-form on the $(n-1)$-manifold without boundary ∂M, hence the integral $\int_{\partial M} \omega|_{\partial M}$ is perfectly defined. since it is quite evident that when we extend the integral over ∂M we are considering the restriction of ω to the border, it is custom to write the last integral simply as

$$\int_{\partial M} \omega.$$

By convention, for manifolds without boundary, i.e. when $\partial M = \varnothing$, we set $\int_{\partial M} \omega \equiv 0$.

We are now ready for the Stokes theorem.

6.6.4 The Stokes theorem

Given a compactly supported $(n-1)$-form ω on a n-dimensional oriented manifold with boundary M, its differential $d\omega$ gives back a n-form on M. The Stokes theorem guarantees the non-intuitive fact that the integral of the $(n-1)$-form ω on the $(n-1)$-manifold ∂M coincides exactly with the integral of the $(n-1)$-form ω on the $(n-1)$-manifold M.

Theorem 6.6.2 (Stokes theorem) *Let M be a n-dimensional oriented manifold with boundary ∂M endowed with the orientation defined in 6.6.1.*

Then, for all compactly supported $\omega \in \mathbb{A}^{n-1}(M)$ we have

$$\boxed{\int_M d\omega = \int_{\partial M} \omega}.$$

Proof. The strategy of the proof consists in first proving the theorem in the case of $\mathbb{R}^n$, which will be extended to generic manifolds without boundary, then in the case of $\mathbb{H}^n$, which will be extended to generic manifolds with boundary, and finally to consider the general case.

Let us start by supposing that $M = \mathbb{R}^n$. In this case, a generic $(n-1)$-form ω on $\mathbb{R}^n$ is the sum of terms of the type

$$\omega = f(x) dx^1 \wedge \cdots \wedge dx^{n-1},$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, we known that this is always possible, modulo a permutation of the coordinates. Since d and $\int_M$ are $\mathbb{R}$-linear operators, the result obtained reasoning on this term can be extended to the general case.

Let us compute the differential of ω:

$$\begin{aligned} d\omega &= \left(\sum_{j=1}^{n} \frac{\partial f}{\partial x^j} dx^j \right) dx^1 \wedge \dots dx^{n-1} \\ &= \frac{\partial f}{\partial x^1} dx^1 \wedge dx^1 \wedge \cdots \wedge dx^{n-1} + \cdots + \frac{\partial f}{\partial x^{n-1}} dx^{n-1} \wedge dx^1 \cdots \wedge dx^{n-1} \\ &\quad + \frac{\partial f}{\partial x^n} dx^n \wedge dx^1 \cdots \wedge dx^{n-1}. \end{aligned}$$

The terms contained in the second line are all null because the wedge product is alternating, so we remain with

$$d\omega = \frac{\partial f}{\partial x^n} dx^n \wedge dx^1 \cdots \wedge dx^{n-1},$$

to bring dx^n at the end of the sequence of wedge products, we must perform $n-1$ flips, so, again by the alternating property of $\wedge$, we can rewrite the previous expression as

$$d\omega = (-1)^{n-1} \frac{\partial f}{\partial x^n} dx^1 \cdots \wedge dx^{n-1} \wedge dx^n.$$

Since we are in $\mathbb{R}^n$, we can integrate $d\omega$ using the Lebesgue measure on $\mathbb{R}^n$ as follows

$$\int_{\mathbb{R}^n} d\omega = (-1)^{n-1} \int_{\mathbb{R}^n} \frac{\partial f}{\partial x^n} dx^1 \dots dx^n,$$

now we apply the Fubini theorem for multiple integrals starting the integration with dx^n:

$$\int_{\mathbb{R}^n} d\omega = (-1)^{n-1} \int_{\mathbb{R}^{n-1}} \left(\int_{-\infty}^{+\infty} \frac{\partial f}{\partial x^n} dx^n \right) dx^1 \dots dx^{n-1},$$

but in the factor between parenthesis we are integrating a derivative, thence

$$\int_{-\infty}^{+\infty} \frac{\partial f}{\partial x^n}(x^1, \dots, x^{n-1}, x^n) dx^n = f(x^1, \dots, x^{n-1}, t)\big|_{t\to-\infty}^{t\to+\infty} = 0,$$

because ω, and so f, has compact support! As a consequence,

$$\int_{\mathbb{R}^n} d\omega = 0.$$

On the other hand, $\mathbb{R}^n$ is a manifold without border, hence

$$\int_{\partial\mathbb{R}^n} \omega = 0,$$

and so the Stokes theorem is proven for $M = \mathbb{R}^n$.

We now consider the case $M = \mathbb{H}^n$, for which the border is $\partial M = \{(x^1, \dots, x^{n-1}, 0) \in \mathbb{R}^n\}$. If we write $\mathbf{x} = (x^1, \dots, x^{n-1})$, the generic $(n-1)$-form on $\mathbb{H}^n$ can be written as

$$\omega(\mathbf{x}, x^n) = \sum_{j=1}^{n} g_j(\mathbf{x}, x^n) dx^1 \wedge \dots \wedge \cancel{dx^j} \wedge \dots dx^n.$$

To compute $\int_{\partial\mathbb{H}^n} \omega$ we must restrict ω to $\partial\mathbb{H}^n$, but on it x^n is constant because it is always 0, so also $dx^n = 0$, it follows that, on $\partial\mathbb{H}^n$, in the previous expression of ω all the terms are null, except for the one in which the differential dx^j that is canceled is exactly dx^n! In conclusion, what we get is

$$\omega(\mathbf{x}, x^n)|_{\partial\mathbb{H}^n} = g_n(\mathbf{x}, 0) dx^1 \wedge \dots \wedge dx^{n-1},$$

and its integral is

$$\int_{\partial\mathbb{H}^n} \omega = (-1)^n \int_{\mathbb{R}^{n-1}} g_n(\mathbf{x}, 0) dx^1 \dots dx^{n-1}, \tag{6.6}$$

where the factor $(-1)^n$ must be introduced because of the convention that we have established in definition 6.6.1, in other terms: in the identification $\partial\mathbb{H}^n \cong \mathbb{R}^{n-1}$ we take the same orientation of $\mathbb{R}^{n-1}$ if n is even and the opposite one if n is odd.

Now let us compute the integral over $\mathbb{H}^n$ of $d\omega$. First of all we write $d\omega$ explicitly: using the same argument as before regarding the alternating property of the wedge product, we get

$$\begin{aligned} d\omega &= \sum_{j=1}^{n} \left(\sum_{h=1}^{n} \frac{\partial g_j}{\partial x^h} dx^h \right) dx^1 \wedge \dots \cancel{dx^j} \dots \wedge dx^{n-1} \\ &= \sum_{j=1}^{n} \frac{\partial g_j}{\partial x^j} dx^j \wedge dx^1 \wedge \dots \wedge dx^{n-1} \\ &= \left(\sum_{j=1}^{n} (-1)^{j-1} \frac{\partial g_j}{\partial x^j} \right) \wedge dx^1 \wedge \dots dx^j \dots \wedge dx^{n-1}, \end{aligned}$$

where, again, the factor $(-1)^{j-1}$ comes from the order flips that must be performed in order to put dx^j in its ordinal position.

Now we must compute $\int_{\mathbb{H}^n} d\omega$. For $1 \leqslant j \leqslant n-1$, the coordinates vary from $-\infty$ to $+\infty$, while for $j = n$ the variability is from 0 to $+\infty$. Let us first concentrate ourselves on the integral w.r.t. the variables with $1 \leqslant j \leqslant n-1$:

$$\int_{-\infty}^{+\infty} \frac{\partial g_j}{\partial x^j}(x^1, \dots, x^n) dx^j = g_j(x^1, \dots, t, \dots, x^n)\Big|_{t\to-\infty}^{t\to+\infty} = 0, \tag{6.7}$$

where of course t is placed at the j-th position and the integral is null because g_j has compact support on $\mathbb{H}^n$.

Using the Fubini theorem we find, for $j = 1, \dots, n-1$:

$$\begin{aligned}\int_{\mathbb{H}^n} \frac{\partial g_j}{\partial x^j}(x^1, \dots, x^n) dx^1 \dots dx^n &= \int_0^{+\infty} \left[\int_{\mathbb{R}^{n-1}} \frac{\partial g_j}{\partial x^j} dx^1 \dots dx^{n-1}\right] dx^n \\ &= \int_0^{+\infty} \left[\int_{\mathbb{R}^{n-2}} \left(\int_{-\infty}^{+\infty} \frac{\partial g_j}{\partial x^j} dx^j\right) dx^1 \dots \cancel{dx^j} \dots dx^{n-1}\right] dx^n,\end{aligned}$$

but thanks to eq. (6.7) the integral between parenthesis in null, and so the whole integral is null:

$$\int_{\mathbb{H}^n} \frac{\partial g_j}{\partial x^j}(x^1, \dots, x^n) dx^1 \dots dx^n, \quad j = 1 \dots, n-1.$$

Instead, when $j = n$, we have

$$\int_0^{+\infty} \frac{\partial g_n}{\partial x^n}(x^1, \dots, x^n) dx^n = g_n(x^1, \dots, x^{n-1}, t)\Big|_{t=0}^{t\to+\infty} = 0 - g_n(x^1, \dots, x^{n-1}, 0) \tag{6.8}$$

again because g_n has compact support in $\mathbb{H}^n$.

Now we can conclude the computation of $\int_{\mathbb{H}^n} d\omega$ quite easily:

$$\begin{aligned}\int_{\mathbb{H}^n} d\omega &= \sum_{j=1}^n (-1)^{j-1} \int_{\mathbb{H}^n} \frac{\partial g_j}{\partial x^j} dx^1 \dots dx^n \\ &\underset{(6.7)}{=} (-1)^{n-1} \int_{\mathbb{H}^n} \frac{\partial g_n}{\partial x^n} dx^1 \dots dx^n \\ &= (-1)^{n-1} \int_{\mathbb{R}^{n-1}} \left(\int_0^{+\infty} \frac{\partial g_n}{\partial x^n} dx^n\right) dx^1 \dots dx^{n-1}.\end{aligned}$$

where, in the last step, we have decomposed the integral, thanks to the Fubini theorem, in the integral over dx^n, which varies between 0 and $+\infty$ and the integral over the remaining variables, which must be extended over $\mathbb{R}^{n-1}$.

Finally, we use eq. (6.8) to write

$$\int_{\mathbb{H}^n} d\omega = (-1)^{n-1} \int_{\mathbb{R}^{n-1}} -g_n(\mathbf{x}, 0) dx^1 \dots dx^{n-1} = (-1)^n \int_{\mathbb{R}^{n-1}} g_n(\mathbf{x}, 0) dx^1 \dots dx^{n-1},$$

but thanks to eq. (6.6) this integral is exactly $\int_{\partial\mathbb{H}^n} \omega$ and so the Stokes theorem is proven also for $M = \mathbb{H}^n$.

Having proved the Stokes theorem for the local models of a manifold with and without boundary, we can now pass to the general case where M is a generic oriented manifold with boundary.

Let (U, φ) a chart of M, then $\varphi(U)$ is either an open set of $\mathbb{R}^n$, in which case we are dealing with a n-inner chart, or an open set of of $\mathbb{H}^n$, in which case we are dealing with a n-border chart.

Taking U connected, it is not restrictive to consider the cases where $\varphi(U) = \mathbb{R}^n$ or $\varphi(U) = \mathbb{H}^n$, respectively. Hence, we can choose an oriented atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}_\alpha$ of M in which all the charts satisfy the previous hypothesis, i.e. for all α, $\varphi_\alpha(U_\alpha) = \mathbb{R}^n$ or $\varphi_\alpha(U_\alpha) = \mathbb{H}^n$, respectively.

We now proceed in the canonical way to compute the integrals, i.e. we first consider a partition of the unity $\{\rho_\alpha\}_\alpha$ subordinated to the covering $\{U_\alpha\}_\alpha$ and, since $\sum_\alpha \rho_\alpha = \mathbf{1}$, we can write any compactly supported $(n-1)$-form on M, ω, as $\mathbf{1}\omega = \sum_\alpha \rho_\alpha \omega$.

Now we use the fact that both d and $\int$ are additive operators, which allows us to prove the Stokes theorem just by showing that it holds for all the terms $\rho_\alpha\omega$, which is a $(n-1)$-form whose compact support is contained in the open U_α, since $\mathrm{supp}(\rho_\alpha) \subset U_\alpha$.

Now we can pass to the actual computations:

$$\int_M d(\rho_\alpha\omega) = \int_{U_\alpha} d(\rho_\alpha\omega) := \int_{\varphi_\alpha(U_\alpha)} (\varphi_\alpha^{-1})^* d(\rho_\alpha\omega),$$

but since the pull-back and d commute we can write

$$\int_M d(\rho_\alpha\omega) = \int_{\varphi_\alpha(U_\alpha)} d((\varphi_\alpha^{-1})^*(\rho_\alpha\omega)),$$

however we are under the hypothesis that $\varphi_\alpha(U_\alpha)$ is either $\mathbb{R}^n$ or $\mathbb{H}^n$, in which cases we have proven that the Stokes theorem holds true, so we can continue the previous formula as follows:

$$\begin{aligned}\int_M d(\rho_\alpha\omega) &= \int_{\varphi_\alpha(U_\alpha)} d((\varphi_\alpha^{-1})^*(\rho_\alpha\omega)) \\ &\underset{\text{Stokes}}{=} \int_{\partial\varphi_\alpha(U_\alpha)} (\varphi_\alpha^{-1})^*(\rho_\alpha\omega) \\ &\underset{(1.6.1)}{=} \int_{\varphi_\alpha(U_\alpha\cap\partial M)} (\varphi_\alpha^{-1})^*(\rho_\alpha\omega) \\ &=: \int_{U_\alpha\cap\partial M} \rho_\alpha\omega,\end{aligned}$$

but since $\mathrm{supp}(\rho_\alpha) \subset U_\alpha$ the last integral can be extended just to ∂M, thus giving

$$\int_M d(\rho_\alpha\omega) = \int_{\partial M} \rho_\alpha\omega,$$

which is exactly the Stokes theorem for the terms $\rho_\alpha\omega$, which, by additivity, implies the validity of the theorem for ω. □

The Stokes theorem has many important consequences that can be easily proven thanks to it. Two of them are collected below.

Corollary 6.6.1 *Let M be a n-manifold without boundary.*

1. *If $\omega \in \mathbb{A}^n(M)$ is a compactly supported exact form, i.e. it exists $\eta \in \mathbb{A}^{n-1}(M)$ such that $\omega = d\eta$, then*
$$\int_M \omega = 0.$$
In particular, if M is compact and $\nu \in \mathbb{A}^n(M)$ is a volume form, then ν is not exact and so the n-th de Rham cohomology group $H^n_{dR}(M) \neq \{0\}$.

2. *More generally, if $S \subset M$ is a compact sub-manifold of M of dimension k without boundary, and $\omega \in \mathbb{A}^k(M)$ is a closed k-form, i.e. $d\omega = 0$ such that $\int_S \omega \neq 0$, then ω is not exact and S is not the border of a sub-manifold of dimension $k+1$ of M.*

Proof.

1. Applying the Stokes theorem we find:
$$\int_M \omega = \int_M d\eta = \int_{\partial M} \eta = 0$$
because $\partial M = \varnothing$.

Regarding the volume form: since ν is nowhere null, its integral cannot be 0, so it cannot be exact. However, $d\nu$ is a $(n+1)$-form on a n-dimensional manifold, so it is null, this shows that ν is closed but not exact, hence $H^n_{\mathrm{dR}}(M) \neq \{0\}$.

2. Being ω exact, we can write as before $\omega = d\eta$ and $\omega|_S = d(\eta|_S)$ because the d and the pull-back commute. As a consequence, using again the Stokes theorem, we have
$$\int_S \omega = \int_S d\eta = \int_{\partial S} \eta = 0$$
because $\partial S = \varnothing$. Since this contradicts the assumption that $\int_S \omega \neq 0$, ω cannot be exact.

Finally, if $S = \partial\tilde{S}$, where $\tilde{S}$ is a sub-manifold of dimension $k+1$ of M, then, again by the Stokes theorem, we would have
$$\int_S \omega = \int_{\partial\tilde{S}} \omega = \int_{\tilde{S}} \omega = 0$$
because ω is closed, so $d\omega = 0$. □

6.6.5 The Gauss-Green theorem in $\mathbb{R}^2$ as a corollary of Stokes theorem

Let $M = D \subset \mathbb{R}^n 2$ be a regular bounded domain. The border of D is a compact sub-manifold of dimension 1 of $\mathbb{R}^2$, hence it is diffeomorphic to S^1. Hence, it exists a smooth, simple, regular, closed curve $\gamma : [0,1] \to \mathbb{R}^2$ which parameterizes ∂D, i.e. $\gamma([0,1]) = \partial D$. It is also possible to suppose without loss of generality that γ preserves the orientation, i.e. that the natural orientation from 0 to 1 defines on ∂D the same orientation as the one induced by D.

If we consider now a 1-form ω on D with coefficients given by two functions $f, g \in \mathscr{C}^\infty(D)$, i.e.
$$\omega = f dx + g dy \in \mathbb{A}^1(D),$$
then we get
$$d\omega = \frac{\partial f}{\partial x}\cancel{dx \wedge dx} + \frac{\partial f}{\partial y} dy \wedge dx + \frac{\partial g}{\partial x} dx \wedge dy + \frac{\partial g}{\partial y}\cancel{dy \wedge dy},$$

since $dy \wedge dx = -dx \wedge dy$ we obtain

$$d\omega = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx \wedge dy.$$

By the Stokes theorem we find that

$$\int_D d\omega = \int_D \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dxdy = \int_{\partial D} \omega = \int_{\partial D} (fdx + gdy) = \int_\gamma (fdx + gdy),$$

which is the classical Gauss-Green theorem.

6.7 Orientation of hypersurfaces

Let us consider a manifold M with or without boundary and a sub-manifold $S \subset M$ (immersed or embedded). Let us recall that a tangent vector field to M along S is a section of the fiber bundle $TM|_S$, the restriction of TM to S, i.e. a function $N : S \to TM$, such that $N_p \in T_pM$[2], for all $p \in S$.

Every tangent vector field $X \in \mathfrak{X}(M)$ can be restricted to a tangent vector field along S, but the reverse, in general, is not true, i.e. not all tangent vector field along S can be obtained by restriction of a tangent vector field $X \in \mathfrak{X}(M)$.

Let us now see what happens when we consider an orientation on M and an immersed hypersurface $S \subset M$ with, as usual, $\iota : S \hookrightarrow M$ representing the natural inclusion of S into M.

We will need the following definition.

Def. 6.7.1 (Internal product or contraction between a k-form and a vector field) *Given a manifold M of dimension N and $k < n$, if $\omega \in \mathbb{A}^k(M)$ and $X \in \mathfrak{X}(M)$, then the internal product, or contraction, between ω and X, written as $X \lrcorner \omega$, is the $(k-1)$-form defined as follows:*

$$(X \lrcorner \omega)(X_1, \ldots, X_{k-1}) := \omega(X, X_1, \ldots, X_{k-1}),$$

for all $X_1, \ldots, X_{k-1} \in \mathfrak{X}(M)$.

Theorem 6.7.1 *Let M be a n-dimensional oriented manifold with or without boundary and an immersed hypersurface $S \subset M$, so $\dim(S) = n - 1$, with or without boundary.*

Let N be a tangent vector field along S which, however, has the property of not being tangent to S in any point, i.e.

$$\forall p \in S, \; N_p \in T_pM \backslash T_pS.$$

Then, S has a unique orientation such that, for all $p \in S$, an ordered set of vectors $(E_1, \ldots, E_{n-1}) \in T_pS$ is an ordered basis of T_pS if and only if $(N_pE_1, \ldots, E_{n-1})$ is an oriented basis of T_pM.

This result can be restated by saying that, if ω is a volume form for M, then $\iota_S^*(N \lrcorner \omega)$ is a volume form for this orientation.

Proof. If $\omega \in \mathbb{A}^n(M)$ is a volume form for M, then $\sigma := \iota_S^*(N \lrcorner \omega)$ is a $(n-1)$-form on S. If we manage to prove that $\sigma_p \neq 0$ for all $p \in S$, then this will automatically prove that σ provides

[2]notice that we do not demand N_p to be tangent to S, it may happen, but this is not necessarily true.

an orientation on S thanks to the equivalence between volume forms and orientation forms proven before in this chapter.

Let now use the hypotheses that, for all $p \in S$, $(E_1, \ldots, E_{n-1}) \in T_pS$ is an ordered basis of T_pS and that $N_p \in T_pM \backslash T_pS$, hence $(N_p, E_1, \ldots, E_{n-1})$ is a basis for T_pM because N_p is surely linearly independent from $E_1, \ldots, E_{n-1}$.

Let us now compute

$$\sigma_p(E_1, \ldots, E_{n-1}) = (N \lrcorner \omega)_p((E_1, \ldots, E_{n-1})) = \omega_p(N, E_1, \ldots, E_{n-1}) \neq 0 \quad \forall p \in S,$$

where in the second step we have used the fact that the differential of the canonical inclusion is the identity and in the last step we have take advantage of the fact that ω is a volume form.

This verifies that also $\sigma_p \neq 0$ for all $p \in S$, hence σ is an orientation form too.

It remains to understand what kind of orientation is induced on S. For that we notice that

$$\sigma_p(E_1, \ldots, E_{n-1}) > 0 \iff \omega_p(N_p, E_1, \ldots, E_{n-1}) > 0,$$

so $(E_1, \ldots, E_{n-1})$ is a positively-oriented basis of S if and only if $(N_p, E_1, \ldots, E_{n-1})$ is a positively-oriented basis for M, which is what we had to check. $\square$

The important consequence of this theorem is that, **in order to endow any hypersurface S of an oriented manifold M with an orientation, it is sufficient to fix a vector field N which is tangent to M but not to S!**

Let us apply the previous theorem to the simple, but very useful, case of spheres. We know that S^n is a hypersurface of $\mathbb{R}^{n+1}$, which has a standard orientation given by the canonical basis.

The vector field on $\mathbb{R}^{n+1}$ given by

$$N := x^i \frac{\partial}{\partial x^i}$$

written in components is nothing but the vector fields which associated to the local components $(x^1, \ldots, x^{n+1})$ considered as a point, the same components $(x^1, \ldots, x^{n+1})$, but this time considered as a vector. So N is a radial field orthogonal (so not tangent) to S^n called the field of **outer normal vectors** of S^n, as depicted in Figure 6.1.

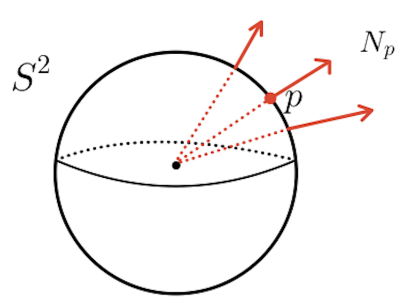


Figure 6.1: Representation of the radial outer field for the sphere S^2.

Thanks to what we have just proven above, the vector field N induces and orientation on S^n defined very simply as follows: $(E_1, \ldots, E_n)$ is a positively-oriented basis for T_pS^n if and only if $(N_p, E_1, \dots, E_n)$ is a positively-oriented basis for $\mathbb{R}^{n+1}$. The orientation of S^n defined by the outer normal vectors is called the **standard orientation of S^n**.

The opposite orientation that can be given to the sphere is clearly provided by the inner normal vectors pointing toward its center.

Corollary 6.7.1 *Let M be a n-dimensional oriented manifold and $S \subset M$ a regular level set of a smooth function $f \in \mathscr{C}^\infty(M)$. Then, S is oriented.*

Proof. The level set theorem assures that S is a sub-manifold of M of dimension $n-1$.

We have the possibility to endow M with a Riemannian metric, let us call it g and define the tangent vector field N that will provide an orientation of S as follows:

$$N := \operatorname{grad} f|_S := \sharp(df)|_S\,.$$

Since the gradient of f is orthogonal, so not tangent, w.r.t. the metric g, to the tangent spaces to the level set S, the previous theorem implies that S is oriented. □

6.7.1 An alternative way to provide an orientation on the border of a manifold

We have seen that, given a n-dimensional oriented manifold with boundary M, we can endow ∂M with an orientation by restricting the ordered atlas of M to the border ones and consider the induced orientation on ∂M if n is even and the opposite one if n is odd.

However, in light of the results proven in this section, there is an alternative way of equipping ∂M with an orientation. In fact, we know that the border ∂M is an embedded hypersurface of M, so we can orient it by fixing a vector field N that is never tangent to ∂M.

Regarding this alternative procedure, we have some standard definitions.

Def. 6.7.2 *Let $p \in \partial M$ and $v \in T_pM$. We say that*

- *v is internal if $v \notin T_p(\partial M)$ and if it exists a smooth curve $\gamma : [0, \varepsilon] \to M$, $\varepsilon > 0$, such that $\gamma(0) = p$ and $\dot\gamma(0) = v$;*
- *v external $-v$ is internal.*

A graphical representation of internal and external vector is depicted in Figure 6.2.

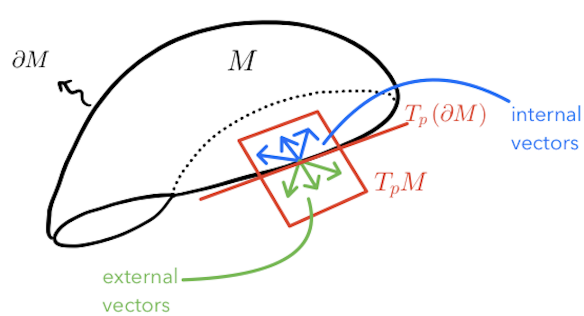


Figure 6.2: Depiction of internal and external vectors for a 2-dimensional surface.

It is possible to prove that it always exists an external vector field N along ∂M, i.e. a vector field $N\partial M \to TM$ such that $N_p \in T_pM$ is an external vector for all $p \in \partial M$, see e.g. [1]. Since, by construction, N is never tangent to ∂M, N determines an orientation of ∂M. We call N an **external vector field**.

Of course we can ask ourselves if the orientation of ∂M depends on the chosen external vector field, the following result guarantees that this is not the case.

Theorem 6.7.2 *Let M be a n-dimensional oriented manifold with boundary. Then ∂M is oriented and all the external vector fields along ∂M determine the same orientation on it.*

Proof. Let $p \in \partial M$ and let $(x^1, \dots, x^n)$ its local coordinates defined by a border chart (U, φ) in p. This means that, in U the border ∂M is described by the hypersurface equation $x^n = 0$, while the points of U internal to M are characterized by the equation $x^n > 0$ because in this way they belong to $\text{int}(\mathbb{H}^n)$.

If now $\tilde{N}$ and N are two external vector field along ∂M, then their last components $\tilde{N}_p^n$ and N_p^n, respectively, are both negative, otherwise they will point toward $\text{int}(\mathbb{H}^n)$ and they will correspond to internal vector fields! As a consequence, if we consider the two bases of T_pM given by

$$(\tilde{N}_p, \partial_1|_p, \dots, \partial_n|_p), \quad (N_p, \partial_1|_p, \dots, \partial_n|_p),$$

by direct computation, it can be verified that the change of basis matrix has determinant $\tilde{N}_p^n / N_p^n > 0$ being a ratio between two strictly negative numbers. Hence, the two bases are equioriented and so they induce the same orientation of $T_p(\partial M)$. □

Now, the only fact that remains to be proven is that the orientation induced on the border by an external vector field is the same as the one defined in 6.6.1. It can be proven that this is indeed the case, see e.g. [1, 10].

We are going to verify this fact in the case $M = \mathbb{H}^n$, as depicted in Figure 6.3, the general proof is a localization of the next computation with the help of a border chart.

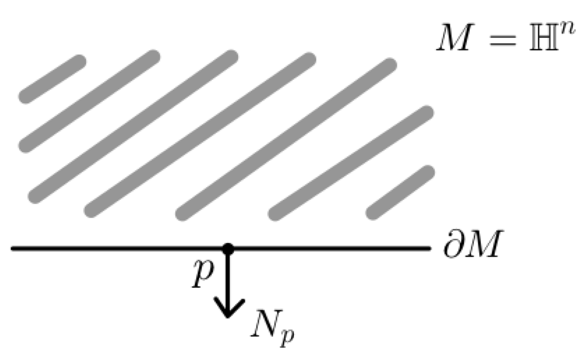


Figure 6.3:

Since the orientation of ∂M does not depend on the particular external vector field along ∂M chosen, we can take the easiest one, i.e.

$$N = -\frac{\partial}{\partial x^n} \cong (0, 0, \dots, -1) = -e_n \in \mathbb{R}^n.$$

If we identify $T_p(\partial M)$ with the subspace of $\mathbb{R}^n$ generated by the vectors $e_1, \dots, e_{n-1}$ of the canonical basis of $\mathbb{R}^n$, then, by definition $(e_1, \dots, e_{n-1})$ is a positively-oriented basis induced by the external vector field N if the basis of T_pM given by $(N, e_1, \dots, e_{n-1})$ is positively-oriented for the canonical orientation of $T_pM \cong \mathbb{R}^n$.

Let us denote with $[N, e_1, \dots, e_{n-1}]$ the equivalence class defined by the basis $(N, e_1, \dots, e_{n-1})$, then we have:

$$[N, e_1, \dots, e_{n-1}] = [-e_n, e_1, \dots, e_{n-1}] = -[e_n, e_1, \dots, e_{n-1}],$$

now notice that each time we exchange two basis vectors, the orientation changes sign, hence

$$[N, e_1, \dots, e_{n-1}] = -(-1)^{n-1}[e_1, \dots, e_{n-1}, e_n] = (-1)^n[e_1, \dots, e_{n-1}, e_n],$$

which means that the orientation on ∂M determined by the external vector field N coincides with the standard orientation of $\mathbb{R}^n$ defined by the canonical basis (with the correct order of the basis vectors) if n is even, while is the opposite one if n is odd.

But this is exactly the definition of orientation given in 6.6.1.

Chapter 7

Riemannian and pseudo-Riemannian manifolds

We start with the introduction of the fundamental concept of Riemannian and pseudo-Riemannian metric.

7.1 Riemannian and pseudo-Riemannian metrics

A scalar product on a vector space allows us to measure the length of vectors and the angles between them. In differential geometry, the typical vector spaces that we have to deal with are the tangent spaces to each point p of a manifold M. If we assign a scalar product to each tangent space T_pM, i.e.

$$\begin{array}{rccc} g_p : & T_pM \times T_pM & \longrightarrow & \mathbb{R} \\ & (v,w) & \longmapsto & g_p(v,w) \end{array}$$

smoothly w.r.t. changes of $p \in M$, then we fix a so-called Riemannian metric on M.

Since the (real-valued) scalar product is bilinear, symmetric, i.e. $g_p(v,w) = g_p(w,v)$ for all $v,w \in T_pM$ and positive-definite, i.e. $g_p(v,v) \geqslant 0$ for all $v \in T_pM$, with $g_p(v,v) = 0$ if and only if $v = 0$, a Riemannian metric on M is nothing but a positive-definite symmetric tensor field on TM of type $(0,2)$, i.e. 2-covariant, as formalized by the definition below.

Def. 7.1.1 (Riemannian metric and manifold) *A Riemannian metric on a manifold M is a positive-definite tensor field $g \in S_2^0(M)$. A Riemannian manifold is a couple (M,g), where g is a Riemannian metric on M.*

The norm canonically induced by the scalar product g_p on T_pM will be denoted with $\| \ \|_p$:

$$\|v\|_p^2 = g_p(v,v) \qquad \forall v \in T_pM.$$

More generally, as it is required in relativistic theories, we can reduce the requests on g by dropping off the property of being positive, but keeping the fundamental property of **non-degeneracy**, i.e. $g_p(v,w) = 0 \ \forall w \in T_pM$ implies $v = 0$, i.e. the only vector g_p-orthogonal to all the other vectors of T_pM is the 0 vector of T_pM, in this case we get a pseudo-Riemannian metric.

Def. 7.1.2 (pseudo-Riemannian metric and manifold) *A pseudo-Riemannian metric on a manifold M is a non-degenerate tensor field $g \in S^0_2(M)$. A pseudo-Riemannian manifold is a couple (M, g), where g is a pseudo-Riemannian metric on M.*

An important concept related with pseudo-Riemannian metrics is their signature.

Def. 7.1.3 (Signature) *Given a pseudo-Riemannian metric g on a* ***connected*** *manifold M of dimension n, we say that g has signature (r, s), $r + s = n$, if the maximal dimension of a subspace of T_pM where g is:*

- *positive-definite is r;*
- *negative-definite is s.*

The definition is well-posed for connected manifolds because, by an argument of continuity, it can be proven that r and s do not depend on the point $p \in M$.

A particularly important situation arising in relativistic theories is that of the signature $(1, n-1)$ or $(n-1, 1)$. In this case one says that g is a **Lorentz metric**, or that g has a **Lorentz signature**.

Since the metric is a symmetric tensor field of type $(0, 2)$, in the local coordinate system $(U, \varphi = (x^1, \dots, x^n))$ it can be written as follows:

$$g(p) = g_{\mu\nu}(p) dx^\mu \otimes dx^\nu,$$

where $g_{\mu\nu} \in \mathscr{C}^\infty(U)$. Thanks to theorem 4.1.1, the coefficients $g_{\mu\nu}$ are explicitly given by

$$g_{\mu\nu}(p) = g_p(\partial_i|_p, \partial_j|_p)$$

because $(dx^i|_p)_i$ is the dual basis of T_p^*M associated to the coordinate basis $(\partial_i|_p)_i$ of T_pM.

It is custom to drop p and write simply:

$$\boxed{g = g_{\mu\nu} dx^\mu \otimes dx^\nu},$$

considering implicit the dependence on p of the matrix of scalar functions $(g_{\mu\nu})$, thus treating is as a:

- real symmetric and positive-definite matrix for a Riemannian metric;
- real symmetric matrix of signature (r, s) for a pseudo-Riemannian metric.

Since real symmetric matrices can be diagonalized, for every $p \in U$, $g_{\mu\nu}$ can always put in the diagonal form $g_{\mu\nu} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, where λ_i is the i-th eigenvalue of $g_{\mu\nu}$.

When $M = \mathbb{R}^n$, the tangent and cotangent bundle are canonically isomorphic and the Euclidean metric induced by the Euclidean scalar product is such that $(g_{\mu\nu}) = I_n$, i.e. the identity matrix of dimension n for all $p \in \mathbb{R}^n$.

Remark about the notation: by symmetry, we could write $g = g_{\mu\nu} dx^\mu \odot dx^\nu$, where $\odot$ is the symmetric product, or, as it is typically done by physicists, $g = g_{\mu\nu} dx^\mu dx^\nu$, which is justified by the fact that the product is symmetric.

Many authors adopt the so-called **Gauss' notation** by writing ds^2 instead of g, so that we usually find the following notation for the metric:

$$\boxed{ds^2 = g_{\mu\nu} dx^\mu dx^\nu}.$$

By definition, the matrix $g_{\mu\nu}$ is invertible, since, in the Riemannian case, it is positive-definite and, in the pseudo-Riemannian case, it is non-degenerated.[1] The inverse is usually denoted as follows:

$${g_{\mu\nu}}^{-1} = g^{\mu\nu},$$

so that:

$$g^{\mu\gamma} g_{\gamma\nu} = \delta^\mu{}_\nu, \quad g_{\mu\gamma} g^{\gamma\nu} = \delta_\mu{}^\nu.$$

Apart from permitting the computation of the scalar product between tangent vectors, *a (pseudo)-Riemannian metric also allows us to canonically identify the tangent and the cotangent bundle* with the help of the following linear isomorphism:

$$\begin{array}{rccc} \flat_p : & T_pM & \xrightarrow{\sim} & T_p^*M \\ & v & \longmapsto & \flat_p(v), \end{array} \qquad \begin{array}{rccc} \flat_p(v) : & T_pM & \longrightarrow & \mathbb{R} \\ & w & \longmapsto & \flat_p(v)(w) := g_p(v,w). \end{array}$$

Being TM the disjoint union of T_pM when we vary the point $p \in M$, we can define a linear isomorphism $\flat$ of bundles $\flat : TM \xrightarrow{\sim} T^*M$ simply by requiring that $\flat|_{T_pM} = \flat_p$ for all $p \in M$.

Let us search for a local expression of $\flat$: let $g = g_{\mu\nu} dx^\mu \otimes dx^\nu$ and let $X = X^h \partial_h$ be a local section of TM, i.e. a local vector field, then the application of $\flat$ to X must give a local section of T^*M, i.e. a local covector field, or a 1-form on M that will be written as $\flat(X) = \alpha_j dx^j$ in local coordinates.

By definition of $\flat$, we have:

$$\flat(X)(\partial_k) = g(X, \partial_k) = (g_{\mu\nu} dx^\mu \otimes dx^\nu)(X^h \partial_h, \partial_k),$$

but $dx^\mu \otimes dx^\nu$ is a symmetric bilinear form, so we can move the coefficients X^h outside and write:

$$\flat(X)(\partial_k) = g_{\mu\nu} X^h (dx^\mu \otimes dx^\nu)(\partial_h, \partial_k).$$

Now, by definition of tensor product of two linear forms (cfr. (4.1)), we have that $(dx^\mu \otimes dx^\nu)(\partial_h, \partial_k) = dx^\mu(\partial_h) dx^\nu(\partial_k) = \delta^\mu{}_h \delta^\nu{}_k$, so:

$$\flat(X)(\partial_k) = g_{\mu\nu} X^h \delta^\mu{}_h \delta^\nu{}_k = g_{hk} X^h.$$

However, we also have:

$$\flat(X)(\partial_k) = \alpha_j dx^j(\partial_k) = \alpha_j \delta^j{}_k = \alpha_k,$$

thus $\alpha_k = g_{hk} X^h$, so, finally:

$$\boxed{\flat(X^h \partial_h) = g_{hk} X^h dx^k}.$$

Since the basis are the (fixed) standard basis of the tangent ant the cotangent bundle, it is custom to omit them and write simply the components, i.e.

$$\flat(X^h) = g_{hk} X^h = g_{kh} X^h,$$

[1] To avoid specifying if we are discussing a Riemannian or pseudo-Riemannian metric, we will simply write (pseudo-)Riemannian metric by meaning that we can refer to both cases.

by symmetry.

In conclusion, we can write:

$$\begin{array}{rccl} \flat: & TM & \xrightarrow{\sim} & T^*M \\ & (X^h) & \longmapsto & \flat(X^h) = (\alpha_k) = g_{kh}X^h. \end{array}$$

This formula explains why it is custom to say that $\flat$ is the isomorphism which transforms the components (X^h) of a local vector field to the components (α_k) of a local 1-form by '**lowering the indices with the metric tensor**'. The symbol $\flat$ ('flat' or 'bemolle') is chosen because in music it lowers in pitch by one semitone.

Analogously, the inverse isomorphism $\flat^{-1} \equiv \sharp : T^*M \xrightarrow{\sim} TM$ act like this:

$$\begin{array}{rccl} \sharp: & T^*M & \xrightarrow{\sim} & TM \\ & (\alpha_k) & \longmapsto & \sharp(\alpha_k) = (X^h) = g^{hk}\alpha_k. \end{array}$$

It is custom to say that $\sharp$ is the isomorphism which transforms the components (α_k) of a local 1-form to the components (X^h) of a vector field by '**raising the indices by using the inverse metric tensor**', since $(X_h = g^{hk}\alpha_k)$. Again, the symbol $\sharp$ ('sharp' or 'diesis') is because in music it highers in pitch by one semitone.

Summarizing, in the presence of a (pseudo-)Riemannian metric we can transform vector fields to 1-forms and vice-versa, simply by applying the metric tensor and its inverse, respectively.

7.1.1 Noticeable example 1: the gradient of a scalar function

Let us apply the isomorphism $\sharp : T^*M \to TM$ to the differential of a smooth scalar function $\phi \in \mathscr{C}^\infty(M)$. We know that $d\phi$ is a section of T^*M, i.e. a 1-form, so, if we apply $\sharp$ to $d\phi$, we obtain a vector field, which turns out to be the generalization of the gradient to manifolds.

Def. 7.1.4 (Gradient of a scalar function) *Given a scalar function $\phi \in \mathscr{C}^\infty(M)$, its gradient* $\mathrm{grad}(\phi) \in \mathfrak{X}(M)$ *is the vector field defined by:*

$$\mathrm{grad}(\phi) := \sharp(d\phi).$$

In local coordinates, if $d\phi = (\partial_j\phi)\, dx^j$, then the action of $\sharp$ on the components is as follows: $\sharp(\partial_j\phi) = g^{ij}\,(\partial_j\phi)$, so that

$$\mathrm{grad}(\phi) = g^{ij}\,(\partial_j\phi)\,(\partial_i x^i),$$

coherently with the fact that $\mathrm{grad}(\phi)$ is a tangent vector, so it must be a linear combination of the ∂_i's. This shows that, for generic manifolds, **the presence of a (pseudo-)Riemannian metric is fundamental in order to define the gradient of a scalar function**.

This fact is hidden for the trivial case of $M = \mathbb{R}^n$ because, as already remarked, in that situation $g^{ij} = g_{ij} = I_n$ and so $\mathrm{grad}(\phi) = (\frac{\partial\phi}{\partial x^1}, \ldots, \frac{\partial\phi}{\partial x^n})$.

7.1.2 Noticeable example 2: symplectic manifolds, the Hamiltonian isomorphism and the Poisson bracket

Riemannian, or pseudo-Riemannian, metrics and manifolds are built via symmetric positive-definite, or non-degenerated, tensor fields of type $(0, 2)$. Another remarkable construction can be obtained by considering anti-symmetric non-degenerated tensor fields of type $(0, 2)$.

Def. 7.1.5 (Simplectic form and manifold) *A closed non-degenerated 2-form $\omega = \omega_{ij} dx^i dx^j$ is called a simplectic form on M and a couple (M, ω) is said to be a simplectic manifold.*

Also for simplectic manifold we can identify the tangent and the cotangent bundles with the analogous of the isomorphism $\sharp$ that, in this setting, is called the **Hamiltonian isomorphism**:

$$\begin{array}{rccl} H: & T^*M & \xrightarrow{\sim} & TM \\ & (\alpha_j) & \longmapsto & H(\alpha_j) = X^i, \ X^i = \omega^{ij}\alpha_j. \end{array}$$

We can repeat the same construction as before with $\sharp$ to obtain a vector field from the differential of a scalar function $\phi \in \mathscr{C}(M)$ but, this time, by using H instead of $\sharp$. What we obtain is $H(d\phi) \in \mathfrak{X}(M)$, which is called **Hamiltonian vector field** of the scalar function ϕ.

Since a 2-form ω takes as input two vector fields, it is interesting to see what happens if we consider the differential of two scalar functions $\phi, \psi \in \mathscr{C}^\infty(M)$, the Hamiltonian vector fields associated to them, i.e. $H(d\phi), H(d\psi)$, and then we apply ω. The result is the so-called **Poisson bracket**:

$$\boxed{\{\phi, \psi\} := \omega(H(d\phi), H(d\psi))}\,.$$

$\mathscr{C}^\infty(M)$ becomes a Lie algebra w.r.t. the Poisson bracket (just as the set of tangent vector fields on M becomes a Lie algebra w.r.t. the Lie bracket).

7.2 Existence of Riemannian metrics

We can now prove the existence of Riemannian metrics.

Theorem 7.2.1 *Every smooth manifold M admits a Riemannian metric.*

Proof. The idea is quite simple: we start with a local Riemannian metric and then we extend it to the whole manifold thanks to a partition of the unity. Let us discuss the technical details.

Consider an atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ of M and a partition of the unity $\{\rho_\alpha\}$ subordinated to the covering $\{U_\alpha\}$ (so that each ρ_α is identically 0 outside U_α).

On U_α it is very easy to induce a metric from the Euclidean metric of $\mathbb{R}^n$. To see how, consider a chart function $\varphi_\alpha : U_\alpha \xrightarrow{\sim} \varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^n$, if $\varphi_\alpha \equiv (x^1, \dots, x^n)$, then we know that the vector fields $(\partial_1, \dots, \partial_n)$ provide a local frame for $TM|_{U_\alpha}$. Given $p \in U_\alpha$ and two tangent vectors $X_p, Y_p \in T_pM$, $X_p = X^i \partial_i\big|_p$, $Y_p = Y^j \partial_j\big|_p$, we define a scalar product between $\partial_i\big|_p$ and $\partial_j\big|_p$ by means of the Euclidean product $\langle\ ,\ \rangle$ of $\mathbb{R}^n$ as follows:

$$g_p^\alpha(\partial_i|_p\,, \partial_j|_p) := \langle \varphi_\alpha(\partial_i|_p), \varphi_\alpha(\partial_j|_p) \rangle = \langle e_i, e_j \rangle = \delta_{ij},$$

recalling that $\partial_i|_p = d\varphi_\alpha^{-1}\big|_p (e_i)$, e_i being the i-th element of the canonical basis of $\mathbb{R}^n$. The extension to any couple X_p, Y_p of tangent vectors in T_pM is performed by linearity:

$$g_p^\alpha(X|_p, Y_p) := X^i Y^j g_p^\alpha(\partial_i|_p\,, \partial_j|_p) = X^i Y^j \delta_{ij} = \sum_{i=1}^n X^i Y^i.$$

g_p^α is then a positive-definite bilinear form for all $p \in M$ and for all α. Now we glue together these scalar products to build a tensor field $g \in T_2^0(M)$ by defining:

$$g_p := \sum_\alpha \rho_\alpha(p) g_p^\alpha, \qquad \forall p \in M.$$

The definition is well-posed because the sum is actually finite since, for all $p \in M$, there is only a finite number of $\rho_\alpha(p)$ different from 0. Plus, $\rho_\alpha(p) \geqslant 0$ for all $p \in M$ and all α, thus the coefficients $\rho_\alpha(p)$ do not modify the positive-definiteness of the forms g_p^α and then g results in a positive-definite symmetric tensor field, i.e. a Riemannian metric on M. □

Remark: the proof just developed **works only to prove the existence of positive-definite (or negative-definite) Riemannian metrics on** M. It does not work if we want to build a pseudo-Riemannian metric on M with signature (r, s). In fact, even if the g^α have the same signature, the g resulting from the sum may not have the same signature and could even be degenerated.

7.3 Riemannian metrics and changes of coordinates

If $(U, \varphi \equiv (x^1, \ldots, x^n))$ and $(\tilde{U}, \tilde{\varphi} \equiv (\tilde{x}^1, \ldots, \tilde{x}^n))$ are two local charts, then in $U \cap \tilde{U}$ the transition functions allow us to express $\tilde{x}$ as a function of x and vice-versa. We already know that the differentials of $\tilde{x}$ and that of x are related by the Jacobian matrix of the function $\tilde{x} = \tilde{x}(x)$:

$$d\tilde{x}^h = \frac{\partial \tilde{x}^h}{\partial x^i} dx^i. \tag{7.1}$$

If we write the Riemannian metric g in terms of the $\tilde{x}$ and x coordinates, we have:

$$g = \tilde{g}_{hk} d\tilde{x}^h \otimes d\tilde{x}^k = g_{ij} dx^i \otimes dx^j,$$

where the matrices $(\tilde{g}_{hk})$ and (g_{ij}) represent the metric in the local coordinate systems $\tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^n)$ and $x = (x^1, \ldots, x^n)$, respectively.

By replacing (7.1) in the expression of g we obtain:

$$g = \tilde{g}_{hk}\, d\tilde{x}^h \otimes d\tilde{x}^k = \tilde{g}_{hk} \left(\frac{\partial \tilde{x}^h}{\partial x^i} dx^i \right) \otimes \left(\frac{\partial \tilde{x}^k}{\partial x^j} dx^j \right) \underset{\text{bilinearity of } \otimes}{=} \left(\frac{\partial \tilde{x}^h}{\partial x^i}\, \tilde{g}_{hk}\, \frac{\partial \tilde{x}^k}{\partial x^j} \right) dx^i \otimes dx^j,$$

which, compared with $g = g_{ij} dx^i \otimes dx^j$ gives:

$$\boxed{g_{ij} = \frac{\partial \tilde{x}^h}{\partial x^i}\, \tilde{g}_{hk}\, \frac{\partial \tilde{x}^k}{\partial x^j}}\ .$$

If we use the matrix notation we can re-write this relationship as follows:

$$(g_{ij}) = \left(\frac{\partial \tilde{x}}{\partial x} \right)^t (\tilde{g}_{hk}) \left(\frac{\partial \tilde{x}}{\partial x} \right),$$

where $\left(\frac{\partial \tilde{x}}{\partial x}\right)$ is the Jacobian matrix of the transition function $\tilde{x}(x)$. This is coherent with the well-known linear algebra result which says that the matrices associated to symmetric bilinear forms transform, after a change of basis, by multiplication with the change of basis matrix on the right and its transposed (not its inverse) on the left.

This fact has an important consequence: the determinant of the matrices associated to the metric g in different coordinate systems are, in general, different, in fact:

$$\det(g_{ij}) = \det(\tilde{g}_{hk}) \left(\det \left(\frac{\partial \tilde{x}}{\partial x} \right) \right)^2,$$

i.e. they are related by the square of determinant of the Jacobian matrix of the transition function $\tilde{x}(x)$. The information that we can assure is that the sign of $\det(g_{ij})$ and $\det(\tilde{g}_{hk})$ is the same.

Chapter 8

Connections on vector bundles

Connections on manifolds are also called, in particular in the Physics literature, covariant derivatives. To motivate the exigence of introducing these objects let us start by showing a problem related to the Lie derivative that can be underlined already in the trivial case in which the manifold M is an open set U in $\mathbb{R}^n$.

8.1 Motivation

Consider two vector fields $X, Y \in \mathfrak{X}(U)$. Since all the tangent spaces in every point $p \in U$ can be identified with $\mathbb{R}^n$, i.e. $T_pU \cong T_p\mathbb{R}^n \cong \mathbb{R}^n$, X and Y can be simply thought as vector-valued functions defined on U: $X, Y : U \subset \mathbb{R}^n \to \mathbb{R}^n$.

Thanks to this identification, the derivative of Y along X in every point $p \in U$ can be identified with the directional derivative of $Y : U \to \mathbb{R}^n$ in the direction defined by the vector $X(p) \equiv X_p$. We write:

$$\partial_X Y\big|_p := D_{X_p} Y(p) \equiv \lim_{\varepsilon \to 0} \frac{Y(p + \varepsilon\, X|_p) - Y_p}{\varepsilon},$$

having used definition (B.3).

Let us examine the properties of ∂_X: for all $X_1, X_2, Y_1, Y_2 \in \mathfrak{X}(U)$, $a, b \in \mathbb{R}$ and $f \in \mathscr{C}^\infty(U)$, we have

1. $\partial_X(aY_1 + bY_2) = a\partial_X Y_1 + b\partial_X Y_2$
2. $\partial_X(fY) = X(f)Y + f\partial_X Y$
3. $\partial_{aX_1+bX_2} Y = a\partial_{X_1} Y + b\partial_{X_2} Y$
4. $\partial_{fX} Y = f\partial_X Y$.

Property 1. follows simply from the linearity of the directional derivative. To understand property 2. notice that $fY : U \to \mathbb{R}^n$, $p \mapsto f(p)Y_p$ is the product of two functions, one real-valued the other vector-valued, defined on U, thus the Leibniz rule must be applied and we get $\partial_X(fY)(p) = D_{X_p} f(p) Y_p + f(p) D_{X_p} Y(p)$, but $D_{X_p} Y(p) = \partial_X Y\big|_p$ and, regarding the first term, we must recall that $X \in \mathrm{Der}(\mathscr{C}^\infty(U))$, i.e. X can be interpreted also as a derivation whose action on the elements of $\mathscr{C}^\infty(U)$ is exactly the directional derivative, i.e. $X_p(f) = D_{X_p} f(p)$. So, $\partial_X(fY)(p) = X_p(f)Y_p + f(p)\, \partial_X Y|_p$ for all $p \in U$, i.e. property 2.

To resume: 1. & 2. $\implies$ $\partial_X(Y)$ **is $\mathbb{R}$-linear but not $\mathscr{C}^\infty(U)$-linear w.r.t.** Y.

Property 3. is an immediate consequences of the linearity of the directional derivative w.r.t. the directional vector, cfr. formula (B.10). To understand property 4. notice that $\partial_{fX}(Y)|_p = D_{(fX)(p)}(Y)(p) = D_{f(p)X_p}Y(p)$, but $f(p) \in \mathbb{R}$ and $X_p \in \mathbb{R}^n$, so the evaluation of fX in p simply gives a scalar multiple of the vector X_p and thus the property follows again from formula (B.10).

To resume: 3. & 4. $\implies$ $\partial_X(Y)$ **is both $\mathbb{R}$-linear and $\mathscr{C}^\infty(U)$-linear w.r.t.** X.

It is crucial to stress that the other operator that we have defined that implements the derivative of a vector field w.r.t. another one, i.e. the Lie derivative $\pounds_X Y$, does not possess property 4., i.e. it is not $\mathscr{C}^\infty(U)$-linear, in fact, thanks to anti-symmetry and Leibniz rule:

$$\pounds_{fX}Y = [fX, Y] = -[Y, fX] = \pounds_Y fX = -Y(f)X - f\pounds_Y X = -Y(f)X + f\pounds_X Y,$$

thus $\pounds_{fX}Y = -Y(f)X + f\pounds_X Y$, i.e. $\pounds_{fX}Y \neq f\pounds_X Y$.

This shows that **the Lie derivative**, in spite of being a fundamental object that permits to determine conditions to show the existence of integral submanifolds, **cannot be considered as the perfect analogue of the directional derivative** of a function defined on an open subset of $\mathbb{R}^n$.

Another limitation related to the Lie derivative is that it allows us to compute the rate of variation of an object w.r.t. a vector field, only when this object is build from the tangent bundle to a manifold: in fact, vector and covector fields and tensors on a manifold are always built by starting from the tangent bundle. Thus, we cannot avoid the problem underlined above also when we take the Lie derivative of general tensors on a manifold.

The aim of connections (actually the linear ones) is solve both problems at once, i.e. to define a $\mathbb{R}$ and $\mathscr{C}^\infty(U)$- linear derivative along a vector field on M of the section of a general vector bundle E on M, not only of the tangent bundle TM.

8.2 Failed approach towards the generalization of the Lie derivative

It is highly instructive to discuss an approach that goes in the direction that we want, but that fails for one reason that will be underlined. The lesson learned from this failure will allow us understanding the correct path to follow.

Consider a generic vector bundle $\pi : E \to M$ of rank r on a manifold M, let $X \in \mathfrak{X}(M)$ and $s : U \to E$ be a section of E on an open set $U \subset M$. As stressed when we have defined the Lie derivative, given the integral curve γ of X passing through a given point $p \in U$, it does not make sense to compute the derivative of s along X in p as follows

$$\lim_{t \to 0} \frac{s(\gamma(t)) - s(\gamma(0))}{t}$$

simply because $s(\gamma(t)) \in E_{\gamma(t)}$ and $s(\gamma(0)) \in E_{\gamma(0)}$, which are two different vector spaces.

In the trivial case $M = \mathbb{R}^n$ all the fibers are canonically isomorphic to each other and we can perform the difference of vectors belonging to different fibers, but this is not possible if M is not trivial.

Vector bundles are not trivial, but they are always locally trivial, so a more refined idea could be to use local triviality to try to extend our definition of derivative. Let us see how far we can go by using this feature. We know that it exists an open cover $\{U_i\}$ with sets U_i small enough such that $E|_{U_i}$ is trivial, i.e. there are diffeomorphisms $\chi_i : E|_{U_i} \xrightarrow{\sim} U_i \times \mathbb{R}^r$. If we compose the section with the local trivializations χ_i as in the following commutative diagram

$$\begin{tikzcd}
E|_{U_i} \arrow[rr, "\chi_i"] & & U_i \times R^r \\
& U \cap U_i \arrow[ul, "s"] \arrow[ur, "\tilde{s}_i = \chi_i \circ s"'] &
\end{tikzcd}$$

then we obtain $\tilde{s}_i(p) = (p, \tilde{s}_i^1(p), \ldots, \tilde{s}_i^r(p))$ $\forall p \in U$, where each $\tilde{s}_i^k(p) : U \cap U_i \to \mathbb{R}$ is a simple smooth scalar function, for all $k = 1, \ldots, r$, and so we can apply any vector field $X \in \mathfrak{X}(M)$ (interpreted as a derivation of $\mathscr{C}^\infty(U \cap U_i)$) to these functions.

This seems to suggest that we could define the derivative $X(s)$ of the section s, on a small open neighborhood U_i of p, by deriving the functions $\tilde{s}_i^k$ as follows: $X(s)|_{U_i} \approx (X(\tilde{s}_i^1), \ldots, X(\tilde{s}_i^r))$, we use the symbol $\approx$ because we will see that this definition is not entirely correct. The problem with this definition is that, even if it is perfectly correct on U_i, we must assure its coherence when we consider another open cover $\{U_j\}$ and local trivializations χ_j on the intersections $U \cap (U_i \cap U_j)$ (that we will denote simply as $U_i \cap U_j$ to avoid a cumbersome notation). The following commutative diagram shows how the situation looks in this case.

$$\begin{tikzcd}
& E|_{U_i \cap U_j} \arrow[dl, "\chi_i"'] \arrow[dr, "\chi_j"] & \\
(U_i \times U_j) \times \mathbb{R}^r & & (U_i \cap U_j) \times \mathbb{R}^r \arrow[ll, "\eta_{ij}"'] \\
& U_i \cap U_j \arrow[ul, "\tilde{s}_i"] \arrow[ur, "\tilde{s}_j"'] \arrow[uu, dotted, bend right, "s"] &
\end{tikzcd}$$

The two Cartesian products $(U_i \cap U_j) \times \mathbb{R}^r$ written on the left and on the right are characterized by different copies of $\mathbb{R}^r$ that host different coordinates of $\tilde{s}_i$ and $\tilde{s}_j$. They are related by the transition functions $\eta_{ij} : U_i \cap U_j \to \mathrm{GL}(r, \mathbb{R})$ and, for all $p \in U_i \cap U_j$, $\eta_{ij}(p)$ is an invertible matrix that represents the change of coordinates from the two copies of R^r, explicitly:

$$\tilde{s}_i(p) = \left(p, \begin{pmatrix} \tilde{s}_i^1(p) \\ \vdots \\ \tilde{s}_i^r(p) \end{pmatrix} \right) = \left(p, \eta_{ij}(p) \begin{pmatrix} \tilde{s}_j^1(p) \\ \vdots \\ \tilde{s}_j^r(p) \end{pmatrix} \right),$$

i.e. $\tilde{s}_i^k = \eta_{ij} \tilde{s}_j^k$, $k = 1, \ldots, r$.

The components $(X(\tilde{s}_i^1), \ldots, X(\tilde{s}_i^r))$ represent the local expressions of a section of $X(s)$ on E if, on $U_i \cap U_j$, they are related by the transition functions η_{ij} as follows:

$$\begin{pmatrix} X(\tilde{s}_i^1) \\ \vdots \\ X(\tilde{s}_i^r) \end{pmatrix} = \eta_{ij} \begin{pmatrix} X(\tilde{s}_j^1) \\ \vdots \\ X(\tilde{s}_j^r) \end{pmatrix},$$

i.e. $X(\tilde{s}_i^k) = \eta_{ij} X(\tilde{s}_j^k)$, $k = 1, \dots, r$.

Let us see if this is really what happens by applying X (thought as a derivation of $\mathscr{C}^\infty(U_i \cap U_j)$) on both sides of the equation $\tilde{s}_i^k = \eta_{ij}\tilde{s}_j^k$. Thanks to the Leibniz-like behavior of X we find:

$$X(\tilde{s}_i^k) = \eta_{ij} X(\tilde{s}_j^k) + X(\eta_{ij})\tilde{s}_j^k, \qquad \forall k = 1, \dots, r,$$

which differs from $X(\tilde{s}_i^k) = \eta_{ij} X(\tilde{s}_j^k)$ because of *the spurious term given by the derivatives of the transition functions* $X(\eta_{ij})$.

This shows that the components $X(\tilde{s}_i^k)$ define a section representing the derivative of the section s along the vector field X if and only if $X(\eta_{ij}) = 0$ for all i, j. However, this, in general, is not true and so this approach defines an non-intrinsic object that depends on the local trivialization used. The only situation in which this construction works it when the transition functions η_{ij} are locally constant, so that their derivatives are null, in this case we talk about a **flat vector bundle**.

8.3 Connections on vector bundles

In the previous we have shown that:

1. the naive definition of the derivative of the section of a vector bundle w.r.t. a vector field as the limit of the incremental ratio makes no sense because we are comparing vectors belonging to different vector spaces;

2. a finer use of the local trivialization of the vector bundle leads us to an object that makes sense, but that cannot be considered as the derivative of a section because, in general, it depends on the trivialization itself.

The conclusion that we reach is that, unlike the Lie derivative, **there is no intrinsic way to define the derivative of the section of a vector bundle by using only the elements already present in the vector bundle structure**. We are forced to **introduce an external structure**, which is provided by the connection, as we define below (we recall that $\Gamma(E)$ is the set of all sections of the vector bundle $\pi : E \to M$).

Def. 8.3.1 (Connection) *A connection on a vector bundle $\pi : E \to M$ is a function*

$$\begin{array}{rccl} \nabla : & \mathfrak{X}(M) \times \Gamma(E) & \longrightarrow & \Gamma(E) \\ & (X, s) & \longmapsto & \nabla(X, s) \equiv \nabla_X s, \end{array}$$

that transforms the couple given by a vector field X on M and a section s of the bundle (E, M, π) in another section $\nabla_X s$ of the same bundle, in such a way that, for all $X, X_1, X_2 \in \mathfrak{X}(M)$, $f, f_1, f_2 \in \mathscr{C}^\infty(M)$, $s, s_1, s_2 \in \Gamma(E)$ and $k_1, k_2 \in \mathbb{R}$, the following properties are satisfied:

1. *$\mathscr{C}^\infty(M)$-linearity w.r.t. the vector field:* $\boxed{\nabla_{f_1 X_1 + f_2 X_2} s = f_1 \nabla_{X_1} s + f_2 \nabla_{X_2} s}$

2. *$\mathbb{R}$-linearity w.r.t. the section:* $\boxed{\nabla_X (k_1 s_1 + k_2 s_2) = k_1 \nabla_X s_1 + k_2 \nabla_X s_2}$

3. *Leibniz property:* $\boxed{\nabla_X (f s) = f \nabla_X s + X(f) s}$.

These properties are obviously inspired by those of the directional derivative of a function defined on an open set of $\mathbb{R}^n$ that we have discussed in section 8.1 and are *imposed by hand* to make $\nabla_X s$ the correct generalization of the directional derivative in the trivial case.

Def. 8.3.2 (Covariant derivative) *The section $\nabla_X s$ is the covariant derivative of the section s along the vector field X.*

There is a special case that deserves a particular attention and a dedicated definition.

Def. 8.3.3 (Linear connection) *A connection on the tangent bundle TM to a manifold M is called a linear connection on M.*

Having defined a connection does not guarantee that such an object exists. In the special case of a globally trivial vector bundle of rank r, i.e. $E = M \times \mathbb{R}^r$, a connection is easily seen to exist. In fact, a section $s \in \Gamma(M \times \mathbb{R}^r)$ can only have this form

$$\begin{array}{rccl} s: & M & \longrightarrow & M \times \mathbb{R}^r \\ & p & \longmapsto & s(p) = (p, (s^1(p), \dots, s^r(p))), \end{array}$$

where $s^i \in \mathscr{C}^\infty(M)$ for all $i = 1, \dots, r$. If $X \in \mathfrak{X}(M)$, then the *canonical section* defined as follows

$$\begin{array}{rccl} \nabla_X s: & M & \longrightarrow & M \times \mathbb{R}^r \\ & p & \longmapsto & s(p) = (p, ((Xs^1)(p), \dots, (Xs^r)(p))), \end{array}$$

can be verified to be a covariant derivative of s along X (by direct verification of the defining properties), so $\nabla_X : \mathfrak{X}(M) \times \Gamma(M \times \mathbb{R}^r) \to \Gamma(M \times \mathbb{R}^r)$, $(X, s) \mapsto \nabla_X s$ is a connection on the trivial bundle $(M \times \mathbb{R}^r, M, \pi)$.

The following results shows, via a constructive proof that makes use of the partition of unity[1], that at least a connection (actually infinite, as we will see later) exists for all vector bundle.

Theorem 8.3.1 *Every vector bundle $\pi : E \to M$ admits a connection.*

Proof. We have just seen that, for a trivial bundle, a connection can be defined as above. We can always find an open cover (U_α) of M that corresponds to a local trivialization of the bundle, i.e. such that the functions χ_α, $E|_{U_\alpha} \xrightarrow{\chi_\alpha} U_\alpha \times \mathbb{R}^r$, are diffeomorphisms.

On $U_\alpha \times \mathbb{R}^r$ there is a canonical connection ∇^0_X as previously defined. Then, the function χ_α allows us to define a connection ∇^α, which depends on the local trivialization, on $E|_{U_\alpha}$.

To define ∇^α we must declare what is the covariant derivative of a section s of $E|_{U_\alpha}$: the first thing we need to do is to compose χ_α with s to obtain a section of the trivial bundle $U_\alpha \times \mathbb{R}^r$, then we can apply ∇^0_X to this section, obtaining another section of $U_\alpha \times \mathbb{R}^r$, by applying χ_α^{-1} we take this section back to $E|_{U_\alpha}$. Thus:

$$\nabla^\alpha_X s := \chi_\alpha^{-1}(\nabla^0_X(\chi_\alpha \circ s)),$$

[1] We notice that this proof cannot be used to guarantee the existence of connections on complex or algebraic manifolds because they do not possess a partition of unity. In fact, there is no alternative proof for those cases, i.e. the theorem is not valid, in general, for vector bundles of complex or algebraic manifolds.

the properties of a connection are easily proven to be satisfied by ∇^α_X thanks to the fact that ∇^0_X is a connection.

The idea to extend the connection ∇^α_X from the restriction of E on U_α to all E consists in smoothly extend it to zero outside U_α and then to smoothly glue together all the ∇^α_X as U_α varies in the cover. This can be achieved thanks to a partition of the unity (ρ_α) subordinated to the cover (U_α). We recall that each ρ_α is a smooth function defined on M whose support is contained in U_α, i.e. $\rho_\alpha \equiv 0$ on $M \backslash U_\alpha$, and that the functions ρ_α sum up to 1.

Thanks to this, for all (global) section s of E on M, we can define its covariant derivative $\nabla_X s$ along X as follows:

$$\boxed{\nabla_X s := \sum_\alpha \rho_\alpha \, \nabla_X|_{U_\alpha} (s|_{U_\alpha}), \quad \text{with} \quad \rho_\alpha \, \nabla_X|_{U_\alpha} := \rho_\alpha \nabla^\alpha_X} \, .$$

It is customary to write the connection associated to the covariant derivative $\nabla_X s$ simply as $\nabla_X = \sum\limits_\alpha \rho_\alpha \nabla^\alpha_X$.

By direct computation, it can be proven that ∇_X just defined verifies all the properties of a connection. Here, we just verify the Leibniz property. For all $f \in \mathscr{C}^\infty(M)$ we have that

$$\begin{aligned}
\nabla_X(fs) &= \sum_\alpha \rho_\alpha \nabla^\alpha_X (f \, s|_{U_\alpha}) \\
&\quad (\nabla^\alpha_X \text{ is a connection}) \\
&= \sum_\alpha \rho_\alpha (f \, \nabla^\alpha_X(s|_{U_\alpha}) + X(f) \, s|_{U_\alpha}) \\
&\quad (f, X(f) : \text{ independents of } \alpha) \\
&= f \sum_\alpha \rho_\alpha \, \nabla^\alpha_X(s|_{U_\alpha}) + X(f) \sum_\alpha \rho_\alpha \, s|_{U_\alpha} \\
&\quad (\rho_\alpha \, s|_{U_\alpha} = \rho_\alpha s) \\
&= f \nabla_X s + X(f) \sum_\alpha \rho_\alpha s = f \nabla_X s + X(f) s \sum_\alpha \rho_\alpha \\
&\quad (\sum_\alpha \rho_\alpha = 1) \\
&= f \nabla_X s + X(f) s.
\end{aligned}$$

Thus, the Leibniz property holds, the others are even simpler to check. $\square$

As expected from the considerations at the beginning of this chapter, **the Lie derivative**

$$\begin{array}{rccc}
\pounds : & \mathfrak{X}(M) \times \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(M) \\
& (X, Y) & \longmapsto & \pounds_X Y = [X.Y],
\end{array}$$

is not a connection on TM. In fact, $\pounds_{fX} \neq f \pounds_X Y$ for $f \in \mathscr{C}^\infty(M)$. Thus, there is a sort of trade-off between Lie derivative and connection w.r.t. their properties: the Lie derivative is intrinsically defined on TM but it fails to be $\mathscr{C}^\infty$-linear, while a connection is not intrinsically defined on a vector bundle (neither on TM) but it has that property. So, there remains a degree of freedom in the choice of a connection. This ambiguity can be eliminated in special cases, e.g. when the vector bundle has a Riemannian structure, as we will see later with the concept of Levi-Civita connection.

It is useful to single out two noticeable properties of a connection ∇_X:

- for all section s of E on M and all point $p \in M$, $\nabla_X s(p)$ depends on the behavior of the section s in a neighborhood of p;

- instead, $\nabla_X s(p)$ depends only on the value of X in p, i.e. on $X_p \in T_pM$, the other tangent vectors in a neighborhood of p assigned by X are totally irrelevant.

Hence, **the behavior of $\nabla_X s(p)$ is local w.r.t. s and point-wise w.r.t. X**. These features of ∇_X are rigorously stated in the following proposition.

Theorem 8.3.2 (Structural properties of $\nabla_X s$) *Let $\pi : E \to M$ be a vector bundle and ∇ a connection on E.*

1. *If $X, \tilde{X} \in \mathfrak{X}(M)$ are such that $X_p = \tilde{X}_p$ and there exists an open neighborhood U of p such that $s, \tilde{s} \in \Gamma(M)$ are coincident on U, i.e. $s|_U = \tilde{s}|_U$, then*
$$\nabla_X s(p) = \nabla_{\tilde{X}} \tilde{s}(p).$$

2. *For all open set $U \subset M$ there exists only one connection on $E|_U$*
$$\begin{array}{rccc} \nabla^U : & \mathfrak{X}(M) \times \Gamma(U) & \longrightarrow & \Gamma(U) \\ & (X, s) & \longmapsto & \nabla^U(X, s), \end{array}$$
such that, for all $p \in U$, $X \in \mathfrak{X}(M)$ and $s \in \Gamma(M)$, we have:
$$\nabla^U_X\big|_U (p) = \nabla_X s(p).$$

3. *If, for all $X \in \mathfrak{X}(M)$ and $s, \tilde{s} \in \Gamma(M)$ it exists a path $\gamma : (-\varepsilon, \varepsilon) \to M$ such that $\gamma(0) = p$, $\gamma'(0) = X_p$ and $s \circ \gamma = \tilde{s} \circ \gamma$, then $\nabla_X s(p) = \nabla_X \tilde{s}(p)$.*

The third property is a refinement of the first one: for two covariant derivatives to coincide in a point is it enough that they coincide on a 'small' arc of path passing through that point and having the vector X_p as tangent vector in p.

8.3.1 Expression of a connection in local coordinates: the Christoffel symbols

Let (U, φ) be a local chart of M that trivializes E, i.e. such that $E|_U \xrightarrow[\chi]{\sim} U \times \mathbb{R}^r$.

By applying the inverse of χ to the couple given by a generic point $p \in U$ and an arbitrary vector of the canonical basis of $\mathbb{R}^r$, i.e. $(p, (0, \ldots, 0, 1, 0, \ldots, 0))$, where the value 1 is in the k-th position, $k \in \{1, \ldots, r\}$, we determine a local basis for $E|_U$, i.e. r local sections defined on U, that we denote with $(e_1, \ldots, e_r) \in \Gamma(U)$ for simplicity,
$$\begin{array}{rccc} e_k : & U & \longrightarrow & E|_U \\ & p & \longmapsto & \chi^{-1}(p, (0, \ldots, 0, 1, 0, \ldots, 0)), \end{array}$$
such that $(e_1(p), \ldots, e_r(p))$ is a basis of the fiber E_p, for all $p \in U$.

Moreover, we know that the local chart (U, φ) determines a local basis of TM given by $(\partial_1, \ldots, \partial_n)$, $n = \dim(M)$.

Let us then consider ∂_j, $j = 1, \ldots, n$, as the vector field w.r.t. we want to define a connection and e_h, $h = 1, \ldots, r$, as the section of E on which this connection acts. Then $\nabla_{\partial_j} e_h$ is again a section of E, by definition of connection. Hence, there must be suitable functions $\Gamma^k_{jh} \in \mathscr{C}^\infty(U)$ such that:

$$\boxed{\nabla_{\partial_j} e_h = \Gamma^k_{jh} e_k}, \qquad j = 1, \ldots, n, \; h, k = 1, \ldots, r, \tag{8.1}$$

notice that three indices are essential: k is the linear combination index and j, h take into account that the connection is defined w.r.t. the vector field ∂_j and it is applied on the basis section e_h.

Def. 8.3.4 (Connection coefficients - Christoffel symbols) *The functions $\Gamma^k_{jh} \in \mathscr{C}^\infty(U)$ appearing in eq. (8.1) are called (local) connection coefficients. In the special case in which $E = TM$ and $r = n$, the connection coefficients are called Christoffel symbols.*

Def. 8.3.5 (Flat connections) *A connection is said to be flat if all its coefficients are identically 0.*

Let us verify that the connection coefficients determine completely the connection. For any $X \in \mathfrak{X}(M)$ and $s \in \Gamma(U)$ we have:

$$X = X^j \partial_j \quad \text{and} \quad s = s^h e_h, \qquad X^j, s^h \in \mathscr{C}^\infty(U),$$

thus, by definition of connection and by using its properties,

$$\nabla_X s = \nabla_X (s^h e_h) = X(s^h) e_h + s^h \nabla_X e_h,$$

$\nabla_X e_h = \nabla_{X^j \partial_j} e_h = X^j \nabla_{\partial_j} e_h = X^j \Gamma^k_{jh} e_k$, thus, by renaming the summation index $X(s^h) e_h = X(s^k) e_k$, we get

$$\nabla_X s = X(s^k) e_k + s^h X^j \Gamma^k_{jh} e_k = (X(s^k) + \Gamma^k_{jh} X^j s^h) e_k.$$

In the literature sometimes we write simply $s = (s^k)$ avoiding the specification of the basis sections e_k. In this case we get the much easier formula to remember:

$$\boxed{\nabla_X (s^k) = X(s^k) + \Gamma^k_{jh} X^j s^h}, \tag{8.2}$$

in fact, it says that the covariant derivative is composed by two term:

- the first term is simply given by the action of X, interpreted as a derivation, applied on s^k (the equivalent of the directional derivative in $\mathbb{R}^n$);

- the additional term, i.e. the correction w.r.t. the classical directional derivative, is provided by a linear combination in which the connection coefficients appear. Thus, *if the connection is flat* (i.e. its coefficients are all 0), then *the covariant derivative and the directional derivative coincide.*

8.3.2 Parallel sections

We now want to discuss the very important concept of parallel sections. In order to examine this, we need to analyze the properties of covariant derivatives in relation with curves. We begin with a definition.

Def. 8.3.6 (Section along a curve) *Let $\pi : E \to M$ be a vector bundle over M and let $\gamma : I \subseteq \mathbb{R} \to M$ be a path in M. A section of E along γ is a $\mathscr{C}^\infty$ function $s : I \to E$ such that $\forall t \in I$, $s(t) \in E_{\gamma(t)}$.*

Such a section is said to be extendable to a local section $s \in \Gamma(E, U)$ if there exist an open neighborhood U of the image of γ and a section $\tilde{s} \in \Gamma(E, U)$ such that $s(t) = \tilde{s}(\gamma(t))$ $\forall t \in I$.

Notation: the set of sections of E along γ forms a vector space, w.r.t. the point-wise linear operations, that is denoted by $\Gamma(E, \gamma)$.

Theorem 8.3.3 *Let $\gamma : I \to M$ be a path in M and ∇ a connection on E. Then, it exists a unique operator $D : \Gamma(E, \gamma) \to \Gamma(E, \gamma)$ such that:*

1. *D is $\mathbb{R}$-linear, i.e.*
$$D(a_1 s_1 + s_2 s_2) = a_1 D(s_1) + a_2 D(s_2), \qquad \forall a_1, a_2 \in \mathbb{R},\ s_1, s_2 \in \Gamma(E, \gamma).$$
2. *D satisfies the Leibniz rule:*
$$D(fs) = f's + fD(s), \qquad \forall f \in \mathscr{C}^\infty(I).$$
3. *If $s \in \Gamma(E, \gamma)$ is extendable and $\tilde{s}$ is an extension of s to an open neighborhood of the image of γ, then we have:*
$$Ds(t) = \nabla_{\gamma'(t)} \tilde{s}.$$

See [1] for the proof.

Def. 8.3.7 (Covariant derivative along a path) *The operator $D : \Gamma(E, \gamma) \to \Gamma(E, \gamma)$ is called the covariant derivative along the path γ. Ds is the covariant derivative of s along the tangent vectors to the path γ.*

Now we have all the information to introduce the concept of parallel section.

Def. 8.3.8 (Parallel section) *Let ∇ be a connection on the vector bundle E over M and let $\gamma : I \to M$ be a path in M. A section $s \in \Gamma(E, \gamma)$ is said to be parallel (along γ) if $Ds \equiv 0$.*

Instinctively, the request $Ds \equiv 0$ could leads us to think that the section s remains 'constant' along γ, but this is not the case and the word parallel is actually more adequate, let us see why.

As t runs in I, the corresponding point in M over the image of the path γ changes, thus when we apply the section s to t we obtain a sequence of vectors belonging to the fibers $E_{\gamma(t)}$, $t \in I$.

Now, since D measures the rate of variation of the section s along γ, the fact that $Ds \equiv 0$ is naturally interpreted as the fact that the vectors $s(t)$ are as similar as possible as we move from one point to another of the image of γ in M.

Since these vectors belong to different fibers and the fibers are not canonically isomorphic vector spaces, being as similar as possible cannot be translated to being constant, i.e. the same vector. It is thus more correct to use the word 'parallel' instead of constant.

Let us use eq. (8.2), i.e. $\nabla_X(s^k) = X(s^k) + \Gamma^k_{jh} X^j s^h$, to further analyze the consequences of the condition $Ds \equiv 0$. Thanks to property 3. of theorem 8.3.3, the action of X in $\gamma(t)$ is simply the derivative of the path γ in t, i.e. $X_{\gamma(t)} = \gamma'(t)$, thus:

$$X(s^k) + \Gamma^k_{jh} X^j s^h = 0 \qquad \Longleftrightarrow \qquad \frac{ds^k}{dt}(\gamma(t)) + \Gamma^k_{jh} \frac{d\gamma^j}{dt}(t) s^h = 0 \qquad \forall k = 1, \ldots, r,$$

where $\frac{ds^k}{dt}(\gamma(t))$ replaces $X(s^k)$ because this is the derivation of the function s^k in the direction given by X, with X tangent to γ in every point. But, thanks to the point 6. of the flux theorem 5.2.3, computing the derivative of a s^k in the tangent direction to γ is the same as evaluating s^k on the points belonging to the curve $\gamma(t)$ and then computing the derivative w.r.t. the parameter t.

Thanks to these identifications, we have written explicitly eq. (8.2) as a system of ordinary differential equations, that, as we recall in the next theorem, always admits a unique solution.

Theorem 8.3.4 ($\exists!$ of solutions of a system of ODE) *Let $I \subseteq \mathbb{R}$ be an interval, $k \geqslant 1$, $t_0 \in I$, $x_0, \ldots, x_{k-1} \in \mathbb{R}^n$, and $A : I \times (\mathbb{R}^n)^k \to \mathbb{R}^n$ a $\mathscr{C}^\infty$ function, linear w.r.t. the variables in $(\mathbb{R}^n)^k$. Then, the Cauchy problem*

$$\begin{cases} \frac{d^k s}{dt^k}(t) = A(t, s(t), \ldots, \frac{d^{k-1}s}{dt^{k-1}}(t)) \\ s(t_0) = x_0 \\ \frac{ds}{dt}(t)(t_0) = x_1 \\ \quad \vdots \\ \frac{d^{k-1}s}{dt^{k-1}}(t_0) = x_{k-1}, \end{cases}$$

admits a unique $\mathscr{C}^\infty$ solution $s : I \to \mathbb{R}^n$.

Thanks to this result, given any point $p \in M$, we can extend any vector $v \in E_p$, keeping it 'parallel to itself', along a curve passing through p, as depicted in the following figure.

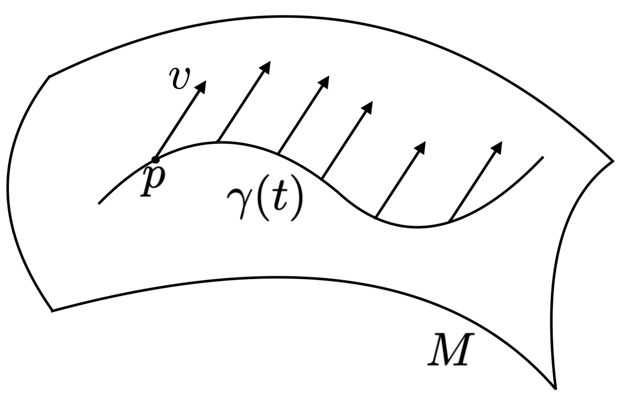


Figure 8.1: A graphic representation of the concept of parallel transport of a vector along a curve thanks to the presence of a connection.

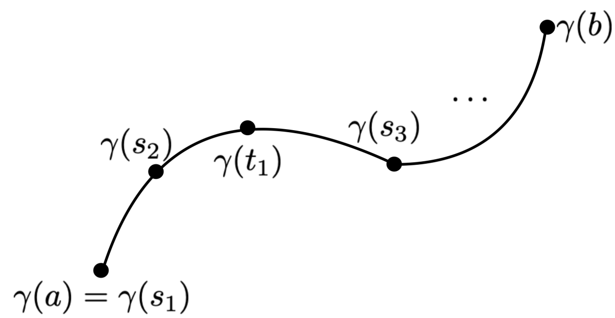
Theorem 8.3.5 *Let $\pi : E \to M$ a vector bundle on M, ∇ a connection on E and $\gamma : [a,b] \to M$ a path in M. If $p = \gamma(a)$, then, for all $v \in E_p$ there exists a unique parallel section $V \in \Gamma(E, \gamma)$ such that $V(a) = v$.*

Proof. The only slightly technical part of the proof consists in the fact that the curve, in general, is not contained in a chart domain. This problem can be fixed by using the compactness of the interval $[a,b]$: the fact that it exists a finite open covering of $[a,b] = \bigcup_{j=1}^{k} [s_j, t_j]$ implies that there is a finite number of charts $(U_1, \varphi_1), \ldots, (U_k, \varphi_k)$, chosen from a local trivialization of E, that cover the image of γ.

Modulo a suitable choice of the covering, we can also suppose that $\gamma([s_j, t_j]) \subseteq \gamma([a,b]) \cap U_j$, for $j = 1, \ldots, k$.

Then, the existence and uniqueness theorem for solutions of a system of ODE quoted before implies that it exists a unique parallel section V_1 along $\gamma|_{[s_1,t_1]}$ such that $V_1(a) = v$.

Thanks again to compactness, we have the freedom to chose the covering of $[a,b]$ as follows: $a = s_1 < s_2 < t_1 < s_3 < t_2 < \cdots < t_{k-1} < t_k = b$, i.e. the sub-intervals that cover $[a,b]$ are partially overlapping (see the picture below).



This trick serves our purposes because, when we solve the system of ODEs in the second open neighborhood, we obtain a unique parallel section V_2 along $\gamma|_{[s_2,t_2]}$ such that $V_2(t_1) = V_1(t_1)$.

By uniqueness, V_1 and V_2 must be equal on $[s_2, t_1]$, so, by gluing together V_1 and V_2, we get a unique parallel section along $\gamma|_{[s_1,t_2]}$.

Following this procedure until $t_k = b$, we obtain a unique parallel section V along γ such that $V(a) = v$. □

This result allows us the possibility to define the extremely useful concept of parallel transport.

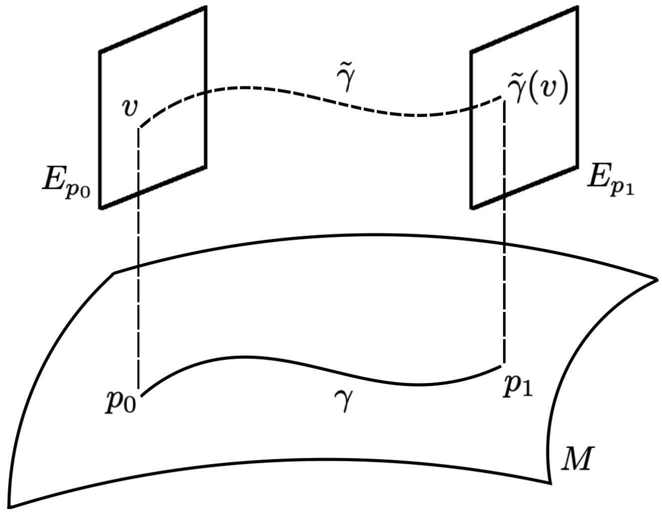
Def. 8.3.9 (Parallel transport) *Let $\pi : E \to M$ a vector bundle on M, ∇ a connection on E and $\gamma : [0,1] \to M$ a path in M with $\gamma(0) = p_0$ and $\gamma(1) = p_1$.*

Given $v \in E_{p_0}$, the only section $V \in \Gamma(E, \gamma)$ parallel along γ and such that $V(0) = v \in E_{p_0}$ is called the parallel extension of v along γ.

The parallel transport along γ is the function:

$$\tilde{\gamma} : E_{p_0} \to E_{p_1}$$

defined by $\tilde{\gamma}(v) = V(1)$, $V \in \Gamma(E, \gamma)$ being the parallel extension of $v \in E_{p_0}$.



The picture above shows the action of parallel transport.

The most important property of the parallel transport is expressed by the following result.

Theorem 8.3.6 *The parallel transport along γ is a linear isomorphism between the vector spaces E_{p_0} and E_{p_1}, the inverse of $\tilde{\gamma}$ being the parallel transport along γ_-, where $\gamma_-(t) := \gamma(1-t)$ is the path that describes that same curve as γ, but traveled in reverse, so: $\tilde{\gamma}^{-1} = \tilde{\gamma}_-$.*

Proof. We have seen that the condition $Dv \equiv 0$, which characterizes parallel sections to a curve, is locally equivalent to:

$$\frac{dV^k}{dt} + \Gamma^k_{jh}(V^j)'s^h = 0 \qquad \forall k = 1, \ldots, r,$$

which is a *linear* system of ODEs. A classical results of the theory of ODEs guarantees that linearity implies that the solution $V(t)$ depends linearly on the initial conditions. This fact is translated in the linearity of the function $\tilde{\gamma} : E_{p_0} \to E_{p_1}$.

Let us now prove that $\tilde{\gamma}^{-1} = \tilde{\gamma}_-$. We denote with D^- the covariant derivative along γ_-. For all section $V \in \Gamma(E, \gamma)$, we set $V^-(t) := V(1-t)$ in such a way that $V^- \in \Gamma(E, \gamma_-)$.

Since $\gamma_-'(t) = \gamma'(1-t) \cdot (1-t)' = -\gamma'(1-t)$, a direct calculation gives $D_t^- V^- = -D_{1-t}V$. Since the only difference between the two covariant derivatives is the sign, it follows that V^- is parallel along γ_- if and only if V is parallel along γ. But then, if V is the parallel extension of $v \in E_{p_0}$ along γ, then V^- is the parallel extension of $V(1) = \tilde{\gamma}(v) \in E_{p_1}$ along γ_-.

This implies that $\tilde{\gamma}_- = \tilde{\gamma}^{-1}$, so $\tilde{\gamma}$ is an isomorphism. □

The *parallel transport is defined also along piece-wise smooth paths*: it is enough to compose the parallel transport along the smooth pieces and use the final value of a piece as the initial condition for the following piece.

8.4 Relationship between connections and differential forms

It is possible to give an alternative definition of a connection, which is more suitable to be used than the previous definition in certain situations. This alternative formulation reveals a strong link between connections and differential forms.

We recall from def. 4.5.3 that a k-form on a manifold M is a section of $\Lambda^k(T^*M)$, i.e. a smooth assignment of an alternating tensor on T^*M and that the vector space of all k-forms on M is written either $\mathbb{A}^k(M)$ or $\Omega^k(M)$.

Consider now a vector bundle $\pi : E \to M$.

Def. 8.4.1 *A k-form with values in E is a section of $\Lambda^k(T^*M) \otimes E$. The vector space of all k-forms with values in E is denoted with either $\mathbb{A}^k(E)$ or $\Omega^k(E)$.*

In local coordinates, the general element of $\mathbb{A}^k(E)$ can be written as

$$\sum_i \omega_i \otimes s_i, \tag{8.3}$$

where s_i are sections of E, i.e. elements of $\mathbb{A}^0(E) \equiv \Gamma(E, M)$, while $\omega_i \in \mathbb{A}^k(M)$ are k-forms on M.

With these definitions and notations, an alternative definition of connection on E can be given as follows.

Def. 8.4.2 (Alternative definition of connection) *A connection on E is a $\mathbb{R}$-linear operator*

$$\nabla : \mathbb{A}^0(E) \equiv \Gamma(E, M) \to \mathbb{A}^1(E)$$

such that

$$\boxed{\nabla(fs) = f\nabla s + df \otimes s} \qquad \forall f \in \mathscr{C}^\infty(M),\ \forall s \in \Gamma(E, M). \tag{8.4}$$

The request expressed in (8.4) is the equivalent of the Leibniz rule in the present context. The first term of (8.4), i.e. $f\nabla s$ is immediate to understand: it is the function f not derived multiplied by the derivative of s, which is provided by ∇ itself.

To comprehend the reason underlying the second term, i.e. $df \otimes s$, notice that we expect s not derived 'multiplied by a derivative of f', and this derivative must provide a 1-form on M. Thanks to (8.3) we see that the only intrinsic way to achieve this is by taking as multiplication the tensor product and as 'derivative' of $f : M \to \mathbb{R}$ its differential, which, as we know, is a 1-form on M.

At first glance, this definition of connection, apart the request of a Leibniz-like behavior just discussed, seems quite unrelated to the original definition 8.3.1 because no vector field enters into play here. To understand the link between the two definitions we must consider the following pairing (which acts on E-values 1-forms and vector fields on M and gives back sections of E):

$$\begin{array}{rccc} \langle\ ,\ \rangle : & \mathbb{A}^1(E) \times \mathfrak{X}(M) & \longrightarrow & \mathbb{A}^0(E) \equiv \Gamma(E, M) \\ & \left(\alpha \equiv \sum_i \omega_i \otimes s_i, X\right) & \longmapsto & \langle \alpha, X \rangle := \sum_i \omega_i(X) s_i, \end{array}$$

perfectly well-defined because ω_i and X are dual objects, one belongs to the tangent and the other to the cotangent space to M, so that $\omega_i(X) \in \mathscr{C}^\infty(M)$.

The relationship between ∇_X and ∇ is then:

$$\nabla_X s = \langle \nabla s, X \rangle. \tag{8.5}$$

Once establish this, let us see how the novel definition of connection can be written in local coordinates. Let $(e_1, \dots, e_r)$ a local frame for E on an open $U \subseteq M$, i.e. a set of r sections of E that, in every point $p \in U$, form a basis of the fiber E_p, then

$$\nabla e_j = \omega_j^k \otimes e_k, \qquad k = 1, \dots, r,$$

where ω_j^k are 1-forms defined on the open U.

What just said is true for every open U, in the particular case when U is a chart domain for M, we have at disposal a local coordinate system $(U, \varphi = (x^1, \dots, x^n))$ and the 1-forms $dx^1, \dots, dx^n$ are a local basis of T^*M, hence we can represent the 1-forms ω_j^k as follows:

$$\omega_j^k = \sum_{i=1}^{n} \Gamma_{ij}^k dx^i$$

for suitable functions $\Gamma_{ij}^k \in \mathscr{C}^\infty(U)$. They are denoted like this because, as we shall see in a moment, they agree with the connection coefficients defined in 8.3.4. To verify this, we select as vector field $X = \partial_i$ and we compute the covariant derivative of e_j w.r.t. X by means of eq. (8.5):

$$\begin{aligned}
\nabla_{\partial_i} e_j &= \langle \nabla e_j, \partial_i \rangle = \langle \omega_j^k \otimes e_k, \partial_i \rangle \\
&= \langle \omega_j^k \otimes e_k, \partial_i \rangle \\
&\quad \text{definition of } \langle\ ,\ \rangle \\
&= \omega_j^k(\partial_i) e_k = \sum_{k=1}^{r} \Gamma_{hj}^k dx^h(\partial_i) e_k = \sum_{k=1}^{r} \Gamma_{hj}^k \delta_i^h e_k \\
&= \Gamma_{ij}^k e_k, \qquad k = 1, \dots, r,\ i, j = 1, \dots, n,
\end{aligned}$$

but then the functions Γ_{ij}^k satisfy eq. (8.1), i.e. the definition of connection coefficients. These considerations justify the following definition.

Def. 8.4.3 (Connection 1-form) *The matrix of 1-forms $\omega = (\omega_j^k)$, $j = 1, \dots, n$, $k = 1, \dots, r$, where*

$$\omega_j^k = \Gamma_{ij}^k dx^i$$

are 1-forms defined on the chart domain $(U, (x^1, \dots, x^n))$, is called the connection 1-form associated to the connection ∇ w.r.t. the local frame selected.

As always, it is important to establish how the expression of ω changes when we change the local reference frame. Since a local frame is a basis of a vector space, if $(\tilde{e}_1, \dots, \tilde{e}_r)$ is another local frame for E on the same chart domain U, there exists an invertible matrix $A = (a_h^k)$ of functions $a_h^k \in \mathscr{C}^\infty(U)$ such that:

$$\tilde{e}_h = a_h^k e_k,$$

where the a_h^k are smooth functions of the point $p \in U$ because the dependence of the fiber E_p on p is smooth.

Let $\tilde{\omega} = (\tilde{\omega}_i^h)$ the connection 1-form of ∇ w.r.t. the local frame $(\tilde{e}_1, \ldots, \tilde{e}_r)$, then, by using the multi-linearity of the tensor product, we have:

$$\nabla \tilde{e}_i = \tilde{\omega}_i^h \otimes \tilde{e}_h = \tilde{\omega}_i^h \otimes a_h^k e_k = a_h^k \tilde{\omega}_i^h \otimes e_k,$$

on the other side, thanks to Leibniz's rule, we also have:

$$\begin{aligned}
\nabla \tilde{e}_i &= \nabla(a_i^k e_k) = a_i^k \nabla e_k + da_i^k \otimes e_k = a_i^k \omega_k^\ell \otimes e_\ell + da_i^k \otimes e_k \\
&\quad \text{indices change: } k \leftrightarrow j,\ \ell \leftrightarrow k \\
&= a_i^j \omega_j^k \otimes e_k + da_i^k \otimes e_k \\
&= (a_i^j \omega_j^k + da_i^k) \otimes e_k.
\end{aligned}$$

Since the vector basis (e_k) of the two expressions of $\nabla \tilde{e}_i$ that we have determined are the same, the coefficients must agree, this implies that:

$$a_h^k \tilde{\omega}_i^h = a_i^j \omega_j^k + da_i^k, \qquad \forall k = 1, \ldots, r,\ i = 1, \ldots, n.$$

We notice that $a_h^k \tilde{\omega}_i^h$ is nothing but the matrix product between A and $\tilde{\omega}$, while (notice the indices position) $a_i^j \omega_j^k$ is the matrix product between ω and A, so, in matrix notation, the previous transformation law can be written as follows:

$$A\tilde{\omega} = \omega A + dA,$$

or

$$\boxed{\tilde{\omega} = A^{-1}\omega A + A^{-1} dA}, \tag{8.6}$$

an expression that has a fundamental importance in gauge field theory.

Example of computation of covariant derivative: let us consider the simple case of a vector bundle of rank 1, i.e. a line bundle (each fiber is a straight line). In this case, the matrix $\omega = (\omega_j^k)$ is a 1×1 matrix of 1-forms, i.e. simply a 1-form

$$\omega = \omega_1^1 = \Gamma_{i1}^1 dx^i = \Gamma_i dx^i.$$

Hence, the connection 1-form in this case is simply a differential form, or covector:

$$\omega = (\Gamma_i) = (\Gamma_1, \ldots, \Gamma_n).$$

If $X = X^j \partial_j \in \mathfrak{X}(U)$ and $s = s^1 e_1 \in \Gamma(E, M)$, where $s^1 \in \mathscr{C}^\infty(U)$ and e_1 is a local basis of E in U, we have:

$$\nabla_X s = (X(s^1) + \Gamma_{j1}^1 X^j s^1) e_1 \equiv (X(s^1) + \Gamma_j X^j s^1) e_1,$$

since 1 is fixed, so the only running index for Γ is indeed j.

If we avoid the specification of the basis element e_1 and we simplify the expression by writing simply s instead of s^1, we get:

$$\nabla_X s = X(s) + \Gamma_j X^j s.$$

Finally, if we choose as particular vector field $X = \partial_i$, then the covariant derivative takes the following form:

$$\boxed{\nabla_{\partial_i} s = \partial_i s + \Gamma_i s},$$

which shows that, *for line bundles, the covariant derivative is simply the partial derivative plus an extra term proportional to the section itself by the* (only) *connection coefficient.*

8.5 Induced connection on tensor bundles

In the same way as we have extended the concept of Lie derivative from vector to tensor fields by forcing the Leibniz rule to be satisfied, we can extend the concept of connection to tensor bundles. The following proposition state this rigorously.

Theorem 8.5.1 *Let M be a smooth manifold and ∇ a connection on TM. Then, it exists a unique way to define a connection ∇ on $T^p_q M$, $\forall p, q$, that satisfies the following properties:*

1. *∇ coincides with the given connection on TM (i.e. it is an actual extension of ∇, this is why we keep the same symbol)*

2. *on $T^0 M \equiv \mathscr{C}^\infty(M)$ the action of ∇ is simply the usual derivation implemented by a vector field, i.e. $\nabla_X(f) = X(f)$, $\forall X \in \mathfrak{X}(M)$*

3. *if $t_j \in T^{h_j}_{k_j}(M)$, $j = 1, 2$, and $X \in \mathfrak{X}(M)$, the following Leibniz rule holds:*

$$\nabla_X(t_1 \otimes t_2) = (\nabla_X t_1) \otimes t_2 + t_1 \otimes (\nabla_X t_2)$$

4. *∇ commutes with contractions.*

Moreover, if $\eta \in T_1(M) \equiv \mathbb{A}^1(M)$ and $X, Y \in \mathfrak{X}(M)$, the following Leibniz rule holds[2]*:*

$$X(\eta(Y)) = (\nabla_X \eta)(Y) + \eta(\nabla_X Y), \tag{8.7}$$

which gives a formula to compute ***the covariant derivative of a 1-form****:*

$$\boxed{(\nabla_X \eta)(Y) = X(\eta(Y)) - \eta(\nabla_X Y)}\ . \tag{8.8}$$

Proof. Let us verify the uniqueness. Suppose that ∇ satisfies the properties 1. $-$ 4. Then, given $\eta \in \mathbb{A}^1(M)$ and $X, Y \in \mathfrak{X}(M)$, we have that $\eta(Y)$ is a function belonging to $\mathscr{C}^\infty(M)$, thus, thanks to 2., $\nabla_X(\eta(Y)) = X(\eta(Y))$.

Now, using the Leibniz rule satisfied by ∇_X, we get eq. (8.7). This shows that the connection on TM determines uniquely the connection on T^*M.

Property 3. determines uniquely the connection on all the tensor bundles $T^h_k M$:

$$\begin{aligned}(\nabla_X t)(\omega^1, \ldots, \omega^h, Y_1, \ldots, Y_k) = {} & X(t(\omega^1, \ldots, \omega^h, Y_1, \ldots, Y_k)) \\ & - \sum_{r=1}^{h} t(\omega^1, \ldots, \nabla_X \omega^r, \ldots, \omega^h, Y_1, \ldots, Y_k) \\ & - \sum_{s=1}^{k} t(\omega^1, \ldots, \omega^h, Y_1, \ldots, \nabla_X Y_s, \ldots, Y_k).\end{aligned}$$

To show the existence, it is enough to define ∇ on T^*M and $T^h_k M$ as above, the fact that it is a connection is tautological because we have defined it by requiring the validity of the Leibniz rule (the other properties are automatically satisfied). □

[2] The coupling $\eta(Y)$ must be thought as of product.

8.5.1 Explicit formulae for covariant derivatives of tensors relatives to linear connections

Given a smooth manifold M of dimension n, let ∇ be a linear connection on M, i.e. a connection on TM, and let $(U, \varphi \equiv (x^1, \dots, x^n))$ a local chart on M. We know that $(\partial_1, \dots, \partial_n)$ is a local frame for TM on the open set U and we can write

$$\boxed{\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k},$$

where Γ^k_{ij} are the Christoffel symbols.

Our aim is to find explicit formulae to compute the covariant derivative of any tensor in the case of a linear connection.

We have seen that, if $X \in \mathfrak{X}(M) = T^1_0(M)$, $X = X^j\partial_j$, then, by Leibniz's rule:

$$\nabla_{\partial_i}X = \nabla_{\partial_i}(X^j\partial_j) = (\partial_i Y^j)\partial_j + X^j\nabla_{\partial_i}\partial_j = (\partial_i X^k)\partial_k + X^j\Gamma^k_{ij}\partial_k = (\partial_i X^k + X^j\Gamma^k_{ij})\partial_k,$$

thus, the **components of the covariant derivative of a vector field** $X \in \mathfrak{X}(M)$ can be explicitly written as follows:

$$\boxed{\nabla_{\partial_i}(X^k) = \partial_i X^k + \Gamma^k_{ij}X^j}, \tag{8.9}$$

i.e. **the sum of the usual derivative, plus an extra term containing the components of the vector field multiplied by the Christoffel symbols of the linear connection**.

Let us now repeat the computation by considering 1-forms, i.e. the cotangent space. Let $(dx^1, \dots, dx^n)$ be a local frame for T^*M on U, then $\nabla_{\partial_i}dx^j = \tilde{\Gamma}^j_{ik}dx^k$, where $\tilde{\Gamma}^j_{ik}$ is another set of Christoffel symbols. Notice that now the running index for the sum, k, is positioned below, while before it was positioned above.

The Christoffel symbols $\tilde{\Gamma}^j_{ik}$ and Γ^j_{ik} are of course related and to make their relation explicit we just have to recall that $\langle dx^j, \partial_k\rangle = \delta^j_k$ which is a constant (either 0 or 1), thus $\partial_i\langle dx^j, \partial_k\rangle = 0$. Recalling that the action of ∇_{∂_i} on a smooth scalar function is the same as the action of ∂_i, we get $0 = \partial_i\langle dx^j, \partial_k\rangle = \nabla_{\partial_i}\langle dx^j, \partial_k\rangle$, so, thanks to Leibniz's rule and the bilinearity of the pairing $\langle\ ,\ \rangle$ we have:

$$\begin{aligned} 0 = \nabla_{\partial_i}\langle dx^j, \partial_k\rangle &= \langle\nabla_{\partial_i}dx^j, \partial_k\rangle + \langle dx^j, \nabla_{\partial_i}\partial_k\rangle = \langle\tilde{\Gamma}^j_{i\ell}dx^\ell, \partial_k\rangle + \langle dx^j, \Gamma^k_{ik}\partial_k\rangle \\ &= \tilde{\Gamma}^j_{i\ell}\langle dx^\ell, \partial_k\rangle + \Gamma^k_{ik}\langle dx^j, \partial_k\rangle = \tilde{\Gamma}^j_{i\ell}\delta^\ell_k + \Gamma^k_{ik}\delta^j_k = \tilde{\Gamma}^j_{ik} + \Gamma^j_{ik}, \end{aligned}$$

which implies $\tilde{\Gamma}^j_{ik} = -\Gamma^j_{ik}$, i.e. **the Christoffel symbols that appear in the covariant derivative of the differential form dx^j are exactly the opposites of those appearing in the covariant derivative of the vector field ∂_j**. This implies that:

$$\boxed{\nabla_{\partial_i}dx^j = -\Gamma^j_{ik}dx^k}.$$

As a consequence, if $\omega = \omega_j dx^j$ is a 1-form, we have:

$$\begin{aligned} \nabla_{\partial_i}\omega = \nabla_{\partial_i}(\omega_j dx^j) &\underset{\nabla_{\partial_i}(\omega_j)=\partial_i(\omega_j)}{=} (\partial_i\omega_j)dx^j + \omega_j\nabla_{\partial_i}dx^j \\ &= (\partial_i\omega_k)dx^k + \omega_j(-\Gamma^j_{ik}dx^k) = (\partial_i\omega_k - \Gamma^j_{ik}\omega_j)dx^k, \end{aligned}$$

thus, the **components of the covariant derivative of a 1-form** $\omega \in \mathbb{A}^1(M)$ can be explicitly written as follows:

$$\boxed{\nabla_{\partial_i}(\omega_k) = \partial_i \omega_k - \Gamma^j_{ik}\omega_j}\,, \tag{8.10}$$

i.e. **the sum of the usual derivative, minus an extra term containing the components of the 1-form multiplied by the Christoffel symbols of the linear connection**.

Since a vector field is a tensor field $X \in T^1_0(M)$ and a 1-form is a tensor field $\omega \in T^0_1(M)$, by comparing eqs. (8.9) and (8.10) it is not difficult to imagine, by multilinearity of the tensor, that the explicit formula for the components of the covariant derivative of a tensor field $t \in T^1_1(M)$ is just the usual derivative with two extra terms proportional to the tensor field, with coefficients given by the Christoffel symbols of the connections with plus and minus sign.

To verify this guess, we write $t = t^h_k \partial_h \otimes dx^k$, where the coefficient functions t^h_k are smooth on U. Then we have:

$$\begin{aligned}
\nabla_{\partial_i} t &= \nabla_{\partial_i}(t^h_k \partial_h \otimes dx^k) = (\partial_i t^h_k)\partial_h \otimes dx^k + t^h_k(\nabla_{\partial_i}\partial_h) \otimes dx^k + t^h_k \partial_h \otimes (\nabla_{\partial_i} dx^k) \\
&\qquad \text{(thanks to (8.9), (8.10))} \\
&= (\partial_i t^h_k)\partial_h \otimes dx^k + t^h_k \Gamma^\ell_{ih}\partial_\ell \otimes dx^k + t^h_k \partial_h \otimes (-\Gamma^k_{i\ell} dx^\ell) \\
&\qquad \text{(exchanging } k \leftrightarrow \ell) \\
&= (\partial_i t^h_k)\partial_h \otimes dx^k + \Gamma^h_{i\ell} t^\ell_k \partial_h \otimes dx^k - \Gamma^\ell_{ik} t^h_\ell \partial_h \otimes dx^k \\
&= (\partial_i t^h_k + \Gamma^h_{i\ell} t^\ell_k - \Gamma^\ell_{ik} t^h_\ell)\partial_h \otimes dx^k
\end{aligned}$$

thus, in components:

$$\boxed{\nabla_{\partial_i}(t^h_k) = \partial_i t^h_k + \Gamma^h_{i\ell} t^\ell_k - \Gamma^\ell_{ik} t^h_\ell}\,, \tag{8.11}$$

which shows that the covariant derivative of a tensor field of type $\binom{1}{1}$ is the usual derivative plus two extra terms involving linear combinations of the tensor components with the Christoffel symbols, notice the difference of sign w.r.t. the position, above or below, of the running index for the sum.

By repeating this same computations for a tensor field of type $\binom{p}{q}$ we get the following explicit formula for the covariant derivative of the components:

$$\boxed{\begin{aligned}
\nabla_{\partial_i}(t^{h_1 h_2 \dots h_p}_{k_1 k_2 \dots k_q}) &= \partial_i t^{h_1 h_2 \dots h_p}_{k_1 k_2 \dots k_q} \\
&\quad + \Gamma^{h_1}_{i\ell} t^{\ell h_2 \dots h_p}_{k_1 k_2 \dots k_q} + \Gamma^{h_2}_{i\ell} t^{h_1 \ell h_3 \dots h_p}_{k_1 k_2 \dots k_q} + \dots \Gamma^{h_p}_{i\ell} t^{h_1 h_2 \dots h_{p-1}\ell}_{k_1 k_2 \dots k_q} \\
&\quad - \Gamma^\ell_{ik_1} - t^{h_1 h_2 \dots h_p}_{\ell k_2 \dots k_q} - \Gamma^\ell_{ik_2} t^{h_1 h_2 \dots h_p}_{k_1 \ell k_3 \dots k_q} - \dots - \Gamma^\ell_{ik_q} t^{h_1 h_2 \dots h_p}_{k_1 k_2 \dots k_{q-1}\ell}.
\end{aligned}}$$

To simplify the heavy notation, in literature we find also the symbol $t_{,i}$ to denote $\partial_i t$ and $t_{;i}$ to denote $\nabla_{\partial_i} t$, so that, for example for a vector field $X = X^k \partial_k$ we find the formula:

$$X^k{}_{;i} = X^k{}_{,i} + \Gamma^k_{ij} X^j.$$

8.5.2 Covariant differential, hessian and divergence

Given a linear connection on M and a tensor field $t \in T^h_k(M)$, we define the covariant version of the differential as follows.

Def. 8.5.1 (Covariant differential) *The covariant differential or* ***total covariant derivative*** *is the operator:*

$$\begin{array}{rccc} \nabla : & T_k^h & \longrightarrow & T_{k+1}^h \\ & t & \longmapsto & \nabla t, \end{array}$$

$$\boxed{(\nabla t)(\omega^1, \ldots, \omega^k, X_1, \ldots, X_k, X_{k+1}) := (\nabla_{X_{k+1}} t)(\omega^1, \ldots, \omega^k, X_1, \ldots, X_k)}\,, \tag{8.12}$$

i.e. the covariant derivative w.r.t. the last vector field. If $\nabla t \equiv 0$, t is said to be a ***parallel tensor field****.*

Thanks to the covariant differential it is possible to define a parallel transport for tensors in the exactly analogous way that we introduced before for vector fields.

Let us now see how it is possible to extend two important objects of calculus in $\mathbb{R}^n$: the hessian and the divergence. In $\mathbb{R}^n$ the hessian is the square matrix that contains the second order partial derivatives of a scalar function; in the case of a smooth scalar function f on a manifold, its covariant derivative coincides with its differential, i.e. $\nabla f = df$ which is not a function anymore, but a differential form, thus, if we want to differentiate a second time, we must necessarily apply the covariant differential! These observations motivates the following definition of hessian.

Def. 8.5.2 (Hessian of a smooth scalar function) *Given $f \in \mathscr{C}^\infty(M)$ and a linear connection on M, the tensor field of type $\binom{0}{2}$ defined as:*

$$\nabla(\nabla f) = \nabla(df)$$

is called the hessian of f.

Let us provide a more explicit expression of the hessian. First of all, since $\nabla(\nabla f)$ is a 2-times covariant tensor, a bilinear form that must be applied to a couple of vector fields $X, Y \in \mathfrak{X}(M)$. Then,

$$\nabla(\nabla f)(X, Y) = \nabla_Y(\nabla(f))(X),$$

having used the definition of covariant differential, eq. (8.12), Y playing the role of the last vector field Y_{k+1}. Since $\nabla(f) = df$, we can rewrite the previous formula as

$$\nabla(\nabla f)(X, Y) = (\nabla_Y(df))(X),$$

but the formula to compute the covariant derivative of a 1-form is provided by eq. (8.8), which gives:

$$\nabla(\nabla f)(X, Y) = Y(df(X)) - df(\nabla_Y X),$$

but, by definition of differential, $df(X) = X(f)$ and $df(\nabla_Y X) = (\nabla_Y X)(f)$, so:

$$\boxed{\nabla(\nabla f)(X, Y) = Y(X(f)) - (\nabla_Y X)(f)}\,, \tag{8.13}$$

which shows that the hessian is not simply the composition of the directional derivative of f w.r.t. to X and then w.r.t. Y, as provided by the first term, but there is also an extra term where the covariant derivative w.r.t. Y appears.

The expression of this explicit formula in coordinates will show us the link with the classical expression of the hessian. If $(U, \varphi \equiv (x^1, \ldots, x^n))$ is a local coordinate system in M, then, if we fix the basis $(\partial_1, \dots, \partial_n)$ of $TM|_U$, and take $X = \partial_i$ and $Y = \partial_j$, then we can associate to $\nabla(\nabla(f))$ a matrix whose (i,j) entry is given by:

$$\nabla(\nabla(f))(\partial_i, \partial_j) = \partial_j(\partial_i f) - (\nabla_{\partial_j}\partial_i)(f),$$

i.e.

$$\boxed{\nabla(\nabla(f))(\partial_i, \partial_j) = \partial^2_{ji} f - \Gamma^k_{ji}\partial_k f},$$

which shows that, if $M = \mathbb{R}^n$ with the classical flat connection $\nabla = d$ characterized by $\Gamma^k_{ji} \equiv 0$, we have that the hessian of $f \in \mathscr{C}^\infty(\mathbb{R}^n)$ is the matrix $\left(\frac{\partial^2 f}{\partial x^j \partial x^i}\right)_{ij}$. Instead, *for a non-trivial manifold with a non-flat connection, an extra term involving the Christoffel symbols appears.*

Let us now pass to the divergence: if X is a vector field on M and ∇ is a linear connection on M, then $\nabla X \in T^1_1(M)$, thus it is perfectly well-defined to contract this tensor of type $\binom{1}{1}$ w.r.t. its only covariant and contravariant index. What we obtain turns out to be the generalization of the classical divergence of a vector field on $\mathbb{R}^n$ to the case of a smooth manifold.

Def. 8.5.3 (Divergence of a vector field) *Given $X \in \mathfrak{X}(M)$ and a linear connection on M, the divergence of X is the smooth scalar function defined as follows:*

$$\operatorname{div}(X) = C^1_1(\nabla X),$$

where C is the contraction operator.

As before, let us make this formula explicit by considering a vector field $X = X^k\partial_k$, then we know that

$$\nabla_{\partial_j} X = (\partial_j X^k + \Gamma^k_{jh} X^h)\partial_k,$$

from this it follows that the covariant differential can be written in every chart domain $(U, \varphi \equiv (x^1, \dots, x^n))$ as:

$$\nabla X = (dX^k + \Gamma^k_{jh} dx^j) \otimes \partial_k \in T^1_1(U).$$

We get the divergence of X by contracting the upper and bottom index, i.e. k and j, respectively, which can be done by renaming both of them as k and considering the implicit sum over k:

$$\boxed{\operatorname{div}(X) = C^1_1(\nabla X) = \partial_k X^k + \Gamma^k_{kh} X^h}.$$

In the trivial case $M = \mathbb{R}^n$ with the flat connection, the Christoffel symbols are identically 0 and we obtain the classical formula of the divergence of a vector field, i.e.

$$\operatorname{div}(X) = \sum_{k=1}^n \frac{\partial X^k}{\partial x^k}.$$

8.6 Compatibility between a linear connection and a (pseudo)-Riemannian metric

In this section we discuss the issue of compatibility between the definition of a linear connection on a manifold M, i.e. a connection defined on the tangent bundle TM of M and a Riemannian metric defined on M itself. We first need to formalize this concept.

Def. 8.6.1 *Let (M, g) be a (pseudo)-Riemannian manifold. A linear connection ∇ on M is compatible with the Riemannian metric g if, for all vector field $X, Y, Z \in \mathfrak{X}(M)$, it holds that:*

$$\nabla_X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z). \tag{8.14}$$

The compatibility equation (8.14) is simply the request that a Leibniz-like behavior holds when ∇_X is applied to the scalar product of vector fields induced by the (pseudo)-Riemannian metric g.

Notice also that, since g is a (bilinear) smooth function, $\nabla_X g(Y, Z) = X(g(Y, Z))$.

The compatibility between a linear connection and a (pseudo)-Riemannian metric can be characterized in five other ways, which are listed in the following result.

Theorem 8.6.1 (Characterizations of compatibility connection-metric) *Let (M, g) be a (pseudo)-Riemannian manifold of dimension n and ∇ a linear connection on M. Then, the following assertions are equivalent.*

1. *∇ is compatible with g.*
2. *$\nabla g \equiv 0$, i.e. g is parallel w.r.t. ∇.*
3. *In all local coordinate system $(x^1, \ldots, x^n)$ it holds that:*
$$\partial_k g_{ij} = g_{\ell j}\Gamma^{\ell}_{ki} + g_{i\ell}\Gamma^{\ell}_{kj} \ . \tag{8.15}$$
4. *For every couple of vector fields V, W along the curve γ in M, it holds that*[3]*:*
$$\frac{d}{dt} g(V, W) = g(DV, W) + g(V, DW) \ . \tag{8.16}$$
5. *For all couple of vector field V, W parallel along γ, $g(V, W)$ is constant along γ.*
6. *The parallel transport defined by ∇ along each curve is an isometry, i.e. it is not only an isomorphism between all tangent spaces on the point traveled by the curve, but it also preserved the norms of tangent vectors and the distances between them.*

Proof. The strategy of the proof to demonstrate the equivalences is the following:

$$1. \iff 2. \quad 2. \iff 3. \quad 1. \implies 4. \quad 4. \implies 5. \quad 5. \implies 6. \quad 6. \implies 1.$$

$\boxed{1. \iff 2.}$: ∇g is the total covariant derivative of a tensor of type $\binom{0}{2}$, hence, since the total covariant derivative increases by one unit the covariance degree of a tensor, so $\nabla g \in T^0_3(M)$ and we must apply it to three vector fields. By definition, if we write $(\nabla g)(Y, Z, X)$ this gives the covariant derivative of g w.r.t. the last vector field applied to the first two, i.e. $(\nabla g)(Y, Z, X) = (\nabla_X g)(Y, Z)$. Now we must use what we have learned about the generalization of covariant derivative to general tensors to write:

$$(\nabla g)(Y, Z, X) = (\nabla_X g)(Y, Z) = X(\langle Y, Z\rangle) - \langle \nabla_X Y, Z\rangle - \langle Y, \nabla_X Z\rangle$$

[3] We recall that DV denoted the covariant derivative of V along the direction tangent to the curve γ, and analogously for DW.

having identified $\langle\,,\,\rangle$ with $g(\,,\,)$. This shows that $\nabla g \equiv 0$ if and only if $X(\langle Y, Z\rangle) = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$, which is exactly the definition of compatibility.

$\boxed{2. \iff 3.}$: this equivalence can be proven by re-writing the previous proof in any local coordinate system, i.e.

$$0 = (\nabla g)(\partial_i, \partial_j, \partial_k) = \partial_k \langle \partial_i, \partial_j\rangle - \langle \nabla_{\partial_k}\partial_i, \partial_j\rangle - \langle \partial_i, \nabla_{\partial_k}\partial_j\rangle,$$

which can be re-written as

$$0 = \partial_k g_{ij} - \Gamma^{\ell}_{ki}\, g_{\ell j} - \Gamma^{\ell}_{kj}\, g_{i\ell},$$

which leads to

$$\partial_k g_{ij} = \Gamma^{\ell}_{ki}\, g_{\ell j} + \Gamma^{\ell}_{kj}\, g_{i\ell}.$$

$\boxed{1. \implies 4.}$: the expression $\frac{d}{dt}\langle V, W\rangle$ means that we apply g to the vector fields V and W defined along γ and then we derive w.r.t. to the curve parameter. But this is equivalent to perform a covariant derivative along the $\dot\gamma$, the tangent vector to the curve γ, so

$$\frac{d}{dt}\langle V, W\rangle = \nabla_{\dot\gamma}\langle V, W\rangle \underset{1.}{=} \langle \nabla_{\dot\gamma} V, W\rangle + \langle V, \nabla_{\dot\gamma} W\rangle \underset{\text{def. of } D}{=} \langle DV, W\rangle + \langle V, DW\rangle.$$

$\boxed{4. \implies 5.}$: if V and W are parallel along γ, then $DV = DW = 0$, hence, by 4. we have

$$\frac{d}{dt}\langle V, W\rangle = 0,$$

and so $\langle V, W\rangle$ is constant.

$\boxed{5. \implies 6.}$: the statement 6. is equivalent to the fact that the parallel transport preserves the metric, which is induced by the scalar product, between a generic couple of vector field defined along any curve. But is is exactly what is stated by 5.

$\boxed{6. \implies 1.}$: we start we the hypothesis that the parallel transport along any curve is an isometry and we must end up with the compatibility between ∇ and g. Let us fix the notation: given any point $p \in M$, let γ the curve such that $\gamma(0) = p$ and $\dot\gamma(0) = X_p$. Let us also $(v_1, \ldots, v_n)$ be a g-orthonormal basis of T_pM. If now we parallel transport each vector v_j along γ, then generate a vector field V along γ itself defined simply by:

$$V(\gamma(t)) = v_j(t),$$

where $v_j(t)$ denotes the result of applying the parallel transport to the tangent vector $v_j \in T_pM = T_{\gamma(0)}M$ to the tangent vector $v_j(t) \in T_{\gamma(t)}M$. Thanks to 6., $(v_1(t), \ldots, v_n(t))$ keeps the property of being an orthonormal basis, but this time of $T_{\gamma(t)}M$, for every value of the parameter t. Having at disposal such a basis, we can write any couple of vector fields Y, Z along γ as

$$Y_{\gamma(t)} = Y^h(t)v_h(t), \quad Z_{\gamma(t)} = Z^k(t)v_k(t).$$

Finally, let us now introduce in our analysis the vector field X along γ that we have defined above as having the property that $X_p = \dot\gamma(0)$ for all choice of $p \in M$. Thanks to this property, the covariant derivative $\nabla_{X_p} \langle Y, Z \rangle$ can be equivalently expressed as follows:

$$\begin{aligned}
\nabla_{X_p} \langle Y, Z \rangle &= \frac{d}{dt} \left\langle Y_{\gamma(t)}, Z_{\gamma(t)} \right\rangle \Big|_{t=0} = \frac{d}{dt} \left\langle Y^h(t) v_h(t), Z^k(t) v_k(t) \right\rangle \Big|_{t=0} \\
&= \frac{d}{dt} \left(Y^h(t) Z^k(t) \langle v_h(t), v_k(t) \rangle \right) \Big|_{t=0} = \frac{d}{dt} \left(Y^h(t) Z^k(t) \delta_{hk} \right) \Big|_{t=0} \\
&= \frac{d}{dt} \left(\sum_{h=1}^{n} Y^h(t) Z^h(t) \right) \Bigg|_{t=0} \quad \text{(the sum was hidden in Einstein's notation)} \\
&= \sum_{h=1}^{n} \frac{dY^h}{dt}(0) Z^h(0) + \sum_{h=1}^{n} Y^h(0) \frac{dZ^h}{dt}(0) \\
&= \left\langle \frac{dY^h}{dt}(0) v_h, Z(0) \right\rangle + \left\langle Y(0), \frac{dY^h}{dt}(0) v_h \right\rangle \text{ (again by orthonormality of } (v_h(t))_{h=1}^{n}) \\
&= \langle D_0 Y, Z \rangle + \langle Y, D_0 Z \rangle \\
&= \left\langle \nabla_{X_p} Y, Z \right\rangle + \left\langle Y, \nabla_{X_p} Z \right\rangle \text{ (again because } \gamma(0) = X_p).
\end{aligned}$$

Since this holds for all $p \in M$, this means that ∇ is compatible with g. □

The consequence of property 2. of the previous theorem is that, when ∇ is compatible with g, then the metric tensor is '*covariantly constant*', which is more precisely expressed by the sentence that g is parallel to ∇.

8.7 The Levi-Civita connection

The request that a connection is compatible with a Riemannian could seem, at first glance, a very strong property to be satisfied. While it is true that many connections are not compatible with a given Riemannian metric, it can nevertheless be proven that there still remain infinitely many connections compatible with any given Riemannian metric. To identify uniquely a connection compatible with a Riemannian metric another object must be introduced. It is the particular tensor field defined below.

Def. 8.7.1 (Torsion of a linear connection) *Given a linear connection ∇ on M, the tensor field τ of type $\binom{1}{2}$, $\tau \in T_2^1(M)$, defined by*

$$\begin{array}{rccl}
\tau: & \mathfrak{X}(M) \times \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(M) \\
& (X, Y) & \longmapsto & \tau(X, Y) := \nabla_X(Y) - \nabla_Y(X) - [X, Y]
\end{array}$$

is called torsion of the linear connection ∇.

It can be seen that the torsion of ∇ just compares the difference $\nabla_X(Y) - \nabla_Y(X)$ with the commutator $[X, Y]$, hence τ is identically 0 when these two terms are exactly the same for all vector fields X, Y.

It could seem a bit strange to declare that τ is a tensor field knowing that neither ∇ nor $[\,,\,]$ are $\mathscr{C}^\infty(M)$-linear, however the particular combination in which they appear in the torsion formula arranges the problem. In fact, we must just verify that τ is bilinear, i.e. it is

$\mathscr{C}^\infty(M)$-linear w.r.t. X and Y. This task is simplified by the fact that it is immediate to see that τ is skew-symmetric:

$$\tau(Y,X) = -\tau(X,Y),$$

hence we can perform the verification just on one argument, say X. For all $f \in \mathscr{C}^\infty(M)$ we have:

$$\begin{aligned}
\tau(fX,Y) &= \nabla_{fX}(Y) - \nabla_Y(fX) - [fX,Y] \\
&= f\nabla_X(Y) - \nabla_Y(fX) - [fX,Y] \quad \text{(linearity w.r.t. the 1st entry of } \nabla) \\
&= f\nabla_X(Y) - f\nabla_Y(X) - Y(f)X - [fX,Y] \quad \text{(Leibniz behavior w.r.t. the 2nd of } \nabla) \\
&= f\nabla_X(Y) - f\nabla_Y(X) - Y(f)X - f[X,Y] + Y(f)X \quad \text{(Leibniz behavior of } [\,,\,]) \\
&= f\nabla_X(Y) - f\nabla_Y(X) - f[X,Y] \\
&= f\tau(X,Y).
\end{aligned}$$

Def. 8.7.2 (Torsion-free, or symmetric, connection) *A linear connection ∇ on M is said to be torsion-free, or symmetric, if its associated torsion tensor field is identically 0.*

The following proposition clarifies why a torsion-free connection is also called symmetric.

Theorem 8.7.1 *Let ∇ be a linear connection on M, then the following assertions are equivalent.*

1. *∇ is torsion-free.*
2. *The Christoffel symbols of ∇ w.r.t. to any local coordinate system are symmetric w.r.t. the two lower indices:*
$$\Gamma^h_{ij} = \Gamma^h_{ji}.$$
3. *For all $f \in \mathscr{C}^\infty(M)$, the hessian of f, $\nabla(\nabla f)$, is a symmetric tensor.*

Proof. It is a simple matter of an explicit computation.

$\boxed{1. \iff 2.}$: fixed a local coordinate system we can write $X = X^h\partial_h$ and $Y = Y^k\partial_k$ and we know that

$$\nabla_X Y = (X(Y^k) + \Gamma^k_{jh}X^jY^h)\partial_k,$$

so, having at disposal this explicit formula, we can compute the torsion

$$\tau(X,Y) = (X(Y^k) + \Gamma^k_{jh}X^jY^h)\partial_k - (Y(X^k) + \Gamma^k_{jh}Y^jX^h)\partial_k - [X,Y],$$

we notice that $(X(Y^k) - Y(X^k))\partial_k$ is nothing but the Lie bracket expressed in local coordinates, hence this term cancels out with $[X,Y]$. Rearranging suitably the names of the indices we arrive to the formula:

$$\tau(X,Y) = X^iY^j(\Gamma^k_{ij} - \Gamma^k_{ji})\partial_k.$$

It follows that $\tau(X,Y) = 0$ for all $X,Y \in \mathfrak{X}(M)$ if and only if $\Gamma^k_{ij} = \Gamma^k_{ji}$ for all i,j,k.

$\boxed{1. \iff 3.}$: recalling eq. (8.13), we have

$$\nabla(\nabla f)(X,Y) = Y(X(f)) - \nabla_Y X(f),$$

hence

$$\nabla(\nabla f)(X,Y) - \nabla(\nabla f)(Y,X) = Y(X(f)) - \nabla_Y X(f) - X(Y(f)) + \nabla_X Y(f),$$

which, by definition of Lie bracket, can be re-written as

$$\nabla(\nabla f)(X,Y) - \nabla(\nabla f)(Y,X) = (\nabla_X Y(f) - \nabla_Y X(f) - [X,Y])(f) = \tau(X,Y)(f).$$

It follows that $\nabla(\nabla f)$ is symmetric if and only if $\tau \equiv 0$. □

We are now ready to state and prove the most important theorem of the section, which is one of the most important results of differential geometry and it bears the name of his discoverer: Tullio Levi-Civita.

Theorem 8.7.2 (Levi-Civita's theorem) *Let (M,g) be a (pseudo)-Riemannian manifold of dimension n. The following statements hold.*

1. *On (M,g) there exists a unique torsion-free connection ∇ compatible with g.*
2. *If $\langle\,,\,\rangle$ denotes the inner product induced by g, then ∇ satisfies[4]:*

   $$\langle \nabla_X Y, Z\rangle = \frac{1}{2}\left(X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle + \langle [X,Y],Z\rangle - \langle [Y,Z],X\rangle + \langle [Z,X],Y\rangle\right),$$

   for all $X,Y,Z \in \mathfrak{X}(M)$.
3. *Let $E_1,\dots,E_n$ be a g-orthonormal local frame for TM, then the previous formula becomes*

   $$\langle \nabla_{E_i} E_j, E_k\rangle = \frac{1}{2}\left(\langle [E_i,E_j],E_k\rangle - \langle [E_j,E_k],E_i\rangle + \langle [E_k,E_i],E_j\rangle\right).$$
4. *The Christoffel symbols of ∇ are given by*

   $$\boxed{\Gamma^k_{ij} = \frac{1}{2} g^{k\ell}\left(\frac{\partial g_{\ell j}}{\partial x^i} + \frac{\partial g_{i\ell}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^\ell}\right)}. \tag{8.17}$$

Proof.

1. Let us start by proving the uniqueness of ∇, which is required to be compatible with g and torsion-free. If we write three times the compatibility with g with a cyclic permutation of the vector fields we obtain the following three equations:

   $$X\langle Y,Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle,$$
   $$Y\langle Z,X\rangle = \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle,$$
   $$Z\langle X,Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle.$$

   Adding together the first two equations and subtracting the third, thanks to the symmetry of the real (pseudo)-scalar product we obtain by a direct computation

   $$\begin{aligned}&X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle = \\ &\langle \nabla_X Z - \nabla_Z X, Y\rangle + \langle \nabla_Y Z - \nabla_Z Y, X\rangle + \langle \nabla_X Y + \nabla_Y X, Z\rangle.\end{aligned}$$

[4] By uniqueness, once the metric g is fixed, there must be a formula that determines the connection.

Now let us introduce the property that ∇ is torsion-free, which implies:

$$\nabla_X Z - \nabla_Z X = [X, Z] = -[Z, X], \quad \nabla_Y Z - \nabla_Z Y = [Y, Z],$$

so the previous equation becomes

$$\begin{aligned}&X \langle Y, Z\rangle + Y \langle Z, X\rangle - Z \langle X, Y\rangle = \\ &- \langle [Z, X], Y\rangle + \langle [Y, Z], X\rangle + \langle \nabla_X Y + \nabla_Y X, Z\rangle .\end{aligned}$$

If we want to apply the torsion-free property to $\nabla_X Y + \nabla_Y X$ we must make appear a difference, which can be easily done as follows:

$$\nabla_X Y + \nabla_Y X = 2\nabla_X Y - \nabla_X Y + \nabla_Y X = 2\nabla_X Y + [Y, X] = 2\nabla_X Y - [X, Y].$$

Finally, we arrive to the formula:

$$\begin{aligned}&X \langle Y, Z\rangle + Y \langle Z, X\rangle - Z \langle X, Y\rangle = \\ &- \langle [Z, X], Y\rangle + \langle [Y, Z], X\rangle + 2 \langle \nabla_X Y, Z\rangle - \langle [X, Y], Z\rangle ,\end{aligned}$$

which leads immediately to

$$\langle \nabla_X Y, Z\rangle = \frac{1}{2} \left(X \langle Y, Z\rangle + Y \langle Z, X\rangle - Z \langle X, Y\rangle + \langle [X, Y], Z\rangle - \langle [Y, Z], X\rangle + \langle [Z, X], Y\rangle\right),$$

thus proving that if a connection ∇ exists and it has the properties of being compatible with g and torsion-free, then necessarily ∇ must satisfy the previous equation, which identifies it uniquely.

To prove its existence, knowing that ∇ must satisfy the last equation, we define it in such a way that the corresponding covariant derivative $\nabla_X Y$ satisfies it, i.e.

$$\begin{array}{rccc}\nabla : & \mathfrak{X}(M) \times \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(M) \\ & (X, Y) & \longmapsto & \nabla(X, Y) = \nabla_X Y,\end{array}$$

with

$$\langle \nabla_X Y, Z\rangle := \frac{1}{2} \left(X \langle Y, Z\rangle + Y \langle Z, X\rangle - Z \langle X, Y\rangle + \langle [X, Y], Z\rangle - \langle [Y, Z], X\rangle + \langle [Z, X], Y\rangle\right),$$

for all $Z \in \mathfrak{X}(M)$. Now we must verify that $\nabla_X Y$ defined in this way is actually a vector field, it is $\mathscr{C}^\infty(M)$-linear w.r.t. X, $\mathbb{R}$-linear w.r.t. Y, it satisfies the Leibniz rule, it is compatible with g and torsion-free. These are a lot of properties to verify, but checking them is quite easy.

A quite clever way to prove that $\nabla_X Y$ is actually a vector field is by showing that the map

$$\begin{array}{ccc}\mathfrak{X}(M) & \longrightarrow & \mathbb{R} \\ Z & \longmapsto & \langle \nabla_X Y, Z\rangle\end{array}$$

is a differential 1-form, i.e. a section of the cotangent bundle T^*M. Since the duality between 1-forms and vector fields is implemented by the metric, which defines the inner product between $\nabla_X Y$ and Z, it follows that $\nabla_X Y$ is the dual object to the differential form, i.e. a vector field. The $\mathbb{R}$-linearity of the map is obvious, what is left to be proven is that it is also $\mathscr{C}^\infty(M)$-linear w.r.t. Z.

This property can be proven as follows, given $f \in \mathscr{C}^\infty(M)$:

$$\begin{aligned}\langle \nabla_X Y, fZ\rangle = \frac{1}{2}(&X\langle Y, fZ\rangle + Y\langle fZ, X\rangle - fZ\langle X, Y\rangle + \langle [X,Y], fZ\rangle \\ &- \langle [Y, fZ], X\rangle + \langle [fZ, X], Y\rangle),\end{aligned}$$

but f is a scalar function, so when the scalar product is evaluated in a point it becomes a real number and it can be taken out, hence the previous expression becomes

$$\begin{aligned}\langle \nabla_X Y, fZ\rangle = \frac{1}{2}(&X(f\langle Y, Z\rangle) + Y(f\langle Z, X\rangle) - fZ\langle X, Y\rangle + f\langle [X,Y], Z\rangle \\ &- \langle [Y, fZ], X\rangle + \langle [fZ, X], Y\rangle),\end{aligned}$$

now we use the Leibniz behavior of X, Y and the Lie-bracket to write

$$X(f\langle Y, Z\rangle) = X(f)\langle Y, Z\rangle + fX\langle Y, Z\rangle, \; Y(f\langle Z, X\rangle) = Y(f)\langle Z, X\rangle + fY\langle Z, X\rangle,$$

$$-\langle [Y, fZ], X\rangle = -\langle Y(f)Z, X\rangle - \langle f[Y,Z], X\rangle, \; \langle [fZ, X], Y\rangle = \langle f[Z,X], Y\rangle -^{5} \langle X(f)Z, Y\rangle,$$

so

$$\begin{aligned}\langle \nabla_X Y, fZ\rangle = \frac{1}{2}(&X(f)\langle Y, Z\rangle + fX\langle Y, Z\rangle + Y(f)\langle Z, X\rangle + fY\langle Z, X\rangle - fZ\langle X, Y\rangle \\ &+ f\langle [X,Y], Z\rangle - \langle Y(f)Z, X\rangle - f\langle [Y,Z], X\rangle + f\langle [Z,X], Y\rangle - \langle X(f)Z, Y\rangle).\end{aligned}$$

Since $X(f), Y(f) \in \mathbb{R}$, the terms highlighted in blue in the previous equation cancel out and we remain with

$$\begin{aligned}\langle \nabla_X Y, fZ\rangle &= \frac{1}{2}(fX\langle Y, Z\rangle + fY\langle Z, X\rangle - fZ\langle X, Y\rangle \\ &\quad + f\langle [X,Y], Z\rangle - f\langle [Y,Z], X\rangle + f\langle [Z,X], Y\rangle) \\ &= f\langle \nabla_X Y, Z\rangle.\end{aligned}$$

Analogously, it can be proven that ∇ is $\mathscr{C}^\infty(M)$-linear w.r.t. X, i.e.

$$\langle \nabla_{fX} Y, Z\rangle = \langle f\nabla_X Y, Z\rangle \quad \forall Z \in \mathfrak{X}(M),$$

and that the Leibniz rule holds true, thus proving that ∇ is indeed a connection.

Now let us check the compatibility with g, i.e.

$$\langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle = X\langle Y, Z\rangle,$$

to check this formula we simply have to use two times the expression satisfied by the covariant derivative, in fact, thanks to the symmetry of the (pseudo)-scalar product and the anti-symmetry of the Lie bracket, we obtain

$$\begin{aligned}\langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle = \frac{1}{2}(&X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle + \langle [X,Y], Z\rangle - \langle [Y,Z], X\rangle \\ &+ \langle [Z,X], Y\rangle + X\langle Y, Z\rangle + Z\langle Y, X\rangle - Y\langle X, Z\rangle - \langle [Z,X], Y\rangle \\ &+ \langle [Y,Z], X\rangle - \langle [X,Y], Z\rangle) = X\langle Y, Z\rangle.\end{aligned}$$

This same strategy can be used to check that ∇ is torsion-free. Putting together all the we have checked, the existence and the uniqueness of the connection ∇ of Levi-Civita's theorem is proven.

[5] notice that X is in the second entry of the Lie bracket, so we must use anti-symmetry to make it act as a derivation over f, this is why the minus sign appears.

2. The proof of the second item is clearly intrinsic in the proof of the first item of the theorem.

3. It is an immediate consequence of item 2. because the first three terms containing the inner products are 0 by orthonormality.

4. Let us finally compute the Christoffel symbols, which is done as usual by considering a local coordinate system $(U, (x^j)_{j=1}^n)$ and replace X, Y, Z with the coordinate vector fields $\partial_i, \partial_j, \partial_\ell$. On one side, by definition of Christoffel symbols we have:

$$\langle \nabla_{\partial_i} \partial_j, \partial_\ell \rangle = \left\langle \Gamma_{ij}^k \partial_k, \partial_\ell \right\rangle = \Gamma_{ij}^k \langle \partial_k, \partial_\ell \rangle = \Gamma_{ij}^k g_{k\ell}.$$

On the other side, the formula satisfied by the covariant derivative is greatly simplified in this case because the vector fields given by the partial derivatives commute, i.e. $[\partial_h, \partial_k] = 0$ for all h, k, hence we get

$$\langle \nabla_{\partial_i} \partial_j, \partial_\ell \rangle = \frac{1}{2}(\partial_i \langle \partial_j, \partial_\ell \rangle + \partial_j \langle \partial_\ell, \partial_i \rangle - \partial_\ell \langle \partial_i, \partial_j \rangle),$$

by symmetry of the (pseudo)-scalar product we can write

$$\langle \nabla_{\partial_i} \partial_j, \partial_\ell \rangle = \frac{1}{2}(\partial_i g_{\ell i} + \partial_j g_{i\ell} - \partial_\ell g_{ij}),$$

identifying the two expressions we get

$$\Gamma_{ij}^k g_{k\ell} = \frac{1}{2}(\partial_i g_{\ell i} + \partial_j g_{i\ell} - \partial_\ell g_{ij}),$$

the only thing that remains to be done is to eliminate $g_{k\ell}$ on the left-hand side, which can be done by applying the inverse matrix of the metric, i.e. $g^{k\ell}$, which implies the formula for the Christoffel symbols stated in the item 4.

□

Thanks to the previous theorem, the following definition is well-posed.

Def. 8.7.3 (Levi-Civita connection) *Let (M, g) be a (pseudo)-Riemannian metric, then the only torsion-free linear connection ∇ compatible with g is called the Levi-Civita connection of M.*

An almost immediate corollary of the Levi-Civita theorem is the following, see [1] for the formal proof.

Corollary 8.7.1 *Let $f : (M, g) \to (\tilde{M}, \tilde{g})$ be an isometry between (pseudo)-Riemannian manifolds. Then the following assertions hold.*

1. *f brings the Levi-Civita connection ∇ of M into the Levi-Civita connection $\tilde{\nabla}$ of $\tilde{M}$ in the following sense:*

$$df(\nabla_X Y) = \tilde{\nabla}_{df(X)} df(Y), \quad \forall X, Y \in \mathfrak{X}(M).$$

2. *If γ be a curve in M, then $fcirc\gamma$ is a curve in $\tilde{M}$. Let D be the covariant derivative along γ induced by ∇ and $\tilde{D}$ the one along $f \circ \gamma$ induced by $\tilde{\nabla}$, then*

$$df(DV) = \tilde{D}(df(V)),$$

for all vector field V in M along γ.

For the Euclidean metric in $\mathbb{R}^n$ (induced by the Euclidean scalar product), all the Christoffel symbols are null, so the Levi-Civita connection is flat, so it is nothing but the standard the directional derivative in $\mathbb{R}^n$ along a vector.

A much more interesting example is given by considering a Riemannian manifold (M, g) and his unique Levi-Civita connection ∇^M together with a submanifold N of M which becomes a Riemannian manifold too when endowed with the metric g reduced to N, denoted with $g|_N$.

For all $p \in N$ let us identify T_pN with a vector subspace of T_pM. Thanks to g we can build the *g-orthogonal projection* $P : TM \to TN$ and define a connection on N as follows:

$$\begin{array}{rccl} \nabla^N : & \mathfrak{X}(N) \subset \mathfrak{X}(M) \times \mathfrak{X}(N) \subset \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(N) \subset \mathfrak{X}(M) \\ & (X, Y) & \longmapsto & \nabla^N(X, Y) := P(\nabla^M_X Y), \end{array}$$

it can be proven that ∇^N defined as above is the Levi-Civita connection of N.

This result can be applied, e.g., to submanifolds immersed in $\mathbb{R}^n$: it is sufficient to take the directional derivative on $\mathbb{R}^n$ as Levi-Civita connection and then use the orthogonal projection on the tangent spaces to endow the submanifold with its own Levi-Civita connection.

8.7.1 The Beltrami-Laplace operator

In the standard differential calculus in $\mathbb{R}^n$, it is well-known that the Laplacian operator can be obtained by taking the trace of the Hessian matrix. We can generalize this to general manifolds by interpreting the trace not simply as the sum of the diagonal elements of a matrix, but using its relationship with bilinear forms.

Let us recall that, given a vector space V of finite dimension n endowed with a scalar product $\langle\, , \rangle$ and a symmetric bilinear form $S : V \times V \to \mathbb{R}$, we cannot compute the trace of S intrinsically, because the matrix associated to S changes w.r.t. changes of bases of V not by similarity (i.e. conjugation with the change of bases matrix and its inverse) but by similarity (i.e. conjugation with the change of bases matrix and its transpose), so *the trace of the matrix associated to a symmetric bilinear form depends on the choice of the bases.*

In order to obtain a mathematical object from S whose trace is intrinsic we have to define the endomorphism $S \in \mathrm{End}(V)$ as follows:

$$\begin{array}{rccl} \hat{S} : & V & \longrightarrow & V \\ & v & \longmapsto & \hat{S}(v), \end{array}$$

such that, for all $w \in V$,

$$\left\langle \hat{S}(v), w \right\rangle = S(v, w).$$

To prove that $\hat{S}$ is the unique endomorphism that satisfies this equation, it is useful to pass in coordinates: let $(v_1, \ldots, v_n)$ be a basis of V and let $g_{ij} = g\langle v_i, v_j\rangle$ be the metric coefficients. We can also compute the coefficients of the symmetric bilinear form S by setting $s_{ij} = S(v_i, v_j)$. Notice that both indices of g are placed below and the same holds for S.

On the contrary, we are going to see that the indices of the endomorphism $\hat{S}$ are placed above and below. In fact, let $v = a^i v_i$ and $w = b^j w_j$, then, since $S(v_i) \in V$ there must exist coefficients $s_i^h \in \mathbb{R}$ such that $S(v_i) = s_i^h v_h$. The relationship between the coefficients s_{ij} and s_i^h is well-known: the indices are raised by applying the inverse of the metric tensor expressed as a matrix, i.e.

$$s_i^h = s_{i\ell} g^{\ell h}.$$

So, we have:

$$\left\langle \hat{S}(v), w \right\rangle = \left\langle \hat{S}(a^i v_i), b^j w_j \right\rangle = a^i b^j \left\langle \hat{S}(v_i), w_j \right\rangle = a^i b^j \left\langle s_i^h v_h, w_j \right\rangle = a^i b^j s_i^h \langle v_h, w_j \rangle,$$

but $s_i^h = s_{i\ell} g^{\ell h}$ and $\langle v_h, w_j \rangle = g_{hj}$, hence,

$$\left\langle \hat{S}(v), w \right\rangle = a^i b^j s_{i\ell} g^{\ell h} g_{hj} = a^i b^j s_{i\ell} \delta^{\ell_j} = a^i b^j s_{ij}.$$

On the other side:

$$S(v, w) = S(a^i v_i, b^j v_j) = a^i b^j S(v_i, v_j) = a^i b^j s_{ij},$$

so

$$\left\langle \hat{S}(v), w \right\rangle = S(v, w).$$

Def. 8.7.4 (Trace of a bilinear symmetric form) *With the notations above, the trace of S is the trace of the endomorphism $\hat{S}$. Explicitly:*

$$\mathrm{Tr}(S) = \mathrm{Tr}(\hat{S}) = s_i^i = s_{ij} g^{ji}.$$

Notice that the trace of S depends not only on the matrix (s_{ij}) of S, but also on the metric g.

The application of these concepts to generalize the Laplacian should now be clear: the Levi-Civita connection on a Riemannian metric is symmetric and if apply it two times we obtain the Hessian, which in this case is a symmetric tensor of type $\binom{0}{2}$, hence its coefficients have two indices below and represent the matrix of a bilinear symmetric form, of which we can compute the trace as just seen. Let us formalize what just argued.

Def. 8.7.5 (Laplace-Beltrami operator) *Let (M, g) be a Riemannian manifold and ∇ the Levi-Civita connection on it. Then, the Laplace-Beltrami operator on M is defined as follows:*

$$\begin{array}{rccl} \Delta : & \mathscr{C}^\infty(M) & \longrightarrow & \mathscr{C}^\infty(M) \\ & f & \longmapsto & \Delta(f) := \mathrm{Tr}(\nabla(\nabla f)). \end{array}$$

In general, we call elliptic operators, those differential operators which generalize the Laplacian. The Laplace-Beltrami operator belongs to such category and it is frequently used in problems related to physical problems where a potential energy is involved.

The expression of the Laplace-Beltrami operator in local coordinates is very easy to obtain because we have already seen that, in a local coordinate system $(x^1, \ldots, x^n)$, the expression of the Hessian of a function $f \in \mathscr{C}^\infty(M)$ is

$$\nabla(\nabla f)(\partial_i, \partial_j) = \frac{\partial^2 f}{\partial_{x^i} \partial_{x^j}} - \Gamma_{ij}^k \frac{\partial f}{\partial_{x^k}},$$

whose trace is obtained by raising one of the indices i or j and then by contracting in order to obtain a scalar function, this of course is done by using the inverse of the metric tensor:

$$\boxed{\Delta f = g^{ij} \frac{\partial^2 f}{\partial_{x^i} \partial_{x^j}} - g^{ij} \Gamma^k_{ij} \frac{\partial f}{\partial_{x^k}}}.$$

Notice that the Christoffel symbols Γ^k_{ij} of the Levi-Civita connection are symmetric in ij.

In $\mathbb{R}^n$ with the Euclidean metric and using the Cartesian coordinates, the connection becomes flat and all Christoffel symbols are null, moreover g^{ij} reduces to the identity matrix δ^{ij}, hence the double sum becomes a single sum and we obtain the usual formula for the Laplacian:

$$\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial {x^i}^2}.$$

8.8 Geodesics

Geodesic are curves in a manifold with very particular properties w.r.t. linear connections.

Def. 8.8.1 (Geodesics of a linear connection) *Let M be a manifold of dimension n and ∇ a linear connection on M. A geodesic for ∇ is a curve $\gamma : [0,1] \to M$ such that $D(\dot{\gamma}) \equiv 0$.*

In more explicit terms, γ is a geodesic if and only if its tangent vector $\dot{\gamma}$ is parallel along γ, i.e. its change from one point of the trajectory of γ to another is performed by the parallel transport.

Remarks:

- The concept of geodesic depends only on the connection and so it can be defined also for non-Riemannian manifolds.

- A geodesic does not depend only on the trajectory of γ, but also on its parameterization.

If (U, φ) is a local chart of M in $\gamma(0)$, then $\varphi \circ \gamma : [0,1] \to \mathbb{R}^n$ is a curve in $\mathbb{R}^n$ that can be expressed through its components: $\varphi \circ \gamma = (\gamma^1, \dots, \gamma^n)$. Hence γ is a geodesic if and only if all the component γ^k satisfy $D(\dot{\gamma}^k) \equiv 0$.

We have already seen that the covariant derivative can be split in two terms: the first is the canonical derivative that derives the first derivative of γ^k, thus giving $\ddot{\gamma}^k$; the second contains the Christoffel symbols evaluated on the curve, i.e. $\Gamma^k_{ij} \circ \gamma$, multiplied by the first derivatives of the components i and j, i.e. $\dot{\gamma}^i \dot{\gamma}^j$. The general equation is thus:

$$\ddot{\gamma}^k + (\Gamma^k_{ij} \circ \gamma)\, \dot{\gamma}^i \dot{\gamma}^j = 0, \qquad k \in 1, \dots, n,$$

which is a shorthand notation for the following system of second order ordinary differential equations:

$$\begin{cases} \ddot{\gamma}^1 + (\Gamma^1_{ij} \circ \gamma)\, \dot{\gamma}^i \dot{\gamma}^j = 0 \\ \qquad\qquad \vdots \\ \ddot{\gamma}^n + (\Gamma^n_{ij} \circ \gamma)\, \dot{\gamma}^i \dot{\gamma}^j = 0 \end{cases}.$$

The *mechanical interpretation* of the condition that defines a geodesic is very important: $\dot{\gamma}$ is the velocity vector along the curve γ and so the covariant derivative $D(\dot{\gamma})$ is the **acceleration** along γ. Hence, requiring $D(\dot{\gamma})$ means that the point that follows travels the geodesic γ **never undergoes to an acceleration**, i.e. **a point travels with constant speed on a geodesic**.

Since, by Newton's second law of dynamics, acceleration is proportional to force, this means that point that travels a geodesic is never subjected to a force. In other terms, if not forced otherwise, the curve described by free particles are geodesics.

To confirm this fact, let us consider $M = \mathbb{R}^n$ endowed with the flat connection, i.e. the standard directional derivative, then all the Christoffel symbols are identically 0 and the geodesic equations are simply given by $\ddot{\gamma}^k = 0$, $k = 1, \ldots, n$. Integrating this second order differential equation we find $\gamma^k(t) = a^k + b^k t$, $k = 1, \ldots, n$, i.e. in this case the geodesics are straight lines in $\mathbb{R}^n$ traveled with constant speed.

The theorem of existence and uniqueness of solutions of a system of ordinary differential equations has as a corollary the following result.

Theorem 8.8.1 *Let ∇ be a linear connection on M. For all $p \in M$ and all $v \in T_pM$ there exist an interval $I \subseteq \mathbb{R}$ containing 0 and a geodesic $\gamma : I \to M$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = v$. Moreover, if $\tilde{\gamma} : \tilde{I} \to M$ is another geodesic that satisfies the same conditions, then γ and $\tilde{\gamma}$ coincide on $I \cap \tilde{I}$.*

8.9 Geodesics in Riemannian manifolds

We have seen that, in order to define in the most general way a geodesic, we do not need a Riemannian manifold, the presence of a connection suffices. However, if M happens to be a Riemannian manifold, then the geodesics assume a further very important role: they allow to endow them with a canonical distance, w.r.t. which they are local minimizers.

We start with this definition in a general manifold.

Def. 8.9.1 (Piecewise regular curve) *Let M be a manifold and $\gamma : [0,1] \to M$ a curve. γ is said to be piecewise regular if it exists a partition $0 = t_0 < t_1 < \cdots < t_k = 1$ such that $\gamma|_{[t_{j-1},t_j]}$ is regular, i.e. smooth and with tangent vector never null, for $j = 1, \ldots, k$.*

Now let us consider a Riemannian manifold.

Def. 8.9.2 (Length of a curve in a Riemannian manifold) *Let $\gamma : [0,1] \to M$ be a piecewise regular curve in a Riemannian manifold (M, g). The arc length of γ is the function $s : [0,1] \to [0, +\infty)$ given by*

$$s(t) := \int_0^t \|\dot{\gamma}(u)\|_{\gamma(t)} du,$$

where, of course, $\| \|_p$, for all $p \in M$, is the norm in T_pM induced by the metric g. The length of γ is then

$$L(\gamma) := \int_0^1 s(t) dt = \int_0^1 \|\dot{\gamma}(u)\|_{\gamma(t)} du.$$

If we further suppose that the Riemannian manifold is connected, we can go even further.

Def. 8.9.3 (Riemannian distance) *Let (M, g) be a connected Riemannian manifold, then, for every couple of points $p, q \in M$, the function $d : M \times M \to [0, +\infty)$ defined by*

$$d(p, q) := \inf\{L(\gamma) \,:\, \gamma : [0, 1] \to M \text{ is a piecewise regular curve with } \gamma(0) = p,\ \gamma(1) = q\},$$

is called the Riemannian distance on M induced by g.

The following proposition assures that the previous definition is well-posed, i.e. not only d is a distance on M, but the topology that it induces is not in conflict with that of M seen as a topological space. For the proof, see e.g. [1] page 383.

Theorem 8.9.1 *Let (M, g) be a connected Riemannian manifold, then the Riemannian distance on M induced by g is actually a distance and the metric topology induced on M coincides with its original topology.*

Hence, a connected Riemannian manifold becomes also a metric space (M, d).

The piecewise regular curves that verify the definition of distance are labeled with a special name.

Def. 8.9.4 (Minimizing and locally minimizing curves) *A piecewise regular curve γ on a connected Riemannian manifold (M, g) that satisfies the equation $d(\gamma(0), \gamma(1)) = L(\gamma)$ is called minimizing. If, for every $t \in (0, 1)$, there exists $\varepsilon > 0$ such that $\gamma|_{[t-\varepsilon,t+\varepsilon]}$ is minimizing, i.e. $d(\gamma(t - \varepsilon), d(\gamma(t - \varepsilon)) = L(\gamma|_{[t-\varepsilon,t+\varepsilon]})$ is called locally minimizing.*

Clearly, every minimizing curve is also locally minimizing, but the other way round is not in general true.

We can now state the most important theorem that connects geodesics with the Riemannian distance. Its prove, highly non-trivial, can be found e.g. in [1] pages 385-393.

Theorem 8.9.2 *Every curve γ on a connected Riemannian manifold (M, g) is locally minimizing if and only if it is a geodesic.*

8.10 Curvature

In this section we introduce the concept of curvature of a manifold M through multiple steps. We start with an operator that can be defined each time we fix a linear connection ∇ on M. Given any two vector fields $X, Y \in \mathfrak{X}(M)$, the operator is

$$\begin{array}{rccl} R_{XY} : & \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(M) \\ & Z & \longmapsto & R_{XY}(Z) := \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z. \end{array}$$

To understand more easily how to interpret R_{XY} it is convenient to avoid writing its action on the generic vector field Z, i.e.

$$R_{XY} = (\nabla_X\nabla_Y - \nabla_Y\nabla_X) - \nabla_{[X,Y]} = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}$$

in this way we see that R_{XY} *measures the difference between the Lie bracket of the covariant derivatives along the vector fields X and Y and the covariant derivative along the vector field given by the Lie bracket of X and Y.*

With a direct computation, it can be verified that R_{XY} is $\mathscr{C}^\infty(M)$-linear w.r.t. X, Y, Z. Thanks to this property, the following map

$$\begin{array}{rccl} R: & \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) & \longrightarrow & \mathfrak{X}(M) \\ & (X, Y, Z) & \longmapsto & R(X, Y, Z) := R_{XY}(Z) \end{array}$$

is trilinear, hence it is a tensor field of type $\binom{1}{3}$.

Def. 8.10.1 (Riemann curvature tensor) *The tensor field of type $\binom{1}{3}$ R is called Riemann curvature tensor.*

As always, let us see what is the expression of R is the local coordinates $(x^1, \dots, x^n)$ of an open set $U \subseteq M$, R can be written as

$$R = R_{ijh}{}^k \, dx^i \otimes dx^j \otimes dx^h \otimes \partial_k, \qquad R_{ijh}{}^k \in \mathbb{R},$$

perfectly coherent because if we apply R written in this way to a triple of vector fields, then we end up with a linear combination of the coordinate vector fields ∂_k, which gives a vector field, as it should be. Let us check this assertion by considering, for simplicity $X = \partial_i$, $Y = \partial_j$, $Z = \partial_h$, then the couplings between the differentials and these vector fields give:

$$R(X, Y, Z) = R_{ijh}{}^k \left\langle dx^i, \partial_i \right\rangle \left\langle dx^j, \partial_j \right\rangle \left\langle dx^h, \partial_h \right\rangle \partial_k = R_{ijh}{}^k \, \partial_k \in \mathfrak{X}(M).$$

If we avoid writing the basis vectors, then the expression of R becomes

$$R = \left(R_{ijh}{}^k \right),$$

we will show that R contains all the information on the curvature of the manifold M, a property that now it is not clear at all.

Let us now show how it is possible to express locally the coefficients $R_{ijh}{}^k$ in terms of the Christoffel symbols. On one side, if when we write $R_{\partial_i \partial_j} \partial_h$ we still obtain a vector field, hence there exist coefficients $R_{ijh}{}^k \in \mathbb{R}$ such that

$$R_{\partial_i \partial_j} \partial_h = R_{ijh}{}^k \partial_k. \tag{8.18}$$

On the other side, if we write explicitly the definition of the operator $R_{\partial_i \partial_j}$ applied to ∂_h we have

$$\begin{aligned} R_{\partial_i \partial_j} \partial_h &= \nabla_{\partial_i}(\nabla_{\partial_j} \partial_h) - \nabla_{\partial_j}(\nabla_{\partial_j} \partial_h) - \nabla_{\cancelto{0}{[\partial_i, \partial_j]}} \partial_h \\ &= \nabla_{\partial_i}(\Gamma^\ell_{jh} \partial_\ell) - \nabla_{\partial_j}(\Gamma^\ell_{ih} \partial_\ell) \\ &= \partial_i(\Gamma^\ell_{jh}) \partial_\ell + \Gamma^\ell_{jh} \nabla_{\partial_i} \partial_\ell - \partial_j(\Gamma^\ell_{ih}) \partial_\ell - \Gamma^\ell_{ih} \nabla_{\partial_j} \partial_\ell t \\ &= \partial_i(\Gamma^k_{jh}) \partial_k + \Gamma^\ell_{jh} \Gamma^k_{i\ell} \partial_k - \partial_j(\Gamma^k_{ih}) \partial_k - \Gamma^\ell_{ih} \Gamma^k_{j\ell} \partial_k \\ &= (\partial_i \Gamma^k_{jh} - \partial_j \Gamma^k_{ih} + \Gamma^\ell_{jh} \Gamma^k_{i\ell} - \Gamma^\ell_{ih} \Gamma^k_{j\ell}) \partial_k. \end{aligned}$$

Comparing the expression just found with eq. (8.18) we find a way to express the coefficients of the Riemann curvature tensor solely in terms of the Christoffel symbols of the connection and their partial derivatives:

$$\boxed{R_{ijh}{}^k = \frac{\partial \Gamma^k_{jh}}{\partial x^i} - \frac{\partial \Gamma^k_{ih}}{\partial x^j} + \Gamma^\ell_{jh} \Gamma^k_{i\ell} - \Gamma^\ell_{ih} \Gamma^k_{j\ell}}. \tag{8.19}$$

By the Levi-Civita theorem, if (M, g) is a (pseudo)-Riemannian manifold and ∇ is the Levi-Civita connection, then by eq. (8.17) the Christoffel symbols can be expressed in terms of the coefficients of the metric and their partial derivatives. Hence, in this case, the coefficients of the Riemann curvature tensor of M can be directly computed from the metric of M, even though with a quite complicated formula, if g has not a simple expression.

Remaining in the case of a (pseudo)-Riemannian manifold (M, g), we can use the metric tensor g to lower the index k of the coefficients $R_{ijh}{}^{k}$, which will be associated to a tensor field $\tilde{R}$ of type $\binom{0}{4}$.

$$\tilde{R}(X, Y, Z, W) := \langle R_{XY}(Z), W\rangle, \qquad \forall X, Y, Z, W \in \mathfrak{X}(M).$$

In a local coordinate system $(x^1, \ldots, x^n)$ the expression of R will be

$$\tilde{R} = \tilde{R}_{ijhk}\, dx^i \otimes dx^j \otimes dx^h \otimes dx^k.$$

To obtain the relationship between the coefficients $\tilde{R}_{ijhk}$ of $\tilde{R}$ and $R_{ijh}{}^{k}$ of R it is enough to apply $\tilde{R}$ to the coordinate vector fields:

$$\tilde{R}_{ijhk} = \tilde{R}(\partial_i, \partial_j, \partial_h, \partial_k) = \langle R(\partial_i, \partial_j, \partial_h), \partial_k\rangle = \left\langle R_{ijh}{}^{\ell}\partial_\ell, \partial_k \right\rangle = R_{ijh}{}^{\ell}\langle \partial_\ell, \partial_k\rangle = R_{ijh}{}^{\ell} g_{\ell k},$$

which confirms that the index k is lowered thanks to the application of the metric g. In the following, for simplicity we will write R_{ijhk} instead of $\tilde{R}_{ijhk}$.

By a direct computation, it can be verified that *the Riemann curvature tensor of a Riemannian manifold is invariant under isometries.* This means that, if $f : (M, g) \to (\tilde{M}, \tilde{g})$ is an isometry between Riemannian manifolds with corresponding Levi-Civita connections ∇ and $\tilde{\nabla}$ and Riemann curvature tensors R and $\tilde{R}$, respectively, then

$$\langle R_{XY} Z, W\rangle_p = \left\langle \tilde{R}_{df_p(X) df_p(Y)} df_p(Z), df_p(W) \right\rangle_{f(p)}, \qquad \forall p \in M.$$

8.10.1 Properties of the Riemann curvature tensor

The Riemann curvature tensor satisfies several properties related to symmetry, anti-symmetry and cyclic formulae, as established in the following theorem.

Theorem 8.10.1 *Let R be the Riemannian curvature tensor field of a linear symmetric connection on the manifold M and let $X, Y, Z, W \in \mathfrak{X}(M)$, then the following assertions hold.*

1. $R_{XY} = -R_{YX}$, *so, in particular* $R_{XX} \equiv 0$.
2. $R_{XY} + R_{YZ} + R_{ZX} \equiv 0$, *called* ***'first Bianchi identity'***.

If, moreover, (M, g) is a Riemannian manifold and ∇ is the Levi-Civita connection, then

3. $\langle R_{XY} Z, W\rangle = -\langle Z, R_{XY} W\rangle$, *so, in particular,* $\langle R_{XY} Z, Z\rangle = 0$, *i.e.* $R_{XY} Z$ *and* Z *are always orthogonal.*
4. $\langle R_{XY} Z, W\rangle = -\langle R_{ZW}, Y\rangle$.

Proof. 1. is evident by definition of R_{XY} and the symmetry property of ∇ is not even necessary.

2. Here instead the symmetry of the connection is fundamental, we recall that the symmetry means that $\nabla_X Y - \nabla_Y X = [X, Y]$.

$$\begin{aligned}
R_{XY} + R_{YZ} + R_{ZX} &= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z + \nabla_Y\nabla_Z X - \nabla_Z\nabla_Y X - \nabla_{[Y,Z]}X \\
&+ \nabla_Z\nabla_X Y - \nabla_X\nabla_Z Y - \nabla_{[Z,X]}Y \\
&= \nabla_X(\nabla_Y Z - \nabla_Z Y) + \nabla_Y(\nabla_Z X - \nabla_X Z) + \nabla_Z(\nabla_X Y - \nabla_Y X) \\
&- \nabla_{[X,Y]}Z - \nabla_{[Y,Z]}X - \nabla_{[Z,X]}Y \\
&= \nabla_X[Y, Z] + \nabla_Y[Z, X] + \nabla_Z[X, Y] - \nabla_{[X,Y]}Z - \nabla_{[Y,Z]}X - \nabla_{[Z,X]}Y \\
&= (\nabla_X[Y, Z] - \nabla_{[Y,Z]}X) + (\nabla_Y[Z, X] - \nabla_{[Z,X]}Y) + (\nabla_Z[X, Y] - \nabla_{[X,Y]}Z) \\
&\underset{\nabla \text{ symm.}}{=} [X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] \\
&\underset{\text{Jacobi id.}}{=} 0.
\end{aligned}$$

3. To prove this property it is actually sufficient to demonstrate that $\langle R_{XY}Z, Z\rangle = 0$ for all $X, Y, Z \in \mathfrak{X}(M)$. In fact, if this is true, then

$$0 = \langle R_{XY}(Z + W), Z + W\rangle = \cancelto{0}{\langle R_{XY}Z, Z\rangle} + \langle R_{XY}Z, W\rangle + \langle R_{XY}W, Z\rangle + \cancelto{0}{\langle R_{XY}W, W\rangle},$$

thus implying

$$\langle R_{XY}Z, W\rangle = -\langle Z, R_{XY}W\rangle .$$

To prove that $\langle R_{XY}Z, Z\rangle = 0$ we use the fact that the Levi-Civita connection is compatible with the metric, so

$$Y(\|Z\|^2) = Y(\langle Z, Z\rangle) = \langle\nabla_Y Z, Z\rangle + \langle Z, \nabla_Y Z\rangle = 2\langle\nabla_Y Z, Z\rangle,$$

hence

$$XY(\|Z\|^2) = 2X\langle\nabla_Y Z, Z\rangle = 2\langle\nabla_X\nabla_Y Z, Z\rangle + 2\langle\nabla_Y Z, \nabla_X Z\rangle .$$

With analogous computations we find

$$YX(\|Z\|^2) = 2\langle\nabla_Y\nabla_X Z, Z\rangle + 2\langle\nabla_X Z, \nabla_Y Z\rangle,$$

and

$$[X, Y](\|Z\|^2) = 2\langle\nabla_{[X,Y]}, Z\rangle .$$

As a consequence,

$$\begin{aligned}
(XY - YZ - [X, Y])(\|Z\|^2) &= 2\langle\nabla_X\nabla_Y Z, Z\rangle - 2\langle\nabla_Y\nabla_X Z, Z\rangle - 2\langle\nabla_{[X,Y]}, Z\rangle \\
&= 2\langle\nabla_X\nabla_Y - \nabla_Y\nabla_X - \nabla_{[X,Y]}Z, Z\rangle = 2\langle R_{XY}Z, Z\rangle \\
&= 0.
\end{aligned}$$

4. To prove this property we must use the first Bianchi identity, cyclically permute the first argument of the inner product. We get:

$$\langle R_{XY}Z, W\rangle + \langle R_{YZ}X, W\rangle + \langle R_{ZX}Y, W\rangle = 0,$$

$$\langle R_{YZ}W, X\rangle + \langle R_{ZW}Y, X\rangle + \langle R_{WY}Z, X\rangle = 0,$$
$$\langle R_{ZW}X, Y\rangle + \langle R_{WX}Z, Y\rangle + \langle R_{XZ}W, Y\rangle = 0,$$
$$\langle R_{WX}Y, Z\rangle + \langle R_{XY}W, Z\rangle + \langle R_{YW}X, Z\rangle = 0.$$

Thanks to the property just proven in item 3 and the symmetry of the inner product, if we sum the last four equations, the terms indicated with the same color cancel each other and we remain with

$$\langle R_{ZX}Y, W\rangle + \langle R_{WY}Z, X\rangle + \langle R_{XZ}W, Y\rangle + \langle R_{YW}X, Z\rangle = 0,$$

now, using the anti-symmetry of R, we can rearrange the previous formula as

$$-\langle R_{XZ}Y, W\rangle - \langle R_{YW}Z, X\rangle + \langle R_{XZ}W, Y\rangle + \langle R_{YW}X, Z\rangle = 0,$$

and, using one last time the property in item 3. we can conveniently re-write the first two terms as follows

$$\langle R_{XZ}W, Y\rangle + \langle R_{YW}X, Z\rangle + \langle R_{XZ}W, Y\rangle + \langle R_{YW}X, Z\rangle = 0,$$

i.e.

$$2\langle R_{XZ}W, Y\rangle + 2\langle R_{YW}X, Z\rangle = 0,$$

so

$$\langle R_{XZ}W, Y\rangle = -\langle R_{YW}X, Z\rangle = \langle R_{WY}X, Z\rangle.$$

□

It is quite direct to verify that, in terms of the coefficients R_{ijhk}, the properties expressed by the previous theorem can be written as follows.

1. $R_{ijhk} = -R_{jihk}$: anti-symmetry w.r.t. the switch of the first two indices.

2. $R_{ijhk} + R_{jhki} + R_{hkij} = 0$: first Bianchi identity.

3. $R_{ijhk} = R_{ijkh}$: anti-symmetry w.r.t. the switch of the last two indices.

4. $R_{ijhk} = R_{hkij}$: symmetry w.r.t. the exchange of the first and last couples of indices.

The number of components of the Riemann curvature tensor (R_{ijhk}) is n^4 for a manifold of dimension n, because there are 4 indices that run, independently, from 1 to n. However, due to symmetry, anti-symmetry and the Bianchi identity, the number of independent components of R can be proven to be reduced from n^4 to $n_{\text{ind}} = n^2(n-1)/12$. Let us analyze the first cases:

- $n = 1$: this is the case of a manifold that coincides with the range of a curve and n_{ind} in this case is 0, so *the Riemann curvature tensor of a curve does not exists.* This is explained by the fact that **a curve has no intrinsic curvature**, we can define an extrinsic concept of curvature for a curve, which depends on how it is embedded in $\mathbb{R}^n$. However the Riemann curvature tensor contains intrinsic information about the curvature of a manifold M, this is why, in the case of a 1D manifold it does not exist.

- $n = 2$: in this case $n_{\text{ind}} = 1$. This describes, e.g., the case of the sphere S^2.

- $n = 3$: in this case $n_{\text{ind}} = 6$, this is the case of the Galilean space.
- $n = 4$: in this case $n_{\text{ind}} = 20$, this is an important case because the spacetime of relativity has dimension 4.
- $n = 5$: in this case $n_{\text{ind}} = 50$, this shows how rapidly the number of independent components of the Riemann curvature tensor grows and how complicated the task of its analysis becomes for manifold of large dimension.

8.10.2 The Ricci tensor and the scalar curvature

By performing a sort of trace, it is possible to 'condensate' the information contained in the Riemann curvature tensor into a simpler one. To do that, it is easier to work in components and contract the first lower index with the (only) upper index, i.e.

$$R_{ij} := R_{hij}{}^{h}.$$

These coefficients are the components of a very important tensor field.

Def. 8.10.2 (Ricci tensor field and Ricci curvature) *The Ricci tensor field* Ric *is a tensor field of type* $\binom{0}{2}$ *whose expression in local coordinates is*

$$\text{Ric} = R_{ij} dx^i \otimes dx^j. \tag{8.20}$$

The Ricci curvature is the quadratic form associated to the Ricci tensor, i.e.

$$\widetilde{\text{Ric}}(X) = \text{Ric}(X, X), \qquad \forall X \in \mathfrak{X}(M).$$

Thanks to the properties of symmetry of R_{ijhk} reported at the end of the previous section, the Ricci tensor is symmetric, i.e.

$$R_{ij} = R_{ji},$$

in fact:

$$R_{ij} = R_{hijk} g^{kh} = R_{jkhi} g^{kh} = -R_{kjhi} g^{hk} = +R_{kjih} g^{hk} = R_{ji}.$$

By computing the trace of the Ricci tensor we obtain a scalar function of $\mathscr{C}^\infty(M)$.

Def. 8.10.3 (Scalar curvature) *Given a Riemannian manifold* (M, g) *with Riemann curvature tensor* R*, the scalar curvature* R *of* M *is defined by*

$$R = g^{ij} R_{ij}.$$

R contains less information about the curvature properties of M w.r.t. the Riemann curvature tensor.

Chapter 9

Principal fiber bundles and applications to field theory

Inspirational epithap wanted...

...

Fibre bundles play a major role in modern theoretical physics, whereas it is in general relativity or in the standard model of particle physics. This chapter will discuss largely fibre bundles and connections. In the first part of this chapter, we will define the notion of fibre bundle, beginning with the general fibre bundles and moving into the specific case of principal bundle where the notions of Lie groups defined in the previous chapter will play a central role. Then we will discuss the specific case (but important) of associated vector bundles. The second part will treat the notion of connections and covariant derivatives that are necessaries tools in gauge theories such as Yang-Mills theory.

9.1 Fibre Bundles

We already have encountered vector bundles, namely the tangent bundle TM and the cotangent bundle T^*M of a differential manifold M. For example, for the tangent space TM there was a natural projection $\pi : TM \to M$ that associate to each vector the point p in M at which it is tangent. The inverse image of any point p of M under π (called the fibre over p) was nothing more that the tangent space T_pM and vector fields could be defined as smooth cross-section of TM. We will generalize these notions in this section, going from the general definition of bundles to the specific case of vector bundles associated to principal bundles.

9.1.1 First definitions

First, let's give the proper definition of a bundle.

Def. 9.1.1 (Bundle) *A C^∞-bundle is the data of a surjective projection $\pi : E \to M$, where E and M are smooth manifolds and π is a C^∞-map. E is called the total space, M the base space and for every $p \in M$, $F_p := \pi^{-1}(\{p\})$ is called the fibre over p.*

In the following, we will sometimes denote a bundle by a greek letter like ξ. In this case, $E(\xi)$ will be the total space of the bundle and $M(\xi)$ will be its base space.

In the bundles we will treat, every fibre over $p \in M$ will be diffeomorphic to the same space F, in which case, we will talk about fibre bundle and F will be called the fibre of the bundle. This motivates the more specific definition.

Def. 9.1.2 (Fibre bundle) *Let F be a smooth manifold. The bundle $\pi : E \to M$ is said to be a fibre bundle if, for each $p \in M$, there is an open neighborhood $U \subset M$ and a diffeomorphism $h : U \times F \to E|_U := \pi^{-1}(U)$, called local trivialization of E such that we have the commutation*

$$\begin{tikzcd} E|_U \arrow[d, "\pi"] & U \times F \arrow[l, "h"'] \arrow[dl, "pr_1"] \\ U & \end{tikzcd}$$

where pr_1 is the projection on U, or say differently $\pi(h(x, y)) = x$, for all $x \in U$ and $y \in F$.

A collection $\{(U_i, h_i)\}_i$ of local trivialization such that the open $\{U_i\}_i$ are covering M is called an atlas of the bundle. We have that if $U_i \cap U_j \neq \emptyset$, then for p in this intersection and $f \in F$, we can defined diffeomorphisms $\psi_{ij}(p) : F \to F$, called the bundle transition functions, such that $h_i^{-1} \circ h_j(p, f) = (p, \psi_{ij}(p)(f))$. These maps satisfy

1. $\psi_{ii}(p) = id_F$
2. $\psi_{ij}(p) = (\psi_{ji}(p))^{-1}$
3. $\psi_{ij}(p) \circ \psi_{jk}(p) = \psi_{ik}(p)$ for all U_i, U_j, U_k such that $U_i \cap U_j \cap U_k \neq \emptyset$.

We recognize what we had for TM where, to each $p \in M$, we had a local coordinate chart (U, ϕ) and we could define the map

$$\begin{array}{rcl} TM|_U & \longrightarrow & \phi(U) \times \mathbb{R}^m \\ v & \longmapsto & (x^1, \cdots x^p, v(x^1), \cdots, v(x^p)) \end{array}$$

The fact that the fibre of TM is $\mathbb{R}^m$, a vector space, tells us that we are in a special fibre bundle : a vector bundle.

Def. 9.1.3 (Vector bundle) *A vector bundle of rank k of base manifold M is a fibre bundle where the fibre is $\mathbb{R}^k$. More specifically, to each $p \in M$, $F_p = \pi^{-1}(\{p\})$ is a vector space of size k and there is an open neighborhood $U \subset M$ of p on which we have the local trivialization h that satisfies $\pi \circ h = pr_1$*

$$\begin{tikzcd} E|_U \arrow[d, "\pi"] & U \times \mathbb{R}^k \arrow[l, "h"'] \arrow[dl, "pr_1"] \\ U & \end{tikzcd}$$

We have in particular that, for each $p \in M$, by the local trivialization, the map $h_p : \{p\} \times \mathbb{R}^k \to \pi^{-1}(\{p\})$ is an isomorphism of vector spaces.

Now let's give some particular example of fibre bundle.

1. One of the most known fibre bundle is the Möbius strip where the base space is the circle S^1, and the fibre can be seen as a closed interval of $\mathbb{R}$. The total space can be regarded as a rectangle where the edge must be identified but in identifying opposite vertices. We can construct in the same spirit the Klein bottle.

   PUT THE FIGURE HERE

2. These two examples are in the case that the total space is "twisted" in some sense. A more simpler example is just considering the total space as the product of the base space with the fibre, and the projection map is just the projection on the base space, i.e. the bundle $pr_1 : M \times F \to M$.

3. If G is a Lie group and H is a Lie subgroup of G, then the bundle $\pi : G \to G/H$, where, $\forall\, g \in G$, we have $\pi(g) := gH$ is a fibre bundle with fibre H.

It is sometimes useful to see a bundle as a subspace of a bundle of reference.

Def. 9.1.4 (Sub-bundle) *We say that a bundle $\tilde{\pi} : \tilde{E} \to \tilde{M}$ is a sub-bundle of a bundle $\pi : E \to M$ if we have $\tilde{E} \subset E$, $\tilde{M} \subset M$ and if $\tilde{\pi}$ is the restriction of π to $\tilde{E}$.*

Now, as vector fields can be seen as cross-section of the tangent bundle TM, let's give the proper definition of a cross-section :

Def. 9.1.5 (Cross-section) *Let $\pi : E \to M$ be a bundle. A cross-section of the bundle is a map $\sigma : M \to E$ such that*

$$\begin{tikzcd} E \arrow[d, shift left, "\pi"] \\ M \arrow[u, shift left, "\sigma"] \end{tikzcd} \qquad \pi \circ \sigma = id_M$$

i.e. that for each point $p \in M$, its image $\sigma(p)$ is in the fibre $F_p = \pi^{-1}(\{p\})$

We can note that in the specific case to a product bundle $\pi : M \times F \to M$, by construction a cross-section σ gives rise to a unique function $\tilde{\sigma} : M \to F$ such that $\forall\, p \in M$, $\sigma(p) = (p, \tilde{\sigma}(p))$.

To end this part, let us give the definition of a bundle map :

Def. 9.1.6 (Bundle map) *Let $\pi_E : E \to M$ and $\pi_{E'} : E' \to N$ two bundles. A homomorphism of bundle is a pair of smooth maps (u, f) with $u : E \to E'$ and $f : M \to N$ such that we have the commutative diagram*

$$\begin{tikzcd} E \arrow[r, "u"] \arrow[d, "\pi_E"] & E' \arrow[d, "\pi_{E'}"] \\ M \arrow[r, "f"] & N \end{tikzcd}$$

i.e. that we have $\pi_{E'} \circ u = f \circ \pi_E$.

In the case of vector bundles, we require in addition that the restriction of u *on the fibres to be linear, i.e. that for each* $p \in M$, $u_p : \pi_E^{-1}(\{p\}) \to \pi_{E'}^{-1}(\{f(p)\})$ *is an homomorphism of vector space.*

We can remark that the commutation $\pi_{E'} \circ u = f \circ \pi_E$ tells us that for all $p \in M$, $u(\pi_E^{-1}(\{p\})) \subset \pi_{E'}^{-1}(\{f(p)\})$ i.e. that the bundle maps sends fibers into fibers.

Now that we have defined bundle maps, a question arise whether it is possible or not to define the pull-back of a bundle. This is given by the following

Def. 9.1.7 (Pull-back) *Let* $\pi : E \to M$ *be a fibre bundle that we will denote by* β *and let* $f : M' \to M$ *be a map, where* M' *is another manifold. We define the pull-back of* β *to be the bundle* $\pi' : E' \to M'$, *denoted by* $f^*(\beta)$, *where*

1. M' *is the base space*
2. $E' := \{(x', e) \in M' \times E / f(x') = \pi(e)\}$
3. $\forall\, (x', e) \in E'$, $\pi'(x', e) = x'$

This gives rise to a bundle map (f_β, f) between the bundle $f^*(\beta)$ and β, where, for all $(x', e) \in E'$, $f_\beta(x', e) = e$. We can note that each fibre of $f^*(\beta)$ is diffeomorphic to the fibre of β so $f^*(\beta)$ is a fibre bundle of fibre F.

9.1.2 Principal bundles

There are special fibre bundles where the fibre is a Lie group G. These bundles have the particularity that we can associate to them, in a way that is to define, general bundles. But first, let's define what is a G-bundle.

Def. 9.1.8 (G-bundle) *Let* G *be a Lie group. We say that* $\pi : E \to M$ *is a* G*-bundle if* G *has a right action on* E *and if* $\pi : E \to M$ *is isomorphic to the bundle* $\rho : E \to E/G$ *where* E/G *is the space of the orbit given by the action of* G *on* E *and* ρ *is the canonical projection on the space of orbits.*

A principal bundle is thus a particular G-bundle in the following sense :

Def. 9.1.9 (Principal bundle) *A principal* G*-bundle is a* G*-bundle where the action of* G *on* E *is free.*

For the rest of this chapter, to emphasize that we have a principal map, we will denote the total space by P instead of E.

Note that in a principal G-bundle $\pi : P \to M$, we have a fibre bundle with fibre G. Indeed, if $x \in M$, and $p \in \pi^{-1}(x)$, $\pi^{-1}(x)$ is the orbit of p under the action of G. By the freedom of its action and by theorem **??**, we get that $\pi^{-1}(x)$ is isomorphic to G.

Now let's give a simple example of principal bundle.

If we consider the product bundle $pr_1 : M \times G \to M$, where the right action of G is simply the right multiplication : $\forall\, p \in M$ and $\forall\, x_0 \in G$, $(p, g_0)g := (p, g_0 g)$. This bundle is called the trivial principal bundle.

We would like to define principal bundle map as bundle maps that would preserve the group action. This is satisfied by requiring the map to be equivariant

Def. 9.1.10 (Principal bundle map) *Let $\pi : P \to M$ and $\tilde{\pi} : \tilde{P} \to \tilde{M}$ be two principal G-bundles and let (u, f) be a bundle map. Then (u, f) is said to be a principal bundle map if $u : P \to \tilde{P}$ is G-equivariant as stated in definition* **??**, *i.e. we have, for all $p \in P$ and $g \in G$.*

$$u(pg) = u(p)g \tag{9.1}$$

*As for equation (**??**) we can generalize this in the case where $\pi : P \to M$ is a principal G-bundle, $\tilde{\pi} : \tilde{P} \to \tilde{M}$ a principal $\tilde{G}$-bundle and $\rho : G \to \tilde{G}$ a group homomorphism. Then the bundle map (u, f) is a principal bundle map if we have, for all $p \in P$ and $g \in G$*

$$u(pg) = u(p)\rho(g) \tag{9.2}$$

There is a particular case where (u, id_M) is a principal map between a pair of principal G-bundle $\pi : P \to M$ and $\tilde{\pi} : \tilde{P} \to M$. Then in this case, u is an isomorphism.

By these principal bundle map, we can define a trivial principal G-bundle.

Def. 9.1.11 (Trivial principal bundle) *A principal G-bundle $\pi : P \to M$ is trivial if there is a principal bundle map from $\pi : P \to M$ to the product bundle $pr_1 : M \times G \to M$.*

There is a special characterization of trivial principal G-bundle when looking at cross-section. Mainly

Theorem 9.1.1 *A principal G-bundle is trivial if, and only if, it possesses a continuous cross-section.*

Proof. To do... □

9.1.3 Associated vector bundles

In this last part of this section, we will see how to associate a general bundle to a principal G-bundle by extended the action of the group on another manifold. First, we will define the G-product.

Def. 9.1.12 (G-product) *Let G be a Lie group and let X and Y two spaces on which G has a right-action given respectively by*

$$\begin{array}{rccl} \rho: & X \times G & \longrightarrow & X \\ & (x,g) & \longmapsto & \rho(x,g) = \rho_g(x) := xg \end{array}, \qquad \begin{array}{rccl} \theta: & Y \times G & \longrightarrow & Y \\ & (y,g) & \longmapsto & \theta(y,g) = \theta_g(y) := yg. \end{array}$$

Then, we can define the right action of G on the product space $X \times Y$ by the map

$$\begin{array}{rccc} \Theta_g : & X \times Y & \longrightarrow & X \times Y \\ & (x, y) & \longmapsto & \Theta_g(x, y) := (\rho_g(x), \theta_g(y)) = (xg, yg) \end{array}, \quad \forall\, g \in G$$

The G-product of X and Y is then the quotient of the product $X \times Y$ on the space of orbits of the action of Θ, i.e. that two elements (x, y) and (x', y') belongs to the same equivalence class if there exists $g \in G$ such that $x' = xg$ and $y' = yg$. We denote the G-product by $X \times_G Y$ and the equivalence class of $(x, y) \in X \times Y$ is written $[x, y]$.

In the case one of the space is G itself, then it can be shown that there is an diffeomorphism between $G \times_G Y$ and Y.

Now we have the key ingredients to define associated bundles.

Def. 9.1.13 (Associated bundle) *Let $\pi : P \to M$ a principal G-bundle and F a smooth manifold on which G acts on the left. We define its associated bundle through the action of G on F by the fibre bundle $\pi_F : P_F \to M$ with fibre F where*

- *$P_F := P \times_G F$ where the right action on this space is defined by*
$$(p, v)g := (pg, g^{-1}v) \tag{9.3}$$
- *π_F is defined by*
$$\pi_F([p, v]) = \pi(p) \tag{9.4}$$

We need to check that this bundle defined in this way is indeed a fibre bundle.

First, let's notice that π_F is well defined. If we take another representent $[p', v']$ of $[p, v]$, we have that there exists $g \in G$ such that $(p', v') = (pg, g^{-1}v)$ hence

$$\pi_F([p', v']) = \pi(p') = \pi(pg) = \pi(p) = \pi_F([p, v])$$

because p and pg belongs to the same orbit hence to the same fibre.

To see that $\pi_F : P_F \to M$ is indeed a fibre bundle. We need to find a local trivialization of P_F. Let's consider the atlas of the principal bundle $\pi : P \to M$ given by $\{(U_i, h_i\})_i$. Then, for every $x \in M$, there exists an open U such that we have the diffeomorphism $h : U \times G \to \pi^{-1}(U)$. We have thus the identification $U \times G \cong \pi^{-1}(U)$. Now let's consider $\pi_F^{-1}(U)$, we have :

$$\pi_F^{-1}(U) = \pi^{-1}(U) \times_G F \cong (U \times G) \times_G F = (U \times G \times F)/G = U \times (G \times_G F)$$

But we have the further identification that $G \times_G F \cong F$. Indeed, there is a diffeomorphism between $G \times_G F$ and F given by the map

$$\begin{array}{rccc} \iota : & G \times_G F & \longrightarrow & F \\ & [g, v] & \longmapsto & vg^{-1} \end{array}$$

- this map is well defined since, given two different representative $[g_1, v_1] = [g_2, v_2]$, then there exists $g \in G$ such that $g_2 = g_1 g$ and $v_2 = v_1 g$. Therefore, $v_2 g_2^{-1} = v_1 g(g_1 g)^{-1} = v_1 g g^{-1} g_1^{-1} = v_1 g_1^{-1}$.

- it is injective since, if $\iota([g,v]) = \iota([g',v'])$ then $vg^{-1} = v'g'^{-1}$. It follows that, applying the element $g^{-1}g'$ to g and v : $[g,v] = [gg^{-1}g', vg^{-1}g'] = [g', v'g'^{-1}g'] = [g', v']$. Hence we get injectivity

- the map is clearly surjective since, for all $v \in F$, we have $\iota([e,v]) = v$.

We finally get that $\pi_F^{-1}(U) \cong U \times F$ so the open covering of M defines also a local trivialization of P_F and the fibre at $x \in M$, $\pi_F^{-1}(x)$, is diffeomorphic to F. Hence $\pi_F : P_F \to M$ is a fibre bundle.

As we did for fibre bundle in general and principal bundle, let's define what is an associated bundle map.

Def. 9.1.14 (Associated bundle map) *Let $\pi : P \to M$ and $\tilde{\pi} : \tilde{P} \to \tilde{M}$ two principal G-bundle with associated bundle $\pi_F : P \times_G F \to M$ and $\tilde{\pi}_F : \tilde{P} \times_G F \to \tilde{M}$ respectively and let (u, f) be a principal bundle map. An associated bundle map (u_F, f) between the pair of associated bundle is defined by*

$$u_F([p,v]) := [u(p), v] \tag{9.5}$$

This is well-defined since

$$u_F([pg, g^{-1}v]) = [u(pg), g^{-1}v] = [u(p)g, g^{-1}v] = [u(p), v] = u_F([p,v])$$

because u is equivariant as a principal bundle map. And it is a bundle map because we have, for all $[p,v] \in P \times_G F$

$$f \circ \pi_F([p,v]) = f \circ \pi(p)$$
$$\tilde{\pi}_F \circ u_F([p,v]) = \tilde{\pi}_F([u(p), v]) = \tilde{\pi}(u(p))$$

and $f \circ \pi(p) = \tilde{\pi}(u(p))$ since (u, f) is a bundle map.

Finally, we will get interest in the particular case of associated vector bundle where we replace the space F by a vector space V and requiring that the action of G on V is linear. Hence we require that the action

$$\begin{array}{rccc} \theta : & G & \longrightarrow & GL(V) \\ & g & \longmapsto & \theta_g \end{array}$$

is a representation of G in V.

More formally, if $\pi : P \to M$ is a principal G-bundle and V is a vector of dimension n on which G acts linearly, then the associated bundle $\pi_V : P \times_G V \to M$ can be given the structure of an n-dimensional real vector bundle. Indeed, if $x \in M$, let $p \in \pi^{-1}(\{x\})$ and define the homeomorphism

$$\begin{array}{rccc} \iota_p : & V & \longrightarrow & \pi_V^{-1}(\{x\}) \\ & v & \longmapsto & \iota_p(v) := [p,v] \end{array} \tag{9.6}$$

Then we define the operations

1. $\iota_p(v_1) + \iota_p(v_2) := \iota_p(v_1 + v_2), \quad \forall\, v_1, v_2 \in V$
2. $\lambda \iota_p(v) := \iota_p(\lambda v), \quad \forall \lambda \in \mathbb{R}, \forall\, v \in V$

This is well defined thanks to the linearity of the action of G on V, since if we take another element $p' \in \pi^{-1}(\{x\})$ such that $\iota'_p(v') = \iota_p(v)$, then we get

$$\begin{aligned}
[p', v'_1] + [p', v'_2] &= [p', v'_1 + v'_2] \\
&= [pg, g^{-1}v_1 + g^{-1}v_2] \\
&= [pg, g^{-1}(v_1 + v_2)] \\
&= [p, v_1 + v_2] \\
&= [p, v_1] + [p, v_2]
\end{aligned}$$

And let's end this section with an important example of principal bundle associated to a vector bundle : the bundle of frames.

Let $\pi : E \to M$ be a vector bundle of rank n. A frame at a point $p \in M$ is an ordered set of basis vectors for the vector space E_p. If we define by $\mathcal{F}_p$ the set of all frames at the point p, the bundle of frames $\mathcal{F}(E)$ is defined to be the disjoint union of all such spaces i.e. that a point in $\mathcal{F}(E)$ is a pair (p, b) where $p \in M$ and $b \in \mathcal{F}_p$ and the projection map $\pi_{\mathcal{F}} : \mathcal{F}(E) \to M$ is the function that takes a frame into the point in M to which it is attached.

In fact, since for each $p \in M$, there is an isomorphism between E_p and $\mathbb{R}^n$, a frame can be seen as a linear isomorphism. Indeed, let $\mathcal{B}$ be the canonical basis for $\mathbb{R}^n$, a frame b of E_p is uniquely determined by the image vectors of the vectors of $\mathcal{B}$ through a suitable isomorphism $\lambda : \mathbb{R}^n \to E_p$ which represents the base change.

We can define a natural free action of $GL(n, \mathbb{R})$ on $\mathcal{F}(E)$. If $\lambda \in \mathcal{F}_p$ represented by a matrix Λ and if α is an automorphism of $\mathbb{R}^n$ represented by a matrix A then we define the right action of $GL(n, \mathbb{R})$ on the fibres $\mathcal{F}_p$ by

$$\begin{array}{rrcl}
\theta : & \mathcal{F}_p \times GL(n, \mathbb{R}) & \longrightarrow & \mathcal{F}_p \\
& (\lambda, \alpha) & \longmapsto & \lambda \circ \alpha = \Lambda A
\end{array}$$

This action is transitive by unicity of base change and can be extended to the bundle $\mathcal{F}(E)$ by

$$\begin{array}{rrcl}
\Theta : & \mathcal{F}(E) \times GL(n, \mathbb{R}) & \longmapsto & \mathcal{F}(E) \\
& ((p, \lambda), A) & \longmapsto & (p, \theta(\lambda, A))
\end{array}$$

We can therefore see that $\pi_{\mathcal{F}} : \mathcal{F}(E) \to M$ is a $GL(n, \mathbb{R})$-principal bundle.

Indeed, the right action is free and the orbits coincide with the fibers. It remains to see that we have a local trivialization. Let $\{(U_i, \varphi_i)\}_i$ be a local trivialization for E where we have the diffeomorphisms $\varphi_i : E|_{U_i} \to U_i \times \mathbb{R}^n$ with restrictions on the fibers $\varphi_{i,p} : E_p \to \{p\} \times \mathbb{R}^n$. Then we can define the maps :

$$\begin{array}{rrcl}
\psi_i : & \mathcal{F}(E)|_{U_i} = \tilde{\pi}^{-1}(U_i) & \longmapsto & U_i \times GL(n, \mathbb{R}) \\
& (p, \lambda) & \longmapsto & (p, \varphi_{i,p} \circ \lambda)
\end{array}$$

These maps are invertible and differentiable with inverse differentiable so they are diffeomorphisms and $\{(U_i, \psi_i)_i\}$ is a local trivialization for $\mathcal{F}(E)$.

Consider a point p of M and U its open neighborhood, and $q$an element of $\mathcal{F}(E)|_U$. We can write $q = (p, \lambda)$. Then we have $\psi(q) = (p, \varphi_p \circ \lambda) = (\pi(q), h(q))$ for a certain function $h : \mathcal{F} \to GL(n, \mathbb{R})$. It remains to see that this map is equivariant w.r.t. the right action of the group :

$$\begin{aligned}
h(\Theta(q, A)) &= h(\Theta((p, \lambda), A) = h(p, \theta(\lambda, A)) = h(p, \lambda \circ \alpha) \\
&= (p, \varphi_p \circ \lambda \circ \alpha) = (p, \theta(\varphi_p \circ \lambda, A)) = \Theta((p, \varphi_p \circ \lambda), A) \\
&= \Theta(h(p, \lambda), A) = \Theta(h(q), A)
\end{aligned}$$

By local trivialization, we have that $\mathcal{F}/GL(n, \mathbb{R}) \cong M$ and each fibre is diffeormorphic to $GL(n, \mathbb{R})$. Hence $\tilde{\pi} : \mathcal{F}(E) \to M$ is a principal fibre of structure group $GL(n, \mathbb{R})$.

A special bundle of frames is the tangent frame bundle (called also the frame bundle of the manifold M) where the vector bundle in consideration is the tangent bundle. In this case, a local section is called a smooth local frame. One important example is that given a local coordinate chart $(U, \varphi = (x^1, \ldots, x^m))$ around a point $p \in M$, we have a basis of T_pM given by $(\partial_1|_p, \ldots, \partial_m|_p)$ so we can define a local section of TM by

$$\begin{array}{rccl}
\partial_i : & U & \longrightarrow & TM \\
& p & \longmapsto & \partial_i(p) := \partial_i|_p
\end{array}$$

The same can be done for the cotangent bundle.

9.2 Connection and parallel transport

The purpose of connections is to compare points belonging to different fibres in a way that is independent of a local trivialization. Hence, we are looking for vector fields that go for one fibre to another. In this section, we will define a connection in two ways, one as a collection of tangent spaces, the other as a differential one-form. We will first give the definition for general bundles then restrict ourselves to principal and associated bundle to conclude with parallel transport and curvature.

9.2.1 Connection of Ehresmann

Connection in a general bundle

First, let give the definition of a vertical subspace.

Def. 9.2.1 (Vertical subspace) *Let $\pi : E \to M$ be a bundle, and let $e \in E$. The vertical subspace V_eE of the tangent space T_eE is defined to be the kernel of the push-forward of π at e i.e.*

$$V_eE := \ker(\pi_*) = \{v \in T_eE, \pi_*(v) = 0\} \tag{9.7}$$

The elements of V_eE are called vertical. As $e \in E$ changes, these subspaces form a C^∞-subbundle VE of the bundle TE.

If E is an n-manifold and M a m-manifold, since π is a projective surjection, it is of constant rank m (i.e. that for all $e \in E$, $\pi_* : T_eE \to T_{\pi(e)}M$ is of rank m), then we have $\dim V_eE = n - m$.

Remark that V_eE can then be seen as the tangent space to the fiber $\pi^{-1}(\{\pi(e)\})$.

Now, as mentioned above, we want to look at vector fields that points away to the fibres, not tangent to it. This motivates the following definition :

Def. 9.2.2 (Ehresmann connection) *A general connection (or Ehresmann connection) on the bundle $\pi : E \to M$ is a smooth assignment to each point $e \in E$ of a vector subspace H_eE of T_eE such that*

$$T_eE = V_eE \oplus H_eE \tag{9.8}$$

The subbundle HE of TE associated to it is called the horizontal subbundle of TE and elements of H_eE are called horizontal.

Hence, this definition means that each vector $w \in T_eE$ can be written in a unique way as $w = v + h$ where $v \in V_eE$ and $h \in H_eE$. To emphasize this, we will sometimes write by $v = \mathrm{ver}(w)$ and $h = \mathrm{hor}(w)$ for respectively the vertical and horizontal components of w.

Also, by the definition of the vertical subspace V_eE of T_eE, we get that the restriction of π_* to H_eE is an isomorphism of vector space and we have $\dim H_eE = \dim T_{\pi(e)}M = m$.

We can give another definition of the connection in terms of a differential one-form. More precisely, we define a connection 1-form on a bundle $\pi : E \to M$ as a linear map

$$\Phi : TE \to VE \tag{9.9}$$

that satisfies

1. $\Phi \circ \Phi = \Phi$ (idempotent)
2. $\mathrm{Im}\Phi = VE$ (surjectivity)

This definition means that we can see the connection one-form as a projection of TE onto VE. In particular, if $e \in E$, then $\Phi_e : T_eE \to V_eE$ is the projection of T_eE into V_eE. The horizontal sub-bundle is thus defined by $HE := \ker \Phi$. These two different definitions are in fact equivalent and the decomposition (9.8) say that for all $e \in E$, Φ_e is simply the projection of T_eE on V_eE parallel to H_eE.

We can remark that, using the notation of ver and hor, for all $w \in T_eE$, we have $\Phi_e(w) = \mathrm{ver}(w)$.

Connection in a principal bundle

These definitions of vertical subspace and connection need to be slightly adapted in the framework of principal bundle to guarantee the action of the Lie group.

Since a principal bundle is a bundle, we have the same definition for the vertical subspace :

Def. 9.2.3 (Principal vectical subsapce) *Let $\pi : P \to M$ be a principal bundle of structure group G. For each $p \in P$, the vertical subspace V_pP of T_pP is defined to be the kernel of the linear push-forward π_* at p*

$$V_pP := \ker \pi_* = \{\tau \in T_pP / \pi_*\tau = 0\}$$

The particularity here is that V_pP can be identified with the Lie algebra $\mathfrak{g}$ of G. Let's introduced the necessary tools to show this.

Let $\Delta : P \times G \to P$ the usual right action of G on P given by $\Delta(p, g) = \delta_g(p) = pg$ where $\delta_g : P \to P$ is the usual diffeomorphism define in the previous chapter. Then, taking $\xi \in \mathfrak{g}$, we can define a curve in G locally : $t \mapsto \exp(t\xi) \in G$, for $t \in \mathbb{R}$ small enough, that passes though e the neutral element of G at $t = 0$ and tangent to ξ. Now letting this curve acts on $p \in P$, we obtain a curve on P given by $t \mapsto \delta_{\exp(t\xi)}(p) = p\exp(t\xi)$. This curve passes through p at $t = 0$ and so it is tangent to a vector of T_pP. This thus associates to each vector of $\mathfrak{g}$ a vector in T_pP by the map

$$\begin{array}{rccl} u_p : & \mathfrak{g} & \longrightarrow & T_pP \\ & \xi & \longmapsto & u_p(\xi) := \left.\dfrac{\mathrm{d}}{\mathrm{d}t} p\exp(t\xi)\right|_{t=0} \end{array} \tag{9.10}$$

By varying $p \in P$, we can define a map that associate to each vector $\xi \in \mathfrak{g}$ a vector field on P denoted by X^ξ whose value at $p \in P$ is given by $X^\xi_p := u_p(\xi)$.

$$\begin{array}{rccl} u : & \mathfrak{g} & \longrightarrow & \mathfrak{X}(P) \\ & \xi & \longmapsto & X^\xi \end{array} \tag{9.11}$$

In other words, we associate to the vector fields on G, whose integral curve is $\sigma_\xi : t \mapsto \exp(t\xi)$, the vector field on P whose integral curve is given by $t \mapsto p\exp(t\xi)$.

As $\mathfrak{g}$ and $\mathfrak{X}(P)$ are Lie algebras, we can notice that this u is a morphism of Lie algebras since it satisfies, for all $\xi, \eta \in \mathfrak{g}$

$$X^{[\xi,\eta]} = u([\xi, \eta]) = [u(\xi), u(\eta)] = [X^\xi, X^\eta]$$

We can thus show the identification mentioned above :

Theorem 9.2.1 *Let $\pi : P \to M$ be a principal bundle of group structure G and let $\mathfrak{g}$ be the Lie algebra of G. Then the map u_p defined above is an isomorphism between $\mathfrak{g}$ and V_pP.*

Proof. First notice that u_p is linear since for each $\xi \in \mathfrak{g}$, $u_p(\xi)$ is a derivative. Now let's prove injectivity and surjectivity.

- For injectivity, suppose that $\xi \in \ker(u_p)$, $\xi \neq 0$. Since the action of G on P is free, the only element that fixes the point p is the neutral element e. Hence, for $t \neq 0$, we will have $p\exp(t\xi) \neq p$ so the curve on P induced by $t \mapsto \exp(t\xi)$ is not constantly equal to p and the vector $u_p(\xi)$ tangent to this curve in p at $t = 0$ will be non-zero. This contradicts the fact that $\xi \in \ker(u_p)$ so we get that $\xi = 0$ and u_p is injective.

- The surjectivity follows immediately. Indeed, by a local trivialization of the principal bundle $\pi : P \to M$, $p \in P$ will correspond to a point $(x, g) \in M \times G$, hence the fibre $\pi^{-1}(\{x\})$ is diffeomorphic to G. By the dimension analysis seen for general bundle, we get that $\dim V_pP = \dim P - \dim M = \dim G$. Hence the restriction of u_p to the codomain V_pP, since u_p is linear, injective and $\dim V_pP = \dim G = \dim \mathfrak{g}$ guarantees that u_p is surjective.

□

Now, let's define the notion of a connection in the special case of a principal bundle.

Def. 9.2.4 (Connection in a principal bundle) *Let $\pi : P \to M$ a principal bundle with group structure G. A connection is a smooth assignment to each point $p \in P$ of the total space of a subspace H_pP of T_pP such that*

1. *$T_pP = V_pP \oplus H_pP$*
2. *the subspace H_pP is invariant by the action of the group G i.e.*

$$\delta_{g*}(H_pP) = H_{\delta_g(p)}P, \text{ for all } p \in P \text{ and } g \in G. \tag{9.12}$$

Another way of defining a connection is as a one-form

Def. 9.2.5 (Connection one-form) *Let $u_p^{-1} : V_pP \to \mathfrak{g}$ being the inverse of the isomorphism between $\mathfrak{g}$ and V_pP. We defined the connection one-form $\omega : TP \to P \times \mathfrak{g}$ of the principal G-bundle $\pi : P \to M$ as a $\mathfrak{g}$ valued one-form defined by, $\forall\, p \in P$:*

$$\begin{array}{rccl} \omega_p : & T_pP & \longrightarrow & \mathfrak{g} \\ & \tau & \longmapsto & \omega_p(\tau) := u_p^{-1}(\Phi_p(\tau)) \end{array} \tag{9.13}$$

where $\Phi_p : T_pP \to V_pP$ is the projection of T_pP on V_pP parallel to H_pP associated to the differential one-form $\Phi : TP \to VP$ defined in the same way as in (9.9).

This one-form satisfies several properties of which

Proposition 9.2.1

1. *If X^ξ is the vector field induced by u on ξ then, for all $p \in P$*

$$\omega_p(X^\xi_p) = \xi \tag{9.14}$$

2. *For all $g \in G$, $p \in P$ and $\tau \in T_pP$, we have*

$$(\delta_g^*\omega)_p(\tau) = Ad_{g^{-1}}(\omega_p(\tau)) \tag{9.15}$$

3. *$h \in H_pP$ if, and only if, $\omega_p(h) = 0$.*

Proof.

1. Since, for all $p \in P$, $X_p^\xi = u_p(\xi) \in V_pP$ by theorem 9.2.1, we have that

$$\omega_p(X_p^\xi) = u_p^{-1}(\mathrm{ver}(X_p^\xi)) = u_p^{-1}(X_p^\xi) = u_p^{-1}(u_p(\xi)) = \xi.$$

2. First, let's see the link between the action on V_pP and the action on $\mathfrak{g}$. Let $\xi \in \mathfrak{g}$ and let $g \in G$, we have

$$\delta_{g*}(u_p(\xi)) = \delta_{g*}\left(\frac{\mathrm{d}}{\mathrm{d}t} p\exp(t\xi)\right)\Bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\delta_g(p\exp(t\xi))\Bigg|_{t=0}$$

But we have

$$\begin{aligned}\delta_g(p\exp(t\xi)) &= p\exp(t\xi)g = pg(g^{-1}\exp(t\xi)g)\\ &= pg(C_{g^{-1}}(\exp(t\xi)) = pg\exp(t\mathrm{Ad}_{g^{-1}}(\xi))\end{aligned}$$

by equation (**??**). Hence

$$\delta_{g*}(u_p(\xi)) = \frac{\mathrm{d}}{\mathrm{d}t} pg\exp(t\mathrm{Ad}_{g^{-1}}(\xi))\Bigg|_{t=0} = u_{pg}(\mathrm{Ad}_{g^{-1}}(\xi)) \tag{9.16}$$

Furthermore, we have, for all $g \in G$, $p \in P$ and $\tau \in T_pP$

$$\begin{aligned}(\delta_g^*\omega)_p(\tau) &= \omega_{\delta_g(p)}(\delta_{g*}\tau)\\ &= u_{pg}^{-1}\circ\Phi_{pg}(\delta_{g*}\tau)\\ &= u_{pg}^{-1}\circ\delta_{g*}\circ\Phi_p(\tau)\end{aligned}$$

Because by (9.12) we have the commutation $\delta_{g*}\circ\Phi_p = \Phi_{pg}\circ\delta_{g*}$. Hence

$$\begin{aligned}(\delta_g^*\omega)_p(\tau) &= u_{pg}^{-1}\circ\delta_{g*}\circ u_p\circ u_p^{-1}\circ\Phi_p(\tau)\\ &= u_{pg}^{-1}\circ\delta_{g*}\circ u_p(\omega_p(\tau))\\ &= u_{pg}^{-1}\circ u_{pg}(\mathrm{Ad}_{g^{-1}}(\omega_p(\tau)))\\ &= \mathrm{Ad}_{g^{-1}}(\omega_p(\tau)).\end{aligned}$$

where in the penultima equality we use equation (9.16)

3. By definition, one of the implication is straightforward. The only if part is also direct because since u_p is an isomorphism between V_pP and $\mathfrak{g}$, if $\omega_p(h) = 0$ then this means that $\Phi(h) = 0$ hence $h \in H_pP$.

□

9.2.2 Covariant derivative and parallel transport

An important tool to compare points in is the use of parallel transport, that is a way to transport points from one fibre to another.

Parallel transport in a principal bundle

Let $\pi : P \to M$ be a principal G-bundle where G is a Lie group equipped with a connection ω. We have seen the basic definitions of vertical fields and horizontal fields. Vertical fields can be easily constructed but there is also a way to generate more explicitly horizontal fields from a given vector field : this is the operation of horizontal lifting. Indeed we have seen that for each $p \in P$, π_* is an isomorphism from H_pP to $T_{\pi(p)}M$. This yields to the following

Def. 9.2.6 (Horizontal lift) *Let X be a vector field on M. Then there exists a unique vector field, called the horizontal lift of X and that we denote by $X^\uparrow$ such that, for all $p \in P$*

1. $\pi_*(X_p^\uparrow) = X_{\pi(g)}$
2. $\Phi_p(X_p^\uparrow) = 0$

We can remark that since, for all $p \in P$, $X_p^\uparrow \in H_pP$, then we have that $\delta_{g*}(X_p^\uparrow) = X_{pg}^\uparrow$, i.e. the operation of horizontal lift is G-equivariant. In fact, this property and the point ii) of the definition guarantee that a vector field on P is the horizontal lifting of a vector field defined on M.

This operation of horizontal lifting satisfied several properties

1. $X^\uparrow + Y^\uparrow = (X+Y)^\uparrow$
2. $(fX)^\uparrow = f \circ \pi X^\uparrow$ for all $f \in C^\infty(M)$
3. $[X,Y]^\uparrow = \mathrm{hor}([X^\uparrow, Y^\uparrow])$

An analogue procedure of horizontal lifting can be applied to curves on the base manifold M.

Def. 9.2.7 (Horizontal lift of a curve) *Let $\sigma : [a,b] \to M$ be a smooth curve. We define the horizontal lifting of σ by the curve $\sigma^\uparrow : [a,b] \to P$ satisfying*

1. $\Phi_p([\sigma^\uparrow]) = 0$ *i.e that the curve is horizontal,*
2. $\pi(\sigma^\uparrow(t)) = \sigma(t)$ *for all $t \in [a,b]$.*

Here the last point is to satisfy the fact that applying the projection π to this curve on P will give us back the curve on M.

As for vector fields, we have seen that to each vector field on M we could associate a unique horizontal lift on P, we have also existence and uniqueness of horizontal lift of a curve

Theorem 9.2.2 *Let $\sigma : [a,b] \to M$ be a smooth curve on M. Then for all point $p \in \pi^{-1}(\{\sigma(a)\})$ of the fibre over $\sigma(a) \in M$, there exists a unique horizontal lift of σ such that $\sigma^\uparrow(a) = p$*

Proof. To do... □

This theorem shows us the spirit of our goal : we want to compare a point $p \in P$ of one fibre with another point $q \in P$ of another fibre. If we call $\pi(p) = m \in M$ and $\pi(q) = n \in M$, having a smooth curve on M, $\sigma : [a, b] \to M$ such that $\sigma(a) = m$ and $\sigma(b) = n$, then the horizontal lift of this curve $\sigma^{\uparrow}$ will allow us going from the fibre $\pi^{-1}(m)$ to the fibre $\pi^{-1}(n)$.

Def. 9.2.8 (Parallele translation) *Let $\sigma : [a, b] \to M$ be a smooth curve of M. We define the parallel translation along σ the map*

$$\begin{array}{rccc} \tau : & \pi^{-1}(\{\sigma(a)\}) & \longrightarrow & \pi^{-1}(\{\sigma(b)\}) \\ & p & \longmapsto & \sigma^{\uparrow}(b) \end{array} \tag{9.17}$$

where $\sigma^{\uparrow}$ is the horizontal lift of σ that passes through p at $t = a$.

Connection and covariant derivative in an associated vector bundle

These notions of connection and parallel transport can be extended in the case of associated vector bundles. Let's see how the definitions of vertical and horizontal subspace (hence of connection) can be induced from the ones on a principal bundle.

Def. 9.2.9 *Let $\pi : P \to M$ be a principal G-bundle equipped with a connection ω and let $\pi_F : P_F \to M$ be its associated bundle through the action of G on F. Let $[p, v] \in P_F = P \times_G F$, we define*

1. *the vertical subspace $V_{[p,v]}(P_F)$ of the tangent space $T_{[p,v]}(P_F)$ by*

$$T_{[p,v]}(P_F) := \ker \pi_{F*} = \{\tau \in T_{[p,v]}(P_F) / \pi_{F*}\tau = 0\} \tag{9.18}$$

2. *the horizontal subspace $H_{[p,v]}(P_F)$ of the tangent space $T_{[p,v]}(P_F)$ by*

$$H_{[p,v]}(P_F) := k_{v*}(H_p P) \tag{9.19}$$

where

$$\begin{array}{rccc} k_v : & P & \longrightarrow & P \times_G F \\ & p & \longmapsto & [p, v] \end{array} \tag{9.20}$$

The horizontal subspace is well defined. Indeed, we can notice that, for all $p \in P$, we have

$$k_{g^{-1}v} \circ \delta_g(p) = k_{g^{-1}v}(pg) = [pg, g^{-1}v] = [p, v] = k_v(p)$$

Hence $k_{g^{-1}v} \circ \delta_g = k_v$. Now if we choose $[p', v']$ another representative of the class $[p, v]$ (i.e. that there exists $g \in G$ such that $(p', v') = (pg, g^{-1}v)$) then, by the characterization of a horizontal subspace in the principal bundle, we have

$$\begin{aligned} k_{v'_*}(H_{p'}P) &= k_{g^{-1}v*}(H_{pg}P) \\ &= k_{g^{-1}v*}(\delta_{g*}(H_p P)) \\ &= (k_{g^{-1}v} \circ \delta_g)_*(H_p P) \\ &= k_{v*}(H_p P) \end{aligned}$$

Having a connection, we can also define the notions of lifting a curve and parallel transporting.

Let $\sigma : [a, b] \to M$ be a smooth curve on M and $[p, v]$ a point in the fibre $\pi_F^{-1}(\{\sigma(a)\})$. By theorem 9.2.2 there is a unique horizontal lift of σ on P such that $\sigma^{\uparrow}(a) = p$. We then define the horizontal lift of σ on $P \times_G F$ that passes to $[p, v]$ at $t = a$ to be the curve

$$\sigma_F^{\uparrow} := [\sigma^{\uparrow}, v] \tag{9.21}$$

Then the parallel translation along σ in the bundle $P \times_G F$ is simply the map

$$\begin{array}{rccl} \tau_F : & \pi_F^{-1}(\{\sigma(a)\}) & \longrightarrow & \pi_F^{-1}(\sigma(b)\}) \\ & [p, v] & \longmapsto & \sigma_F^{\uparrow}(b) = [\sigma^{\uparrow}(b), v] \end{array} \tag{9.22}$$

where $\sigma^{\uparrow}$ is such that $\sigma^{\uparrow}(a) = p$.

In the case of a vector bundle, this parallel transport allows to define a derivative of a cross-section in a way that is independent of any choice of a local trivialization : this is the covariant derivative.

Def. 9.2.10 (Covariant derivative) *Let $\pi : P \to M$ be a principal bundle with structure group G and let V be a vector space on which G acts. Consider a smooth curve $\sigma : [0, \varepsilon] \to M$ in M such that $\sigma(0) = x_0$ and let τ_V^t be the parallel translation map going from the fibre $\pi_V^{-1}(\{\sigma(t)\})$ to the fibre $\pi_V^{-1}(\{x_0\})$. Then if $\psi : M \to P \times_G V$ is a cross-section, we define the covariant derivative of ψ at x_0 in the direction σ by*

$$\nabla_\sigma \psi := \lim_{t \to 0} \left(\frac{\tau_V^t \psi(\alpha(t)) - \psi(x_0)}{t} \right) \tag{9.23}$$

Considering two curves σ_1 and σ_2 tangent at x_0, we have that $\nabla_{\sigma_1} \psi = \nabla_{\sigma_2} \psi$. Hence, we can extend the definition of the covariant derivative to tangent vectors in $T_{x_0} M$ by defining that if $v \in T_{x_0} M$ is a tangent vector and σ is one of these representative curve, then

$$\nabla_v \psi := \nabla_\sigma \psi$$

and going further, we define the covariant derivative along a vector field X on M by

$$(\nabla_X \psi)(x_0) := \nabla_{X_{x_0}} \psi$$

∇_X is a linear operator. It also satisfies the following properties

1. $\nabla_X (f\psi) = f \nabla_X \psi + X(f)\psi$
2. $\nabla_{fX+Y} \psi = f \nabla_X \psi + \nabla_Y \psi$

The covariant derivative of a cross-section gives a cross-section. By expressing it in local coordinates, we will make appear the Christoffel symbols.

Let $\pi : E \to M$ be a vector bundle of rank r and consider $(U, \varphi = x^1, \dots, x^m)$ a local coordinate chart such that there is a local trivialization $h : U \times \mathbb{R}^r \to E|_U$. We can define a local basis of r local section defined by

$$\begin{array}{rccl} e_k : & U & \longrightarrow & E|_U \\ & p & \longmapsto & h(p, (0, \dots, 0, 1, 0, \dots, 0)) \end{array}$$

Thus, for each $p \in U$, $(e_1(p), \dots, e_r(p))$ forms a basis for the fibre E_p.

Now we want to apply the covariant derivative to these sections along the vector fields formed by the standard local frame of TM $(\partial_1, \dots, \partial_m)$ given by the local chart. The new sections obtained can be expressed in the local basis $(e_k)_k$ so we define

$$\nabla_{\partial_j} e_h = \Gamma^k_{jh} e_k$$

where $\Gamma^k_{jh} \in C^\infty(U)$. These functions are called the connection coefficients or, in the case of the tangent bundle, they are known as the Christoffel symbols.

We can then express the covariant derivative in local coordinates : let $s \in \Gamma(U)$ a local section and $X \in \mathfrak{X}(M)$ a vector field. Then we have the decomposition $X = X^j \partial_j$ and $s = s^h e_h$. The covariant derivative of s along X is then

$$\begin{aligned} \nabla_X s = \nabla_X (s^h e_h) &= X(s^h) e_h + s^h \nabla_X e_h \\ &= X(s^h) e_h + s^h \nabla_{X^j \partial_j} e_h \\ &= X(s^h) e_h + s^h X^j \nabla_{\partial_j} e_h \\ &= X(s^h) e_h + \Gamma^k_{jh} s^h X^j e_k \\ &= X(s^k) e_k + \Gamma^k_{jh} s^h X^j e_k \\ &= (X(s^k) + \Gamma^k_{jh} s^h X^j) e_k \end{aligned}$$

Actually, there is another standard definition of these connections in terms of connection one-forms that we can write in a matrix. The connection one-form defined in this way is a matrix of one-forms $\omega = [\omega^k_k]$ where

$$\omega^k_j = \Gamma^k_{ij} \underset{\cdot}{\mathrm{x}}^i$$

are one-forms defined on the coordinate chart $(U, (x^1, \dots, x^m))$ which is associated to ∇ w.r.t. the local frame.

Curvature

Let $\pi : P \to M$ a principal G-bundle and let $HP = \{H_p P / p \in P\}$ its connection. We have already defined $\Phi : TP \to VP$ the vertical projection. In the same spirit, let's define $h : TP \to HP$ the horizontal projection, i.e. for all $p \in P$, $h_p : T_p P \to H_p P$ is the projection on $H_p P$ parallel to $V_p P$. By this, we can define the exterior covariant derivative of a form.

Def. 9.2.11 (Exterior covariant derivative) *If ω is a k-form on P, we define the exterior covariant derivative $D\omega$ to be horizontal $(k+1)$-form defined by*

$$D\omega(X_1, \dots, X_{k+1}) = \mathrm{d}\omega(hX_1, \dots, hX_{k+1}) \tag{9.24}$$

where $X_1, \dots, X_{k+1}$ are vector fields on P.

and the curvature

Def. 9.2.12 (Curvature) *Let ω be a connection one-form defined on the principal G-bundle $\pi : P \to M$. We define the curvature two-form Ω to be the exterior covariant derivative of the connection, i.e.*

$$\Omega = D\omega \tag{9.25}$$

We can give an explicit formula of this curvature two-form through the Cartan structure equation :

Theorem 9.2.3 (Cartan structure equation) *Let ω be a connection one-form and let $\Omega = d\omega$ be its curvature two-form. If X and Y are vector fields on P, we have, for all $p \in P$:*

$$\Omega_p(X_p, Y_p) = \mathrm{d}\omega_p(X_p, Y_p) + [\omega_p(X_p), \omega_p(Y_p)] \tag{9.26}$$

Proof. Since we have the direct sum $TP = VP \oplus HP$, and since we have linear functions in the equation, it suffices to show the equation in three simple case.

1. X and Y are horizontal. This case is the easiest because we have then that $\omega(X) = 0$ and $\omega(Y) = 0$. For the remaining term, we have by definition :

$$\Omega_p(X_p, Y_p) = D\omega(X_p, Y_p) = \mathrm{d}\omega(h_p(X_p), h_p(Y_p)) = \mathrm{d}\omega(X_p, Y_p)$$

   since X and Y are horizontal.

2. X and Y are vertical. By (9.11), there exist $\xi, \eta \in \mathfrak{g}$ such that $X = X^\xi$ and $Y = Y^\eta$. Now by equation (**??**) we have

$$\mathrm{d}\omega(X^\xi, Y^\eta) = X^\xi(\omega(Y^\eta)) - Y^\eta(\omega(X^\xi)) - \omega([X^\xi, Y^\eta])$$

   But the point iii) of proposition 9.2.1 guarantees that $\omega_p(X_p^\xi) = \xi$ and $\omega_p(Y_p^\eta) = \eta$ so they are constants and applying a vector fields annihilate them. For the second term, we have that $\omega([X^\xi, Y^\eta]) = \omega(X^{[\xi,\eta]}) = [\xi, \eta]$ and hence the right hand side of equation (9.26) vanishes. On the other hand, the left hand side is automatically 0 since X and Y are vertical.

3. X is horizontal and Y is vertical. Since X is horizontal, then $\omega(X) = 0$ so the commutator vanishes and $\Omega(X, Y) = 0$ because Y is vertical. It remains to show that $\mathrm{d}\omega(X, Y) = 0$. Now, doing the same procedure as in the previous case, there exists $\eta \in \mathfrak{g}$ such that $Y = X^\eta$ and we can write

$$\mathrm{d}\omega(X, X^\eta) = X(\omega(X^\eta)) - X^\eta(\omega(X)) - \omega([X, X^\eta]) = -\omega([X, X^\eta])$$

   because $\omega(X^\eta) = \eta$ is constant and $\omega(X) = 0$ because X is horizontal.

   But we have the Lie derivative $[X, X^\eta] = -\mathfrak{L}_{X^\eta} X = \lim_{t \to 0} \dfrac{\delta_{\exp(t\eta)*}(X) - X}{t}$. Hence if X is horizontal, the right action on it gives also a horizontal vector field, and so the difference by vector space structure. This proves that $[X, X^\eta]$ is horizontal and then $\omega([X, X^\eta]) = 0$

□

To finish this chapter, let's rewrite this formula in a local way.

If we denote by $\mathcal{B} = \{v_1, \cdots, v_n\}$ a basis of the Lie algebra $\mathfrak{g}$. Then equation (9.26) can be rewritten as :

$$\Omega^a = \mathrm{d}\omega^a + \frac{1}{2} c^a_{bc} \omega^b \wedge \omega^c \tag{9.27}$$

Indeed, let write in the basis $\mathcal{B}$, $\omega = \omega^a v_a$ and $\Omega = \Omega^a v_a$. Hence, we have

$$\mathrm{d}\omega(X,Y) = \mathrm{d}\omega^a(X,Y) v_a \tag{9.28}$$

And

$$\omega(X), \omega(Y)] = [\omega^b(X) v_b, \omega^c(Y) v_c] = \omega^b(X)\omega^c(Y)[v_b, v_c] = \omega^b(X)\omega^c(Y) c^a_{bc} v_a$$

But

$$\begin{aligned} c^a_{bc}\omega^b(X)\omega^c(Y) &= \frac{1}{2}\left[c^a_{bc}\omega^b(X)\omega^c(Y) + c^a_{cb}\omega^c(X)\omega^b(Y)\right] \\ &= \frac{1}{2}\left[c^a_{bc}\omega^b(X)\omega^c(Y) - c^a_{bc}\omega^c(X)\omega^b(Y)\right] \\ &= \frac{1}{2} c^a_{bc}\omega^b(X) \wedge \omega^c(Y) \end{aligned}$$

Finally

$$[\omega(X), \omega(Y)] = \frac{1}{2} c^a_{bc}\omega^b(X) \wedge \omega^c(Y) v_a \tag{9.29}$$

And adding (9.28) and (9.29) we get the result.

We also have the famous Bianchi identity

$$D\Omega = 0.$$

An example : connection in a straight bundle

Let $\pi : E \to M$ a complex vector bundle of rank 1 (called a straight bundle where each fiber is diffeomorphic at $\mathbb{C}$) on a base manifold M and place a connection on E

$$\nabla : \mathfrak{X}(M) \times \Gamma(E) \longrightarrow \Gamma(E).$$

If $U \subset M$ is an open, any local section of E will be $e_1 \in \Gamma(E)$ such that $\pi \circ e_1 = id_U$; in particular if e_1 it is not canceled out on U, then it constitutes a local frame for E: we have that every other section $s \in \Gamma(E)$ will be written in the form $s = s^1 e_1$ with $s^1 \in C^\infty(U)$ with values in $\mathbb{C}$. If we then identify U with the domain of a local chart (U, φ) associated to the local coordinates $x^1, \ldots, x^m$ on M, then the local description of the connection will be

$$\nabla_{\partial_i} e_1 = \Gamma_i e_1, \tag{9.30}$$

If we take $X = X^j \partial_j \in \mathfrak{X}(U)$ and $s = s^1 e_1 \in \Gamma(E)$ we find that

$$\begin{aligned} \nabla_X s = \nabla_X(s^1 e_1) &= X(s^1) e_1 + s^1 \nabla_{X^j \partial_j} e_1 = X(s^1) e_1 + s^1 (X^j \nabla_{\partial_j} e_1) \\ &= X(s^1) e_1 + s^1 X^j \Gamma_j e_1 = (X(s^1) + X^j \Gamma_j s^1) e_1 \end{aligned}$$

and, without indicating the local frame, we get the simpler formula, $\nabla_X s^1 = X(s^1) + X^j \Gamma_j s^1$; if in addition $X = \partial_i$, then

$$\nabla_{\partial_i} s^1 = \partial_i s^1 + \Gamma_i s^1. \tag{9.31}$$

What is also interesting is how the connection change when we change the local frame. If $e_1, \tilde{e}_1$ are two different local frame of E on the open U we can express one in terms of the other as $\tilde{e}_1 = he_1$ with $h \in C^\infty(U)$ with complex values, then we will have $\nabla_{\partial_i} e_1 = \Gamma_i e_1$ and $\nabla_{\partial_i} \tilde{e}_1 = \tilde{\Gamma}_i e_1$; in particular the connection coefficients transform according to the following law:

$$\begin{aligned}\nabla_{\partial_i} \tilde{e}_1 &= \nabla_{\partial_i}(he_1) = (\partial_i h)e_1 + h\Gamma_i e_1 = (\partial_i h + h\Gamma_i)e_1 \\ &= (\partial_i h + h\Gamma_i)h^{-1}\tilde{e}_1 = (\Gamma_i + h^{-1}\partial_i h)\tilde{e}_1,\end{aligned}$$

whence it is obtained

$$\Gamma_i' = \Gamma_i + h^{-1}\partial_i h.$$

We also know that a connection one-form is a matrix $\omega = \left[\omega_i^j\right]$ where ω_i^j are one-forms. In this case ω is a matrix of order 1, therefore it will be

$$\omega = \omega_1^1 = \Gamma_{i1}^1 \underset{\cdot}{\mathrm{x}}^i = \Gamma_i \underset{\cdot}{\mathrm{x}}^i.$$

And using $\Omega = D\omega$, we can find the curvature 2-form $\Omega = \Omega_{ij} \underset{\cdot}{\mathrm{x}}^i \wedge \underset{\cdot}{\mathrm{x}}^j$:

$$\begin{aligned}\Omega &= (\partial_j \Gamma_i \underset{\cdot}{\mathrm{x}}^j) \wedge \underset{\cdot}{\mathrm{x}}^i = (\partial_j \Gamma_i - \partial_i \Gamma_j)\underset{\cdot}{\mathrm{x}}^j \wedge \underset{\cdot}{\mathrm{x}}^i \\ &= (\partial_i \Gamma_j - \partial_j \Gamma_i)\underset{\cdot}{\mathrm{x}}^i \wedge \underset{\cdot}{\mathrm{x}}^j,\end{aligned}$$

from which we get

$$\Gamma_{ij} = \partial_i \Gamma_j - \partial_j \Gamma_i. \tag{9.32}$$

9.3 An application in Physics : the case of electromagnetism

In this section, we will treat a remarkable application of the fibre bundles and connections theory in a physical context: gauge theories. We will see that the Lagrangian of a free relativistic particle is invariant under the global action of group $U_1(\mathbb{C})$; by introducing a principal bundle with this structural group on the $\mathbb{R}^4$ manifold - spacetime - it will be possible to introduce a covariant derivative (i.e. a connection) that safeguards its shape even for a local type of action and subsequently defines a curvature. In this way it will be possible to reinterpret the force exerted by the electromagnetic field as the physical manifestation of this curvature.

9.3.1 Some recall on electromagnetism

Electromagnetism is the study of the electromagnetic force, an interaction between electrically charged particles, one of the four fundamental interactions. Originally, the electricity and magnetism were seen as two different forces but works done in particular by Maxwell and Faraday show that these two forces can be seen as two faces of a same interaction : this yields to electromagnetism. One of its particularities is that it is compatible with special relativity.

In the following, we will place ourselves in the spacetime $\mathbb{R}^4$ seen as a (pseudo)Riemannian manifold equipped with the Minkowski metric η with signature $(+,-,-,-)$.

The spacetime components $\mathbf{x} \in \mathbb{R}^4$ will be denoted x^μ with $x^0 = ct$ where c is the speed of light.We will boldly indicate the points of spacetime while we will mark the vectors in ordinary three-dimensional space with arrows. The Greek indices (as in x^μ) will vary from 0 to 3, indicating the spacetime components and in particular the variable $x^0 = ct$ - denoting c the speed of light.

The electric field and magnetic field will be indicating by $\vec{E} = \vec{E}(t, \vec{x}) = (E^1, E^2, E^3)$ and $\vec{B} = \vec{B}(t, \vec{x}) = (B^1, B^2, B^3)$. As for the charge density and current density $\rho = \rho(t, \vec{x})$ and $\vec{\jmath} = \vec{\jmath}(t, \vec{x}) = (j^1, j^2, j^3)$ can be put together in the quadrivector $\daleth = j^\mu = (c\rho, \vec{\jmath})$ called quadricurrent.

Now let us recall Maxwell's equations in their classical differential form that show the link between the magnetic field and electric field :

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0} \tag{9.33a}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{9.33b}$$

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \tag{9.33c}$$

$$\vec{\nabla} \times \vec{B} - \frac{1}{c^2}\frac{\partial \vec{E}}{\partial t} = \mu_0 \vec{\jmath} \tag{9.33d}$$

Since we have the identities of the analysis $\vec{\nabla} \times (\vec{\nabla} F) = 0$ and $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{F}) = 0$, then by the last identity, equation (9.33b) suggests that we can introduce a vector field $\vec{A} = \vec{A}(t, \vec{x}) = (A^1, A^2, A^3)$ such that $\vec{B} = \vec{\nabla} \times \vec{A}$, called potential magnetic field vector; in this way we have

$$\vec{B} = \begin{bmatrix} B^1 \\ B^2 \\ B^3 \end{bmatrix} = \begin{bmatrix} \partial_2 A^3 - \partial_3 A^2 \\ \partial_3 A^1 - \partial_1 A^3 \\ \partial_1 A^2 - \partial_2 A^1 \end{bmatrix} = \vec{\nabla} \times \vec{A}. \tag{9.34}$$

Now expressing the magnetic field in terms of this potential magnetic field vector in equation (9.33c) we obtain

$$0 = \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = \vec{\nabla} \times \vec{E} + \frac{\partial}{\partial t}(\vec{\nabla} \times \vec{A}) = \vec{\nabla} \times \left(\vec{E} + \frac{\partial \vec{A}}{\partial t}\right),$$

and by the first identity, we can introduce a scalar field $\varphi = \varphi(t, \vec{x})$, called scalar potential such that

$$\vec{E} = \begin{bmatrix} E^1 \\ E^2 \\ E^3 \end{bmatrix} = \begin{bmatrix} -\partial_t A^1 - \partial_1 \varphi \\ -\partial_t A^2 - \partial_2 \varphi \\ -\partial_t A^3 - \partial_3 \varphi \end{bmatrix} = -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\varphi. \tag{9.35}$$

Now we can combine these two potentials into a single potential quadrivector of the electromagnetic field :

$$\mathbf{A} = A^\mu = (A^0, \vec{A}) = \left(\frac{\varphi}{c}, \vec{A}\right).$$

This potential quadrivector is not univocally determined : indeed is we consider the quadigradient of a scalar function $\Lambda : \mathbb{R}^4 \to \mathbb{R}$ there is an invariance (called gauge invariance) under the following transformation:

$$\left(\frac{\phi}{c}, \vec{A}\right) \longrightarrow \left(\frac{\phi'}{c}, \vec{A}'\right) = \left(\frac{\phi}{c} - \frac{1}{c}\frac{\partial \Lambda}{\partial t}, \vec{A} + \vec{\nabla}\Lambda\right).$$

The gauge consists in the choice of the Λ function. Two of the main choice of gauge is the Coulomb's gauge where we impose

$$\vec{\nabla} \cdot \vec{A} = 0$$

The other (the one that we will take here) is the Lorentz's gauge where we choose a Λ such that $\partial_\mu A^\mu = 0$, i.e.

$$\frac{1}{c^2}\frac{\partial \phi'}{\partial t} + \vec{\nabla} \cdot \vec{A} = 0$$

In this case, Maxwell's equations become

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\frac{\phi}{c} - \Delta\phi = \Box\phi = \mu_0 c\rho = \mu_0 J^0$$

$$\frac{1}{c^2}\frac{\partial^2 \vec{A}}{\partial t^2} - \Delta\vec{A} = \Box\vec{A} = \mu\vec{\jmath}.$$

where Δ indicates the Laplacian in the spatial components and $\Box$ indicates the Dalembertian operator. These two equations can be put together in the form

$$\Box A^\mu = \mu_0 J^\mu.$$

9.3.2 Elements of analytical mechanics

The dynamic of a physical quantity can be described by a physical quantity, called the action, on which we do a variational principle.

$$S_{A\to B} = \int_A^B \mathcal{L}(s)\mathrm{d}s$$

where s represent the position of the system in phase space. The least action principle states that, on all possible trajectories, the one taken effectively is the one making the action extremal. Usually, for a material point, the phase space is constituted with the position $\vec{x}$ and the speed $\dot{\vec{x}}$ of the material point and the action is written

$$S\left[\vec{x}(\cdot)\right] = \int_{t_0}^{t_1} L\left(\vec{x}(t), \dot{\vec{x}}(t)\right)\mathrm{d}t,$$

where $L : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ is an appropriate Lagrangian function.

In classical mechanics, the Lagrangian is given as the difference between kinetic energy and potential energy:

$$L\left(\vec{x},\dot{\vec{x}}\right)=\frac{1}{2}m|\dot{\vec{x}}|^2-V(\vec{x}),$$

The Hamilton's variational principle states that the evolution of a system minimizes the action : this principle leads to Euler-Lagrange equations that describe the motion of the system:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}^i}-\frac{\partial L}{\partial x^i}=0.$$

In the case of field theory, the notion of particle transforms to the concept of fields that can be seen as functions on space time with complex value. A particle is then an expression of an excited state of the physical field. In this case, we take into account the interaction of the particle in the Lagrangian by terms of interaction with the fields. If the fields are denoted by $\psi_k : \mathbb{R}^4 \to \mathbb{C}$, then we have

$$L(\mathbf{x},\dot{\mathbf{x}})=\int_\Omega \mathcal{L}\left(\psi_k(\mathbf{x}),\partial_\mu\psi_k(\mathbf{x})\right)\mathrm{d}\mathbf{x}$$

where $\mathcal{L}(\psi_k(\mathbf{x}),\partial_\mu\psi_k(\mathbf{x}))$ is the Lagrangian density and Ω is a subset of spacetime.

As before, we can define the action by the integral of the Lagrangian between to instant and doing Hamilton's principle, we derive the Euler-Lagrange equations in this case

$$\partial_\nu\frac{\partial\mathcal{L}}{\partial(\partial_\nu\psi_k)}=\frac{\partial\mathcal{L}}{\partial\psi_k}. \tag{9.37}$$

Now, the importance of investigating the action is that we can deduce conservation laws by looking at invariance of the action. This is the famous Noether's theorem

Theorem 9.3.1 *When the equations of motion (or, equivalently, the action S) are invariant under a continuous symmetry, there is a conserved current when the equation of motions are satisfied.*

Note that we can have two types of symmetries : symmetries of the Lagrangian under a transformation of the coordinates and internal symmetries of the Lagrangian that do not correspond to transformations on spacetime coordinates but characteristic of a field. In the following development, a field will be seen as a section of a bundle on $\mathbb{R}^4$ and an internal symmetry will be when the group action will be on the fiber points and not the point of the base space.

In our case of interest, we will look at the particles of the electromagnetic field : a photon γ. To specify its state, we need its spacetime position on the manifold $\mathbb{R}^4$ but also the polarization of the wave. If you do a rotation on the direction of oscillation, then the photon will still be at the same coordinates : an internal symmetry is then given by the action of the group of rotation $U_1(\mathbb{C})$ acting on the polarization plane.

9.3.3 $U_1(\mathbb{C})$ gauge theory

We now begin to introduce the Abelian gauge theory with gauge group $U_1(\mathbb{C})$ which describe electromagnetism.

We consider the Lagrangian density of a free relativistic particle which do not interact with other fields or other particles. In this case, $\mathcal{L}$ will depend only in the field of the particle and its derivatives. If we call $\psi : \mathbb{R}^4 \to \mathbb{C}$ the complex scalar field of the particle of mass m, then the Lagrangian is given by

$$\mathcal{L}(\psi, \overline{\psi}, \partial_\mu \psi, \partial_\mu \overline{\psi}) = \frac{1}{2}\lambda \left[\eta^{\mu\nu} \partial_\mu \psi \partial_\nu \overline{\psi} - \frac{m^2 c^2}{\hbar^2} \psi \overline{\psi} \right], \tag{9.38}$$

This Lagrangian has the particularity that it is invariant if we act on the field by an element of the group $U_1(\mathbb{C})$. Indeed, let $\alpha \in \mathbb{R}$ and let do the transformations :

$$\psi \longrightarrow \psi' = \mathrm{e}^{\mathrm{i}\alpha} \psi, \qquad \overline{\psi} \longrightarrow \overline{\psi'} = \mathrm{e}^{-\mathrm{i}\alpha} \overline{\psi}, \tag{9.39}$$

then the Lagrangian $\mathcal{L}' = \mathcal{L}(\psi', \overline{\psi'}, \partial_\mu \psi', \partial_\mu \overline{\psi'})$ transforms as follows

$$\begin{aligned} \mathcal{L}' &= \frac{1}{2} \left[\eta^{\mu\nu} \partial_\mu \psi' \partial_\nu \overline{\psi'} - \frac{m^2 c^2}{\hbar^2} \psi' \overline{\psi'} \right] \\ &= \frac{1}{2} \left[\eta^{\mu\nu} \partial_\mu (\mathrm{e}^{\mathrm{i}\alpha} \psi) \partial_\nu (\mathrm{e}^{-\mathrm{i}\alpha} \overline{\psi}) - \frac{m^2 c^2}{\hbar^2} (\mathrm{e}^{\mathrm{i}\alpha} \psi)(\mathrm{e}^{-\mathrm{i}\alpha} \overline{\psi}) \right] \\ &= \frac{1}{2} \left[\mathrm{e}^{\mathrm{i}\alpha} \mathrm{e}^{-\mathrm{i}\alpha} \eta^{\mu\nu} \partial_\mu \psi \partial_\nu \overline{\psi} - \frac{m^2 c^2}{\hbar^2} \psi \overline{\psi} \right] = \mathcal{L}. \end{aligned}$$

Now one of the problem in our model is that we have to take into account the locality principle i.e. that distant objects cannot have instantaneous mutual influence. In this case, we can ask how to compare physically the field in two distinct points $\psi(\mathbf{x})$, $\psi(\mathbf{y})$.

Mathematical description of the problem

One way to solve this problem mathematically is to attach to each point $\mathbf{x}$ of spacetime a copy of $\mathbb{C}$ to which belong the value of $\psi(\mathbf{x})$. To do so, we introduce a fibre bundle structure of base manifold $\mathbb{R}^4$: in particular, we introduce a $U_1(\mathbb{C})$-principal bundle $\pi : P \to \mathbb{R}$ to which we associate a vector bundle $\pi_L : L \to \mathbb{R}^4$ where $L = P \times_{U_1(\mathbb{C})} \mathbb{C}$ which is a line bundle, on which $U_1(\mathbb{C})$ acts with the classical multiplication.

In this case, the field ψ can be reinterpreted as a differential section of L. More precisely, to each open U of $\mathbb{R}^4$, there is $s = \psi : U \to L|_U = \pi^{-1}(U)$ such that $\pi_L \circ s = id_U$ and for all $\mathbf{x}$, $\psi(\mathbf{x}) \in L_{\mathbf{x}} = \pi^{-1}(\{x\}) \cong \mathbb{C}$.

As for the example made in the previous section, if we have a local frame $e_1 \in \Gamma(L)$, the section s will be written $s = \psi^1 e_1$, with $\psi^1 \in C^\infty(U)$ with value in $\mathbb{C}$. In all the following, we will identify the section with the field.

Now, a difference about what was done before is that it is reasonable to assume that the group action $U_1(\mathbb{C})$ depends on the point where it is applied. In fact, mathematically, this dependence is to take into account that the value $\psi(\mathbf{x})$ belong to the fibre $L_{\mathbf{x}}$. Then the problem of comparing to field at two distinct points can be overcome by defining a connection on L and then do a parallel translation to identify the 2 fibers $L_{\mathbf{x}}$ and $L_{\mathbf{y}}$.

Hence here, we replace the multiplication $e^{i\alpha}$ by $e^{i\alpha(\mathbf{x})}$ where $\alpha : \mathbb{R}^4 \to \mathbb{R}$. The transformations of the field are then

$$\psi \longrightarrow \psi' = e^{i\alpha(\mathbf{x})}\psi, \qquad \overline{\psi} \longrightarrow \overline{\psi'} = e^{-i\alpha(\mathbf{x})}\overline{\psi}. \tag{9.40}$$

Now we need to see if the Lagrangian keeps its invariance if we apply this new transformation. For the derivatives of the field, we have the following transformations :

$$\begin{aligned}
\partial_\mu \psi' &= \partial_\mu \left(e^{i\alpha(\mathbf{x})}\psi \right) = i\partial_\mu \alpha(\mathbf{x}) e^{i\alpha(\mathbf{x})}\psi + e^{i\alpha(\mathbf{x})}\partial_\mu \psi, \\
\partial_\mu \overline{\psi'} &= \partial_\mu \left(e^{-i\alpha(\mathbf{x})}\overline{\psi} \right) = -i\partial_\mu \alpha(\mathbf{x}) e^{-i\alpha(\mathbf{x})}\overline{\psi} + e^{-i\alpha(\mathbf{x})}\partial_\mu \overline{\psi};
\end{aligned}$$

The invariance of the Lagrangian is lost !

In fact, the lost of the invariance can be understood as the fact that the standard directional derivative is no longer well defined on the field ψ seen as a section. The solution is to introduce a connection on the bundle L that will give us a covariant derivative ∇_μ to replace ∂_μ. This will give a properly derivative of the field ψ.

To do so, we introduce a field, called gauge field, which transforms in a way to preserve the invariance of the Lagrangian. We denote $\mathbf{A} = A_\mu$ (in fact, we are looking at the coefficients of the differential form corresponding to the field A^μ by $A_\mu = \eta_{\mu\nu}A^\nu$).

The transformation that we ask (and that we can retrieve, see below) is the following

$$A_\mu \longrightarrow A'_\mu = A_\mu - \frac{1}{q}\partial_\mu \alpha(\mathbf{x}). \tag{9.41}$$

where here q is the charge of the particle[1]. Then the connection is $\nabla : \mathfrak{X}(\mathbb{R}^4) \times \Gamma(L) \to \Gamma(L)$ which maps $(\partial_\mu, s) \longmapsto \nabla_\mu s$, where the operator ∇_μ is defined by the position

$$\nabla_\mu := \partial_\mu + iqA_\mu.$$

Since we are in the case of a bundle of rank 1, we have the identification

$$\Gamma_\mu = iqA_\mu$$

We can thus replace ∂_μ in the Lagrangian by ∇_μ to get

$$\mathcal{L}(\psi, \overline{\psi}, \partial_\mu\psi, \partial_\mu\overline{\psi}) \to \mathcal{L}(\psi, \overline{\psi}, \partial_\mu\psi, \partial_\mu\overline{\psi}, A_\mu) = \mathcal{L}(\psi, \overline{\psi}, \nabla_\mu\psi, \nabla_\mu\overline{\psi}).$$

We can remark that the action of $U(1)$ on ψ induce in fact an action on the local frame e_1. Indeed, if e_1 and $\tilde{e}_1$ are two local bases of L, then any section $s \in \Gamma(L)$ can be written $s = \psi^1 e_1 = \tilde{\psi}^1 \tilde{e}_1$. Hence, we have

$$\psi^1 e_1 = s = \tilde{\psi}^1 \tilde{e}_1 = e^{i\alpha(\mathbf{x})}\psi \tilde{e}_1 \iff e_1 = e^{i\alpha(\mathbf{x})}\tilde{e}_1,$$

[1] In general, if $\sigma : U \to P$ is a local section of a G-principal bundle $P \to M$, and if we define $A := \sigma^*(\omega)$ where ω is a connection of the principal bundle, then for every principal automorphism ϕ, there exists some $\Omega : U \to G$ such that for all $x \in U$, $\sigma(x) = \phi \circ \sigma(x)\Omega(x)$. In the case, the transformation of the local representative A can be written :

$$A_\mu(x) \mapsto \Omega(x)A_\mu(x)\Omega^{-1}(x) + \Omega(x)\partial_\mu\Omega^{-1}(x).$$

i.e. that the e_1 transform with the action of $U_1(\mathbb{C})$ as

$$e_1 \mapsto \tilde{e}_1 = \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} e_1.$$

We can also note that $U(1)$ also acts on the gauge potential, and this will allow us to get the transformation given by (9.41). Indeed, if we have two local bases of L, e_1 and $\tilde{e}_1$ such that $\tilde{e}_1 = \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})}$ then we have $\Gamma_\mu = \mathrm{i}qA_\mu$ and $\Gamma'_\mu = \mathrm{i}qA'_\mu$. Let's see how Γ_μ and Γ'_μ are related :

$$\begin{aligned}
\Gamma'_\mu \tilde{e}_1 = \nabla_\mu \tilde{e}_1 = \nabla_\mu(\mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} e_1) &= \partial_\mu(\mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})}) e_1 + \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} \nabla_\mu e_1 \\
&= -\mathrm{i}\partial_\mu \alpha(\mathbf{x}) \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} e_1 + \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} \Gamma_\mu e_1 \\
&= -\mathrm{i}\partial_\mu \alpha(\mathbf{x}) \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \tilde{e}_1 + \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} \Gamma_\mu \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \tilde{e}_1 \\
&= \left(\Gamma_\mu - \mathrm{i}\partial_\mu \alpha(\mathbf{x})\right) \tilde{e}_1,
\end{aligned}$$

hence we have $\Gamma'_\mu = \Gamma_\mu - \mathrm{i}\partial_\mu \alpha(\mathbf{x})$.

Now, it remains to see if indeed the new Lagrangian is invariant by the action of $U_1(\mathbb{C})$.

Since we have introduced a new field, we need to take it into account in the Lagrangian. We need to add another term $\mathcal{L}_{\mathbf{A}}$ that depends only on the gauge potential $\mathbf{A}$ and its derivatives and we construct it so that it is invariant under the transformation of the gauge potential. The new Lagrangian of electrodynamics is the following

$$\mathcal{L}_{ED}(\psi, \overline{\psi}, \nabla_\mu \psi, \overline{\nabla_\mu \psi}, \mathbf{A}, \partial_\mu \mathbf{A}) = \mathcal{L}(\psi, \overline{\psi}, \nabla_\mu \psi, \overline{\nabla_\mu \psi}) + \mathcal{L}_{\mathbf{A}}(\mathbf{A}, \partial_\mu \mathbf{A}).$$

Let's see if the first term of the Lagrangian is now invariant by the action under the transformations (9.40) and (9.41).

The new Lagrangian is written

$$\mathcal{L}(\psi, \overline{\psi}, \nabla_\mu \psi, \overline{\nabla_\mu \psi}) = \frac{1}{2}\left[\eta^{\mu\nu} \nabla_\mu \psi \overline{\nabla_\nu \psi} - \frac{m^2 c^2}{\hbar^2} \psi \overline{\psi}\right]$$

Let's see how the terms $\nabla_\mu \psi$ transforms by the action of the group :

$$\begin{aligned}
\nabla'_\mu \psi' = (\partial_\mu + \mathrm{i}qA'_\mu)\psi' &= (\partial_\mu + \mathrm{i}qA_\mu - \mathrm{i}\partial_\mu \alpha(\mathbf{x}))\left(\mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \psi\right) \\
&= \mathrm{i}\partial_\mu \alpha(\mathbf{x}) \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \psi + \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \partial_\mu \psi + \mathrm{i}qA_\mu \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \psi - \mathrm{i}\partial_\mu \alpha(\mathbf{x}) \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \psi \\
&= \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \left(\partial_\mu + \mathrm{i}qA_\mu\right) \psi = \mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})} \nabla_\mu \psi
\end{aligned}$$

Similarly, the term $\overline{\nabla_\mu \psi}$ is transformed according to the law

$$\overline{\nabla_\mu \psi} \longrightarrow \overline{\nabla'_\mu \psi'} = \mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})} \overline{\nabla_\mu \psi}.$$

Finally, we have

$$\begin{aligned}
\mathcal{L}' &= \mathcal{L}(\psi', \overline{\psi'}, \nabla'_\mu \psi', \overline{\nabla'_\mu \psi'}) \\
&= \frac{1}{2}\left[\eta_{\mu\nu} \nabla'_\mu \psi' \overline{\nabla'_\nu \psi'} - \frac{m^2c^2}{\hbar^2}\psi'\overline{\psi'}\right] \\
&= \frac{1}{2}\left[\eta^{\mu\nu}\left(\mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})}\nabla_\mu\psi\right)\left(\mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})}\overline{\nabla_\mu\psi}\right) - \frac{m^2c^2}{\hbar^2}\left(\mathrm{e}^{\mathrm{i}\alpha(\mathbf{x})}\psi\right)\left(\mathrm{e}^{-\mathrm{i}\alpha(\mathbf{x})}\overline{\psi}\right)\right] \\
&= \frac{1}{2}\left[\eta^{\mu\nu}\nabla_\mu\psi\overline{\nabla_\nu\psi} - \frac{m^2c^2}{\hbar^2}\psi\overline{\psi}\right] = \mathcal{L},
\end{aligned}$$

Hence the new Lagrangian is now invariant by the action of the group $U(1)$.

Curvature and electromagnetic field

Now that we have a connection, we can look at the corresponding curvature 2-form $\Omega = \Omega_{\mu\nu} \underset{\cdot}{\mathrm{x}}^\mu \wedge \underset{\cdot}{\mathrm{x}}^\nu$ given by the formula (9.32)

$$\Omega_{\mu\nu} = \partial_\mu(\mathrm{i}qA_\nu) - \partial_\nu(\mathrm{i}qA_\mu) = \mathrm{i}q(\partial_\mu A_\nu - \partial_\nu A_\mu) = \mathrm{i}qF_{\mu\nu},$$

where we have set $F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu$[2]. It is the coefficients of an antisymmetric 2-covariant tensor. Hence $F = F_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$ is a differential 2-form. It is also an exact form because it can be written as the exterior derivative of the 1-form $A = A_\mu \mathrm{d}\mu$

$$\begin{aligned}
F = \mathrm{d}A &= (\partial_\nu A_\mu \mathrm{d}x^\nu) \wedge \mathrm{d}x^\mu \\
&= (\partial_\nu A_\mu - \partial_\mu A_\nu)\mathrm{d}x^\nu \wedge \mathrm{d}x^\mu = F_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu.
\end{aligned}$$

We can remark that this tensor is invariant under the action (9.41) of the group $U(1)$

$$\begin{aligned}
F_{\mu\nu} \longrightarrow F'_{\mu\nu} &= \partial_\mu A'_\nu - \partial_\nu A'_\mu \\
&= \partial_\mu\left(A_\nu - \frac{1}{q}\partial_\nu\alpha(\mathbf{x})\right) - \partial_\nu\left(A_\mu - \frac{1}{q}\partial_\mu\alpha(\mathbf{x})\right) \\
&= \partial_\mu A_\nu - \frac{1}{q}\partial_\mu\partial_\nu\alpha(\mathbf{x}) - \partial_\nu A_\mu + \frac{1}{q}\partial_\nu\partial_\mu\alpha(\mathbf{x}) \\
&= \partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}.
\end{aligned}$$

In terms of matrix representation, $F = \left[F_{\mu\nu}\right]$, can be written in the form

$$F = \begin{bmatrix} F_{00} & F_{01} & F_{02} & F_{03} \\ F_{10} & F_{11} & F_{12} & F_{13} \\ F_{20} & F_{21} & F_{22} & F_{23} \\ F_{30} & F_{31} & F_{32} & F_{33} \end{bmatrix}.$$

[2] In the case the group action is an non-abelian group, we would have additional terms with the commutator on the field. Mainly, we would have

$$F_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu - \mathrm{i}q[\mathbf{A}_\mu, \mathbf{A}_\nu]$$

.

Now seeing the components A_μ as the components of the potential quadrivector of the electromagnetic field $\mathbf{A} = A^\mu = (\phi/c, \overrightarrow{A})$, and remembering that we can identify $\partial_0 = \frac{1}{c}\partial_t$, this matrix can be expressed in terms of the electric and magnetic fields. We have in particular

$$\begin{aligned}
F_{01} &= \partial_0 A_1 - \partial_1 A_0 = -\frac{1}{c}\partial_t A^1 - \partial_1\left(\frac{\phi}{c}\right) = \frac{1}{c}E^1 \\
F_{02} &= \partial_0 A_2 - \partial_2 A_0 = -\frac{1}{c}\partial_t A^2 - \partial_2\left(\frac{\phi}{c}\right) = \frac{1}{c}E^2 \\
F_{03} &= \partial_0 A_3 - \partial_1 A_0 = -\frac{1}{c}\partial_t A^3 - \partial_3\left(\frac{\phi}{c}\right) = \frac{1}{c}E^3 \\
F_{12} &= \partial_1 A_2 - \partial_2 A_1 = -\partial_1 A^2 + \partial_2 A^1 = -B^3 \\
F_{13} &= \partial_1 A_3 - \partial_3 A_1 = -\partial_1 A^3 + \partial_3 A^1 = B^2 \\
F_{23} &= \partial_2 A^3 - \partial_3 A_2 = -\partial_2 A^3 + \partial_3 A^2 = -B^1,
\end{aligned}$$

Finally we get the tensor

$$F = \begin{bmatrix} 0 & E^1/c & E^2/c & E^3/c \\ -E^1/c & 0 & -B^3 & B^2 \\ -E^2/c & B^2 & 0 & -B^1 \\ -E^3/c & -B^2 & B^1 & 0 \end{bmatrix}$$

which is the Faraday tensor. Hence the electromagnetic fields is the manifestation of the curvature associated to the connection on the fibre bundle $\pi : L \to \mathbb{R}^4$. [3]

Maxwell's equations

Finally, let's go back to Lagrangian and see how we can retrieve Maxwell's equations.

In the Lagrangian, we have still one unknown in the choice of the added term of the Lagrangian $\mathcal{L}_{\mathbf{A}}$. This term must be a scalar and should be invariant under the action of the group $U(1)$. A solution is to put

$$\mathcal{L}_{\mathbf{A}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{9.42}$$

Finally, the total expression of the Lagrangian is

$$\begin{aligned}
\mathcal{L}_{ED} = \mathcal{L} + \mathcal{L}_{\mathbf{A}} &= \frac{1}{2}\left[\eta^{\mu\nu}\nabla_\mu\psi\overline{\nabla_\nu\psi} - \frac{m^2c^2}{\hbar^2}\psi\overline{\psi}\right] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\
&= \frac{1}{2}\left[\eta^{\mu\nu}(\partial_\mu + \mathrm{i}qA_\mu)\psi(\partial_\nu - \mathrm{i}qA_\nu)\overline{\psi} - \frac{m^2c^2}{\hbar^2}\psi\overline{\psi}\right] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\
&= \frac{1}{2}\eta^{\mu\nu}[\partial_\mu\psi\partial_\nu\overline{\psi}] + \frac{1}{2}\eta^{\mu\nu}\mathrm{i}q[A_\mu\psi\partial_\nu\overline{\psi} - A_\nu\overline{\psi}\partial_\mu\psi - \mathrm{i}qA_\mu A_\nu\psi\overline{\psi}] - \frac{m^2c^2}{\hbar^2}\psi\overline{\psi} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}.
\end{aligned}$$

Now, the choice of $\mathcal{L}_{\mathbf{A}}$ was made so that with this Lagrangian, we could retrieve Maxwell's equation. Indeed, let go in vacuum where there is no particle. In this case, the Lagrangian reduce to the term $\mathcal{L}_{\mathbf{A}}$. First we get

[3] The curvature is physically expressed as the force of the electromagnetic field.

$$
\begin{aligned}
\mathcal{L}_{ED} = \mathcal{L}_{\mathbf{A}} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = -\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&= -\frac{1}{4}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu - \partial_\nu A_\mu \partial^\mu A^\nu + \partial_\nu A_\mu \partial^\nu A^\mu) \\
&= -\frac{1}{2}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu).
\end{aligned}
$$

And the Euler-Lagrange equations associated with this Lagrangian are :

$$
\partial_\nu \frac{\partial \mathcal{L}_{\mathbf{A}}}{\partial(\partial_\nu A_\mu)} = \frac{\partial \mathcal{L}_{\mathbf{A}}}{\partial A_\mu}. \tag{9.43}
$$

Since the Lagrangian does not dependent on the components of the gauge potential, the second term vanishes. For the first term, we have

$$
\frac{\partial \mathcal{L}_{\mathbf{A}}}{\partial(\partial_\nu A_\mu)} = -\frac{1}{2}(\partial^\nu A^\mu - \partial^\mu A^\nu) = -\frac{1}{2}F^{\nu\mu} = \frac{1}{2}F^{\mu\nu},
$$

Hence, Euler-Lagrange equations are finally given by

$$
\partial_\nu F^{\mu\nu} = 0. \tag{9.44}
$$

This equation in fact is a covariant form of two of the Maxwell's equations.

Indeed, take $\mu = 0$. In this case we obtain Maxwell-Gauss equation :

$$
0 = \partial_\nu F^{0\nu} = \partial_0 F^{00} + \partial_i F^{0i} = -\frac{1}{c}\partial_i E^i \iff \vec{\nabla} \cdot \vec{E} = 0;
$$

Now taking $\mu = i$, this will lead to Maxwell-Ampere equation :

For example, if $i = 1$, we find

$$
0 = \partial_\nu F^{1\nu} = \partial_0 F^{10} + \partial_j F^{1j} = \frac{1}{c^2}\partial_t E^1 + \partial_2(-B^3) + \partial_3 B^2,
$$

which is equivalent to $\partial_2 B^3 - \partial_3 B^2 - \frac{1}{c^2}\partial_t E^1 = 0$ and is the first component of the vector equation

$$
\vec{\nabla} \times \vec{B} - \frac{1}{c^2}\frac{\partial \vec{E}}{\partial t} = 0;
$$

The other components are obtained by taking $i = 2, 3$.

It remains to find the two last equations. This will be done thanks to Bianchi's identity. From the curvature 2-form $\Omega = \mathrm{i}qF_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$, we have

$$
\begin{aligned}
0 = D\Omega = \mathrm{d}\Omega &= (\partial_\alpha \Omega_{\beta\gamma}\mathrm{d}x^\alpha) \wedge \mathrm{d}x^\beta \wedge \mathrm{d}x^\gamma \\
&= (\partial_\alpha \Omega_{\beta\gamma} - \partial_\alpha \Omega_{\gamma\beta} + \partial_\beta \Omega_{\gamma\alpha} - \partial_\beta \Omega_{\alpha\gamma} + \partial_\gamma \Omega_{\alpha\beta} - \partial_\gamma \Omega_{\beta\alpha})\mathrm{d}x^\alpha \wedge \mathrm{d}x^\beta \wedge \mathrm{d}x^\gamma \\
&= 2(\partial_\alpha \Omega_{\beta\gamma} + \partial_\beta \Omega_{\gamma\alpha} + \partial_\gamma \Omega_{\alpha\beta})\mathrm{d}x^\alpha \wedge \mathrm{d}x^\beta \wedge \mathrm{d}x^\gamma,
\end{aligned}
$$

Hence we get $\partial_\alpha \Omega_{\beta\gamma} + \partial_\beta \Omega_{\gamma\alpha} + \partial_\gamma \Omega_{\alpha\beta} = 0$, and this leads ultimately to

$$
\partial_\alpha F_{\beta_\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0. \tag{9.45}
$$

We can then derive the two last equations :

- Maxwell-Flux is obtained by only looking at space variables i.e. taking $\alpha\beta\gamma = ijk$. We get
$$0 = \partial_1 F_{23} + \partial_2 F_{31} + \partial_3 F_{12} = \partial_1(-B^1) + \partial_2(-B^2) + \partial_3(-B^3),$$
which is nothing less than
$$\vec{\nabla} \cdot \vec{B} = 0;$$
- For Maxwell-Faraday, we take $\alpha\beta\gamma = 0ij$. In this case, setting for example $i = 2$ and $j = 3$, we have
$$0 = \partial_0 F_{23} + \partial_2 F_{30} + \partial_3 F_{02} = \frac{1}{c}\partial_t(-B^1) + \frac{1}{c}\partial_2(-E^3) + \frac{1}{c}\partial_3(E^2),$$
i.e. $\partial_3 E^2 - \partial_2 E^3 + \frac{1}{c}\partial_t B^1$, which is the first component of the vector equation
$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0.$$
and the two other components are obtained by taking $\alpha\beta\gamma = 0\ 1\ 2$ and $\alpha\beta\gamma = 0\ 1\ 3$.

Part II:

Homogeneous spaces and hyperbolic geometry

The ‘uproar of the Boeotians’.
(Attributed to) Carl Friedrich Gauss

Chapter 10

Homogeneous spaces

A homogeneous space X is to be understood as a space that remains stable under a group of transformations G and such that its points are all 'connected' by the transformations of G. In the theory of homogeneous spaces, *the main attention is concentrated on the transformations of* G *and not on the elements of* X, the reason underlying this is is given by the stabilizer-orbit theorem, which says that X *can be reconstructed via a suitable quotient of* G. This result will allow us to exhibit extremely important examples of homogeneous spaces.

10.1 Preliminaries : group actions and linear transformation groups

10.1.1 Group actions

In this section we shall consider G to be a group and X a non-empty set. 1_G denotes the neutral element of G.

Def. 10.1.1 *The action of a group* G *on* X *is given by an operation*

$$\begin{array}{rccl} \eta : & G \times X & \longrightarrow & X \\ & (g,x) & \longmapsto & \eta(g,x) := g \cdot x \end{array}$$

which verifies, for all $x \in X$ *and* $g, h \in G$*:*

1. $1_G \cdot x = x$
2. $g \cdot (h \cdot x) = (gh) \cdot x$.

If we fix any element $g \in G$, the group action η induces a bijective function on X by $\eta_g : X \to X$, $x \mapsto \eta_g(x) = \eta(g,x)$, its inverse being obviously $\eta_{g^{-1}}$. This remark shows that, if G acts on X, then it can be seen as a subgroup of $\mathrm{Sym}(X)$, the group of all bijective functions on X and the action η can be equivalently characterized by the group homomorphism $\tilde{\eta} : G \to \mathrm{Sym}(X)$, $g \mapsto \tilde{\eta}(g) := \eta_g$. In fact, requiring $\tilde{\eta}$ to be a group homomorphism we assure that $\tilde{\eta}(1_G) = Id_X$, hence $\tilde{\eta}(1_G)(x) = x \ \forall x \in X$, and $\tilde{\eta}(gh) = \tilde{\eta}(g) \circ \tilde{\eta}(h) = \eta_g \circ \eta_h$, so $\tilde{\eta}(gh)(x) = (gh) \cdot x \ \forall x \in X$.

Example 10.1.1 *Some basic examples of group actions are listed below.*

1. The usual multiplication by a scalar belonging to the field $\mathbb{K}$ on which a vector space V is defined is a group action by interpreting $\mathbb{K}$ as G and V as X:

$$\lambda \cdot \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} = \begin{pmatrix} \lambda v^1 \\ \vdots \\ \lambda v^n \end{pmatrix}, \qquad \lambda \in \mathbb{K},\ v = (v^1, \dots, v^n)^t \in V.$$

2. The usual matrix multiplication of $GL(n, \mathbb{R})$ on the vector space $\mathbb{R}^n$ is a group action.

3. $X = \{1, 2, 3\}$. Then, any subgroup of S_3, the group of all permutations of X, for example $A_3 = \{Id, (123), (132)\}$, operates as a group action on X.

4. $X = D_{\mathbb{R}}(0, 1) = \{(x, y)^t \in \mathbb{R}^2 :\ x^2 + y^2 \leqslant 1\}$, the unit disk in $\mathbb{R}^2$, and

$$G = SO(2) = \left\{ \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix},\ \vartheta \in [0, 2\pi) \right\},$$

   the group of rotations in $\mathbb{R}^2$. Then, G operates on $D_{\mathbb{R}}(0, 1)$ by matrix multiplication.

5. $X = D_{\mathbb{C}}(0, 1) = \{z \in \mathbb{C} :\ |z| \leqslant 1\}$, the unit disk in $\mathbb{C}$, and

$$G = U(1) := \{e^{i\theta} :\ \theta \in [0, 2\pi)\},$$

   the group of rotations in $\mathbb{C}$. Then, G operates on $D_{\mathbb{C}}(0, 1)$ by matrix multiplication.

We now define the most important subspaces of X and G associated to the action of a group: the orbit and the stabilizer, respectively.

Def. 10.1.2 *Let the group G act on the set X and fix any $x \in X$.*

1. *The G-**orbit of** x is the subset of X given by:*

$$\mathrm{Orb(x)} = \{g \cdot x : g \in G\} \subset X,$$

   i.e. all the elements $y \in X$ that can be connected to x by a transformation $g \in G$: $y = g \cdot x$.

2. *The **stabilizer group of** x (or **isotropy subgroup**, or **little group** of x) is given by:*

$$\mathrm{Stab}(x) = \mathrm{G}_x = \{g \in G : g \cdot x = x\} \subset G,$$

   i.e. the set of transformations of G that act as the identity on x, leaving it unaltered.

The use of the word group for the stabilizer of $x \in X$ is not accidental, one can easily prove that $\mathrm{Stab}(x)$ is a subgroup of G.

Def. 10.1.3 (G-homogeneous space) *We say that X is a G-homogeneous space (or that G operates transitively on X) if it exists at least one $x \in X$ such that $X = \mathrm{Orb}(x)$, i.e for all $y \in X$ there exists an element $g \in G$ such that $g \cdot x = y$.*

It is easy to see that the request of existence of at least one element x of X whose G-orbit is the whole X is equivalent to the fact that the G-orbits of all the elements of X are the whole X. In fact, consider two arbitrary elements $y, \overline{x} \in X$, then there exist $g, \overline{g} \in G$ such that $g \cdot x = y$ and $\bar{g} \cdot x = \bar{x}$, i.e. $x = \bar{g}^{-1} \cdot \bar{x}$, so $(g\bar{g}^{-1}) \cdot \bar{x} = y$ thus also $\mathrm{Orb}(\bar{x}) = X$. Hence, **a G-homogeneous space has only one G-orbit: X itself**!

Often, a G-homogeneous space X is defined by requiring that, for any couple $x, y \in X$, there exists at least an element $g \in G$ such that $g \cdot x = y$. The two definitions are of course equivalent.

Consequently, a homogeneous space is fully 'connected' by the group that operates upon it: any point of X reaches any other point via a group transformation. This property is often popularized by saying that, set-theoretically speaking, *in a homogeneous space, no point is more important than other*, which explains the adjective 'homogeneous'.

Example 10.1.2 Consider again the group $U(1)$, the unit complex disk $D_{\mathbb{C}}(0,1)$ and its contour $\partial D_{\mathbb{C}}(0,1) = \{z \in \mathbb{C} : |z| = 1\}$. Then, $\partial D_{\mathbb{C}}(0,1)$ **is trivially $U(1)$-homogeneous** because for any couple of points z, w on the unit circle in $\mathbb{C}$ separated by the angle θ, we have that $w = e^{i\theta}z$.

However, $D_{\mathbb{C}}(0,1)$ **is not $U(1)$-homogeneous**, in fact for any $z, w \in D_{\mathbb{C}}(0,1)$ and any $\theta \in [0, 2\pi)$, if we write $w = e^{i\theta}z$ then $|w| = |z|$, thus it is enough to consider two elements inside the unit disk with different modulus, e.g. $z = \frac{1}{2}$ and $w = \frac{2}{3}i$, to exhibit a couple of points of $D_{\mathbb{C}}(0,1)$ that cannot be connected by a transformation of $U(1)$.

The same considerations can be repeated in the real case to prove that the contour of the real unit disk, i.e. $\partial D_{\mathbb{R}}(0,1) = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 = 1\} \cong S^1$**, is $SO(2)$-homogeneous**, and, of course, also $O(2)$-homogeneous, but the disk itself $D_{\mathbb{R}}(0,1)$ **is not $SO(2)$-homogeneous**. In spite of the fact that $D_{\mathbb{R}}(0,1)$ is not homogeneous w.r.t. rotations, we will see that **it is homogeneous w.r.t. hyperbolic rotations**.

Example 10.1.3 The subgroup H of a group G is always a G-homogeneous space. In fact, the unit element 1_G of G belongs to H and it is connected with all the other elements of H. To see this, take any $h \in H$, then h belongs also to G, so $h = 1_G \cdot h$, which shows the transitivity of G on H.

We now come to the most important result of this section. To introduce it, we first recall that, given a group G, a fixed element $g \in G$ and a subgroup H of G, the **left coset** of H in G relative to g is the set:

$$gH := \{gh \ : \ h \in H\}.$$

For all fixed $g \in G$, belonging to the g-left coset of H is an equivalence relationship on G, thus, as g varies in G, we subdivide G into disjoint subsets, the cosets gH. The union of these classes is the **quotient space** G/H:

$$G/H := \{gH, \ g \in G\} \equiv \{\{gh \ : \ h \in H\}, \ g \in G\}.$$

G/H is a group if and only if H is a **normal subgroup** of G, where H is called normal if it is stable under conjugation by elements of G, i.e. if $\forall h \in H$ and $\forall g \in G$ it holds $ghg^{-1} \in H$.

Clearly, the easiest case is represented by $H = \{1_G\}$, in this situation it is evident that $G/1_G \cong G$.

To introduce the paramount important orbit-stabilizer theorem, we first notice that, for every fixed $x \in X$, the map

$$\begin{array}{ccc} G & \longrightarrow & \mathrm{Orb}(x) \\ g & \longmapsto & g \cdot x \end{array}$$

is surjective by definition of orbit but, in general, is not injective. However, as the following result says, **if we quotient G on the stabilizer of x, then we remove all possible redundancy and we remain with a bijection**.

Theorem 10.1.1 (Orbit-stabilizer theorem) *Let $x, y \in X$ and G a group acting on X.*

1. *The map*

$$\begin{array}{ccc} G \diagup_{\mathrm{Stab}(x)} & \stackrel{\sim}{\longrightarrow} & \mathrm{Orb}(x) \\ g\,\mathrm{Stab}(x) & \longmapsto & g \cdot x \end{array} \tag{10.1}$$

   is bijective.

2. $\mathrm{Orb}(x) = \mathrm{Orb}(y) \implies \exists g \in G$ *such that* $g\,\mathrm{Stab}(x)\,g^{-1} = \mathrm{Stab}(y)$, *i.e. if the orbits of two elements of X coincide, then their stabilizers are conjugated by an element g of G, and, as such, they are isomorphic to each other.*

Proof.

1. First of all, let us check that the application (10.1) is well-defined, i.e. it does not depend on the choice of the representative in the equivalence class g $\mathrm{Stab}(x)$. If $h \in g\,\mathrm{Stab}(x)$, there exists $k \in \mathrm{Stab}(x)$ such that $h = gk$, then $h \cdot x = g \cdot (k \cdot x) = g \cdot x$.

Injectivity of (10.1): let $g, h \in G$ such that $g \cdot x = h \cdot x$, we must prove that this implies $g\,\mathrm{Stab}(x) = h\,\mathrm{Stab}(x)$. To do this, notice that $(h^{-1}g) \cdot x = x$, so $h^{-1}g \in \mathrm{Stab}(x)$, i.e. $g \in h\,\mathrm{Stab}(x)$. However, g belongs also to $g\,\mathrm{Stab}(x)$ because $1_G \in \mathrm{Stab}(x)$, hence g belongs to the intersection of the equivalence classes $h\,\mathrm{Stab}(x)$ and $g\,\mathrm{Stab}(x)$, which, however, are disjoint. Thus, the only possibility that remains valid is that $g\,\mathrm{Stab}(x) = h\,\mathrm{Stab}(x)$.

Surjectivity of (10.1): any $y \in \mathrm{Orb}(x)$ is written as $g \cdot x = y$ for some $g \in G$, but then it is the image of (10.1) because any element of $g\,\mathrm{Stab}(x)$ can be written as gk, with $k \in \mathrm{Stab}(x)$, so $(gk) \cdot x = g \cdot (k \cdot x) = g \cdot x = y$.

2. We assume $\mathrm{Orb}(x) = \mathrm{Orb}(y)$, then there is a $g \in G$ such as $y = g \cdot x \iff g^{-1}y = x$. Now suppose $h \in \mathrm{Stab}(x)$ and observe that:

$$(ghg^{-1}) \cdot y = (gh) \cdot (g^{-1} \cdot y) = (gh) \cdot x = g \cdot (h \cdot x) = g \cdot x = y.$$

Consequently, $g\,\mathrm{Stab(x)}g^{-1} \in \mathrm{Stab}(y)$, i.e. $g\,\mathrm{Stab(x)}g^{-1} \subseteq \mathrm{Stab}(y)$. By interchanging the roles of x and y, we find the opposite inclusion $\mathrm{Stab}(y) \subseteq g\,\mathrm{Stab}(x)\,g^{-1}$, so $g\,\mathrm{Stab}(x)\,g^{-1} = \mathrm{Stab}(y)$.
□

If X is G-homogeneous, then $\mathrm{Orb}(x) = X$ for all $x \in X$, thus the orbit-stabilizer theorem implies the following, fundamental, result.

Corollary 10.1.1 *If X is a G-homogeneous space, then, for any fixed $x \in X$:*

1. *the map*
$$\begin{array}{ccc} G \big/ _{\mathrm{Stab}(x)} & \xrightarrow{\sim} & X \\ g\,\mathrm{Stab}(x) & \longmapsto & g \cdot x \end{array} \tag{10.2}$$
is bijective, i.e.
$$X \cong G \big/ _{\mathrm{Stab}(x)}, \tag{10.3}$$
*so, **every G-homogeneous space can be identified with a suitable set of transformations**.*

2. *the stabilizers of all elements of X are conjugated, and thus isomorphic, to each other.*

In Figure 10.1.1 we provide a graphical interpretation of a homogeneous space.

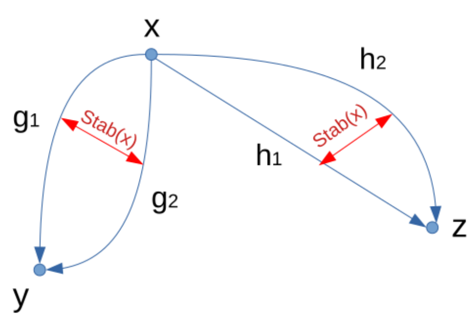


Figure 10.1: Fixed $x \in X$, every other element in X can be viewed as a transformation acting on x, i.e we identify y with all the transformations of G that allow us to pass from x to y modulo the transformations of the stabilizer in x. In the picture $g_2 = g_1 k$, with $k \in \mathrm{Stab}(x)$.

Example 10.1.4 *The straight lines in $\mathbb{R}^n$ are $\mathbb{R}$-homogeneous spaces.*

Let $L = \{u_0 + \lambda v : \lambda \in \mathbb{R}\}$ be the straight line in $\mathbb{R}^n$ passing through u_0 with direction v, u_0 and v are fixed in $\mathbb{R}^n$. Then the group $(\mathbb{R}, +)$ operates transitively on the set L via the action
$$\begin{array}{rccc} \eta_v : & \mathbb{R} \times L & \longrightarrow & L \\ & (\lambda, u) & \longmapsto & \eta_v(\lambda, u) := u + \lambda v. \end{array}$$
Of course, the stabilizer at any point of L is $\{0\}$ because any other $\lambda \neq 0$ will modify the vector u on L. The orbit-stabilizer theorem gives us the bijection $L \cong \mathbb{R} \big/ _{\{0\}} \cong \mathbb{R}$.

Remark 10.1.1 If X a G-homogeneous space, it is often interesting to search for a subgroup H of G whose action on X is still transitive and whose stabilizer is reduced to the unit element 1_G. If such a subgroup exists, then all the equivalence classes that compose the quotient group w.r.t. H are reduced to a single representative and so the orbit-stabilizer group implies that
$$H \cong X, \quad H \text{ transitive on } X \text{ with trivial stabilizer.}$$

10.2 Linear transformation groups and spheres

We remind the definitions of the real and complex linear transformation groups. The symbol $\langle\ ,\ \rangle$ will denote the real or complex Euclidean scalar product, respectively.

Def. 10.2.1 ***(Real matrix groups)***

- $GL(n,\mathbb{R}) = \{n \times n \text{ real invertible matrix}\}$ *(general linear group)*
- $GL^+(2,\mathbb{R}) = \{g \in GL(2,\mathbb{R})\ :\ \det(g) > 0\}$
- $SL(n,\mathbb{R}) = \{g \in GL(n,\mathbb{R}) : \det(g) = 1\}$ *(special linear group)*
- $O(n) = \{g \in GL(n,\mathbb{R}) : \forall x, y \in \mathbb{R}^n,\ \langle gx, gy\rangle = \langle x, y\rangle\}$ *(orthogonal group)*
- $SO(n) = \{g \in O(n) : \det(g) = 1\}$ *(special orthogonal group).*

Def. 10.2.2 ***(Complex matrix groups)***

- $U(n) = \{g \in GL(n,\mathbb{C}) : \forall x, y \in \mathbb{C}^n,\ \langle gx, gy\rangle = \langle x, y\rangle\}$ *(unitary group)*
- $SU(n) = \{g \in U(n) : \det(g) = 1\}$ *(special unitary group).*

We also remind that $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\} \subset \mathbb{R}^n$ is the $(n-1)$-dimensional sphere in $\mathbb{R}^n$ and $S^{2n-1} = \{z \in \mathbb{C}^n : \|z\| = 1\} \subset \mathbb{C}^n$ is the $(2n-1)$-dimensional real sphere in $\mathbb{C}^n$.

Remark 10.2.1 In finite dimension, a $n \times n$ matrix (real or complex) corresponds to a linear applications $f : E \longmapsto E$ with $E = \mathbb{R}^n$ or $\mathbb{C}^n$. However even if $\mathbb{C}^n \simeq \mathbb{R}^{2n}$, one should not mix up $\mathbb{R}$-linear and $\mathbb{C}$-linear maps. A classical counter example is provided by $f : \mathbb{C} \to \mathbb{C}$, $z \mapsto \overline{z}$, which is $\mathbb{R}$-linear but not $\mathbb{C}$-linear.

In the case of orthogonal and unitary group, we have the equivalent definitions :

$$\begin{aligned} g \in O(n,\mathbb{R}) &\iff g^t g = Id_n \\ g \in U(n) &\iff g^\dagger g = Id_n, \end{aligned}$$

where $g^\dagger = \overline{g}^t$ is the adjoint of g. This is easily shown by noticing that

$$\begin{aligned} \langle g^t g x, y\rangle = \langle gx, gy\rangle = \langle x, y\rangle \quad \forall x, y \in \mathbb{R}^n &\iff \langle g^t g x - x, y\rangle = 0 \quad \forall x, y \in \mathbb{R}^n \\ &\iff g^t g x = x \quad \forall x \in \mathbb{R}^n \\ &\iff g^t g = I_n, \end{aligned}$$

and equivalently in the case of the unitary group.

We introduce next some non-Euclidean transformation groups based on the Lorentzian product.

Def. 10.2.3 *We define the* ***Lorentzian (or Minkowski) scalar product*** *on* $\mathbb{R}^{n+1}$ *and* $\mathbb{C}^{n+1}$ *as :*

$$\begin{aligned} \langle x, y\rangle_L &= \sum_{i=1}^{n} x_i y_i - x_{n+1} y_{n+1} = \langle \tilde{x}, \tilde{y}\rangle - x_{n+1} y_{n+1} \qquad x, y \in \mathbb{R}^{n+1} \\ &= \sum_{i=1}^{n} x_i \overline{y_i} - x_{n+1} \overline{y_{n+1}} = \langle \tilde{x}, \tilde{y}\rangle - x_{n+1} y_{n+1} \qquad x, y \in \mathbb{C}^{n+1}, \end{aligned}$$

where $\tilde{x} = (x_1, \ldots, x_n)^t$, $\tilde{y} = (y_1, \ldots, y_n)^t$ and $\langle\ ,\ \rangle$ is the Euclidean product in $\mathbb{R}^n$ or $\mathbb{C}^n$.

The linear groups of signature $(n, 1)$, also called ***Lorentzian signature****, are the following:*

- $O(n,1) = \{g \in GL(n+1, \mathbb{R}) : \forall x, y \in \mathbb{R}^n,\ \ \langle gx, gy\rangle_L = \langle x, y\rangle_L\}$
- $SO(n,1) = \{g \in O(n,1) :\ \ \det(g) = 1\}$
- $U(n,1) = \{g \in GL(n+1, \mathbb{C}) : \forall x, y \in \mathbb{C}^n,\ \ \langle gx, gy\rangle_L = \langle x, y\rangle_L\}$
- $SU(n,1) = \{g \in U(n,1) :\ \ \det(g) = 1\}$.

The Lorentzian scalar product can be defined by using the Euclidean scalar product by noticing that it holds:

$$\langle x, y\rangle_L = \langle \eta x, y\rangle \qquad \text{with} \qquad \eta = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix}. \tag{10.4}$$

Similarly, the orthogonal and unitary group of signature $(n, 1)$ can be re-defined through the conditions below:

$$g \in M(n+1, \mathbb{R}),\ \ g \in O(n,1) \quad \Longleftrightarrow \quad g^t \eta g = \eta \tag{10.5}$$

$$g \in M(n+1, \mathbb{C}),\ \ g \in U(n,1) \quad \Longleftrightarrow \quad g^\dagger \eta g = \eta. \tag{10.6}$$

Thanks to Binet's theorem, for all matrices $g \in O(n)$, $O(n,1)$, $U(n)$ or $U(n,1)$, it holds that $|\det(g)| = 1$.

10.3 Homogeneity of spheres under the group of rotations

The simplest and most intuitive homogeneous spaces are represented by spheres. We have already seen that the circle S^1 is homogeneous under the action of rotations, $SO(2)$ in the real case, $U(1)$ in the complex one. In what follows, we shall see that this result can be extended to higher dimensions.

Notation: in the whole section $(e_j)_{j=1,..,n}$ will denote the canonical basis of $\mathbb{R}^n$ or $\mathbb{C}^n$. Clearly, each e_j belongs to the sphere S^{n-1} since their Euclidean norm is 1.

10.3.1 Spheres in $\mathbb{R}^n$

Before starting, it is worth mentioning that the action of $GL(n, \mathbb{R})$ and its subgroups on $\mathbb{R}^n$ will be the usual matrix multiplication in this section. This will not always be the case in homogeneous spaces, as we will see later on.

Theorem 10.3.1 *Let $n \geqslant 2$.*

1. *S^{n-1} is $SO(n)$-homogeneous*
2. $\mathrm{Stab}(e_n) = \left\{ \begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix} : h \in SO(n-1) \right\} \cong SO(n-1)$

By the orbit-stabilizer theorem, we get:

$$S^{n-1} \cong SO(n) \diagup_{SO(n-1)} \Longleftrightarrow\ S^{n-1} \cong \left\{ \left\{ g \begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix} : h \in SO(n-1) \right\},\ g \in SO(n) \right\}.$$

Proof.

1. To prove that $SO(n)$ operates transitively on S^{n-1}, we have to show that it exists at least one element of S^{n-1} that can be connected to all the other elements of S^{n-1} via transformations of $SO(n)$. We are going to show that this element is e_1, i.e. that $\forall x \in S^{n-1}\ \exists g \in SO(n)$ such that $ge_1 = x$.

Fixed $x \in S^{n-1} \subset \mathbb{R}^n$, thanks to the Gram-Schmidt orthonormalization procedure, we can find $x_2, x_3, \ldots, x_n \in S^{n-1} \subset \mathbb{R}^n$ such that $(x, x_2, \ldots, x_n)$ is an orthonormal basis for $\mathbb{R}^n$.

If we use the vectors $(x, x_2, \ldots, x_n)$ as columns of a matrix A, then we know that $A \in O(n)$ and that $\det(A) = \pm 1$. To guarantee a determinant equal to 1, we slightly modify A by considering the matrix

$$g = \begin{pmatrix} | & | & & | \\ x & x_2 & \ldots & \epsilon x_n \\ | & | & & | \end{pmatrix},$$

with $\epsilon = \pm 1$ chosen such that $\det(g) = 1$, in this way $g \in SO(n)$. By direct computation we get $ge_1 = x$, but x was arbitrarily chosen in S^{n-1}, so the action of $SO(n)$ is transitive on S^{n-1}. Notice that the matrix g depends on x because the Gram-Schmidt orthonormalization is initiated by x itself.

2. Let us search under which conditions it is possible to build a matrix such that

$$g = \begin{pmatrix} h & b \\ c & d \end{pmatrix} \in SO(n),$$

with $h \in M(n-1, \mathbb{R})$, $b \in M((n-1) \times 1, \mathbb{R})$, $c \in M(1 \times (n-1), \mathbb{R})$ and $d \in \mathbb{R}$ that satisfies $ge_n = e_n$. The set of these matrices will give $\mathrm{Stab}(e_n)$.

First of all we notice that $g \in SO(n) \iff g^{-1} = g^t$, thus $ge_n = e_n \implies e_n = g^t e_n$. By direct computation we have:

- $ge_n = \begin{pmatrix} b_1 \\ \vdots \\ b_{n-1} \\ d \end{pmatrix}$, so $ge_n = e_n \iff \begin{pmatrix} b_1 \\ \vdots \\ b_{n-1} \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \iff b = 0_{M((n-1)\times 1)}$ and $d = 1$

- $g^t e_n = \begin{pmatrix} c_1 \\ \vdots \\ c_{n-1} \\ d \end{pmatrix}$, so $e_n = g^t e_n \iff \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_{n-1} \\ d \end{pmatrix} \iff c = 0_{M(1\times(n-1))}$ and $d = 1$.

Hence, the desired matrix is $g = \begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix}$, which belongs to $SO(n)$ if and only if $h \in SO(n-1)$, i.e. $\det(h) = 1$ and $h^t h = I_{n-1}$. This shows that $\mathrm{Stab}(e_n) \cong SO(n-1)$ and, since all stabilizers of a homogeneous space are isomorphic to each other, $S^{n-1} \cong SO(n)/SO(n-1)$. □

10.3.2 Spheres in $\mathbb{C}^n$

The results and proofs are the nearly identical for spheres in $\mathbb{C}^n$ as for those in $\mathbb{R}^n$. The only difference is that we need take some precautions with the determinant of $\mathbb{C}$-linear applications.

Theorem 10.3.2 *Let $n \geqslant 2$.*

1. *S^{2n-1} is $SU(n)$-homogeneous*

2. *$Stab(e_n) = \left\{ \begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix} : h \in SU(n-1) \right\} \cong SU(n-1)$.*

By the orbit-stabilizer theorem, we get:

$$S^{2n-1} \cong SU(n) \diagup_{SU(n-1)} \iff S^{2n-1} \cong \left\{ \left\{ g \begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix} : h \in SU(n-1) \right\},\ g \in SU(n) \right\}.$$

Proof.

1. Fix an arbitrary $z \in S^{2n-1} \subset \mathbb{C}^n$ and apply again the Gram-Schmidt orthonormalization procedure to find $z_2, \ldots, z_n \in S^{2n-1} \subset \mathbb{C}^n$ such that $(z, z_2, \ldots, z_n)$ is an orthonormal basis for $\mathbb{C}^n$. Also, let $g_\theta \in U(n)$,

$$g_\theta = \begin{pmatrix} | & | & \cdots & | \\ z & z_2 & \cdots & e^{i\theta} z_n \\ | & | & \cdots & | \end{pmatrix}, \quad \theta \in [0, 2\pi).$$

Then we have again $g_\theta e_1 = z$, for all $\theta \in [0, 2\pi)$ and, thanks to the properties of the determinant, $\det(g_\theta) = e^{i\theta} \det(g_0) \in S^1$, where $\det(g_0) = e^{i\varphi}$. We finish by choosing $\theta = -\varphi$ so that $\det(g_\theta) = e^{-i\varphi} e^{i\varphi} = 1$, in order to have $g_\theta \in SU(n)$.

2. Exactly the same proof as in the real case, we simply need to replace the transpose matrix by the adjoint matrix and $SO(n)$ by $SU(n)$. □

Remark 10.3.1 Since all n-spheres of different radius are isomorphic, we have exactly the same results for spheres of positive radius, $S^n_R = \{x \in \mathbb{R}^{n+1} : \|x\| = R\}$. This fact will be useful later.

10.4 Homogeneity of the open unit ball: relationship between projective spaces and hyperbolic rotations

We have seen that the contour of the unit disk in $\mathbb{R}^2$ and $\mathbb{C}$ is a homogeneous space under the rotation group $SO(2)$ and $U(1)$, respectively, but that the unit disks $D_{\mathbb{R}}(0,1)$ and $D_{\mathbb{C}}(0,1)$ are not homogeneous under the action of these groups.

In this section we are going to show that $D_{\mathbb{R}}(0,1)$ and $D_{\mathbb{C}}(0,1)$ and, more generally, the open unit ball in $\mathbb{R}^n$ and $\mathbb{C}^n$, are homogeneous spaces w.r.t. the groups $SO(n,1)$ and $SU(n,1)$, the Lorentzian analogues of $SO(n)$ and $SU(n)$, whose action is implemented by *hyperbolic rotations*. In order to show this, it is useful to embed $\mathbb{R}^n$ and $\mathbb{C}^n$ in the real or complex projective space, respectively. For this reason, we begin by discussing the action of the general linear group on projective spaces.

10.4.1 The action of group $GL(n+1,\mathbb{R})$ on $\mathbb{RP}^n$

We have already seen in chapter 1 the real projective space[1]

$$\mathbb{RP}^n = \mathbb{R}^{n+1}\backslash\{0\} \ \big/_{\mathbb{R}^\times} \equiv \left\{ \left\{ \lambda \begin{pmatrix} u_1 \\ \vdots \\ u_{n+1} \end{pmatrix}, \ \lambda \neq 0 \right\}, \ \begin{pmatrix} u_1 \\ \vdots \\ u_{n+1} \end{pmatrix} \neq 0 \right\},$$

its twin brother is the complex projective space:

$$\mathbb{CP}^n = \mathbb{C}^{n+1}\backslash\{0\} \ \big/_{\mathbb{C}^\times},$$

$\mathbb{R}^\times$ and $\mathbb{C}^\times$ being $\mathbb{R}$ and $\mathbb{C}$ without their 0 element. Here we are going analyze more thoroughly the projective space, for the sake of a smoother reading, we will fix our attention only on the real projective space, knowing that everything we will write in this subsection also holds true for the complex projective space, simply by replacing $\mathbb{R}$ with $\mathbb{C}$ and $\mathbb{R}^\times$ with $\mathbb{C}^\times$.

Notation: in this section, the equivalence class $\mathbb{R}^\times \cdot u \in \mathbb{RP}^n$, $u \in \mathbb{R}^{n+1}\backslash\{0\}$, will be denoted by :

$$[u] = \begin{bmatrix} u_1 \\ \vdots \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} \lambda u_1 \\ \vdots \\ \lambda u_{n+1} \end{bmatrix}.$$

Notice now that, for every $u \in \mathbb{R}^n\backslash\{0\}$, the following map[2] is clearly an injection of $\mathbb{R}^n$ into $\mathbb{RP}^n$:

$$\begin{array}{ccc} \mathbb{R}^n & \hookrightarrow & \mathbb{RP}^n \\ u & \mapsto & \begin{bmatrix} u \\ 1 \end{bmatrix}. \end{array} \tag{10.7}$$

Since every element $v \in \mathbb{RP}^n$ can be written as $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \\ v_{n+1} \end{bmatrix}$, we have either $v_{n+1} \neq 0$, and so $v = \begin{bmatrix} u \\ 1 \end{bmatrix}$, with $u \in \mathbb{R}^n$, or $v_{n+1} = 0$, and so $v = \begin{bmatrix} u \\ 0 \end{bmatrix}$, with $u \in \mathbb{R}^n\backslash\{0\}$. Thus, the injection (10.7) becomes a bijection between $\mathbb{R}^n$ and the set $\left\{ \begin{bmatrix} u \\ 1 \end{bmatrix} : u \in \mathbb{R}^n \right\}$. As a consequence, we can split the real projective space in the following disjoint union:

$$\begin{aligned} \mathbb{RP}^n &= \left\{ \begin{bmatrix} u \\ 1 \end{bmatrix} : u \in \mathbb{R}^n \right\} \sqcup \left\{ \begin{bmatrix} u \\ 0 \end{bmatrix} : u \in \mathbb{R}^n\backslash\{0\} \right\} \\ &\underset{(10.7)}{\cong} \mathbb{R}^n \sqcup \mathbb{RP}^{n-1}. \end{aligned}$$

Of course, we can iterate the splitting on the second set, obtaining:

$$\mathbb{RP}^n \cong \mathbb{R}^n \sqcup \mathbb{R}^{n-1} \sqcup \cdots \sqcup \mathbb{R}^1 \sqcup \mathbb{RP}^0.$$

[1] Geometrically speaking, we have seen in chapter 1 that the projective spaces is isomorphic to the set of straight lines passing through 0 in $\mathbb{R}^n$.

[2] The choice of setting to 1 the last coordinate is an arbitrary, yet usual, choice. Changing the position of 1 leads to an isomorphic decomposition in the following.

The description of $\mathbb{RP}^0$ deserves a special discussion: $\mathbb{RP}^0$ is the quotient space $\mathbb{R}\backslash\{0\} \diagup_{\mathbb{R}^\times}$, i.e.

$$\mathbb{RP}^0 = \{\{\lambda u,\ \lambda \neq 0\},\ u \neq 0\} \cong \{[1]\}$$

a set containing a single $\mathbb{R}^\times$-equivalence class, canonically chosen to be [1]. In the projective geometry literature, [1] is denoted with ∞ and called **the point at the infinite**. So, to resume:

$$\mathbb{RP}^n \cong \mathbb{R}^n \sqcup \mathbb{R}^{n-1} \sqcup \cdots \sqcup \mathbb{R}^1 \sqcup \{\infty\}. \tag{10.8}$$

We can now start with the definition of **the action of** $GL(n+1,\mathbb{R})$ **on** $\mathbb{RP}^n$:

$$\begin{array}{ccc} GL(n+1,\mathbb{R}) \times \mathbb{RP}^n & \longrightarrow & \mathbb{RP}^n \\ (g,[u]) & \mapsto & g\cdot[u] := [gu], \end{array}$$

which is well-defined because, thanks to the $\mathbb{R}$-linearity of g, for all $\lambda \in \mathbb{R}^\times$ we have:

$$g\cdot[\lambda u] = [g(\lambda u)] = [\lambda(gu)] = [gu] = g\cdot[u],$$

so the choice of the representative u in a class in $\mathbb{RP}^n$ does not impact the action.

However, we notice that this action is not stable when restricted on $\mathbb{R}^n$, interpreted as a subset of $\mathbb{RP}^n$ via the injection (10.7). To see this, take

$$g = \begin{pmatrix} A & b \\ c & d \end{pmatrix} \in GL(n+1,\mathbb{R})$$

with $A \in GL(n,\mathbb{R})$, $c \in M(1\times n,\mathbb{R})$, $b \in M(n\times 1,\mathbb{R})$ and $d \in \mathbb{R}$, then, by direct computation[3]:

$$g\cdot u \underset{(10.7)}{\cong} g\cdot\begin{bmatrix} u \\ 1 \end{bmatrix} = \begin{bmatrix} Au+b \\ cu+d \end{bmatrix} = \begin{bmatrix} \frac{Au+b}{cu+d} \\ 1 \end{bmatrix} \underset{(10.7)}{\cong} \frac{Au+b}{cu+d} \in \mathbb{R}^n \iff cu+d \neq 0,$$

however, not all the matrices of $GL(n+1,\mathbb{R})$ satisfy the constraint $cu+d \neq 0$, e.g. for all $u \in \mathbb{R}^n\backslash\{0\}$, the matrix $g = \begin{pmatrix} A & 0 \\ \frac{u^t}{\|u\|^2} & -1 \end{pmatrix}$ with $A \in GL(n,\mathbb{R})$ belongs to $GL(n+1,\mathbb{R})$ but:

$$g\cdot u \underset{(10.7)}{\cong} g\cdot\begin{bmatrix} u \\ 1 \end{bmatrix} = \left[\begin{pmatrix} A & 0 \\ \frac{u^t}{\|u\|^2} & -1 \end{pmatrix}\begin{pmatrix} u \\ 1 \end{pmatrix}\right] = \begin{bmatrix} Au \\ \frac{u^t u}{\|u\|^2} - 1 \end{bmatrix} = \begin{bmatrix} Au \\ 0 \end{bmatrix} \in \mathbb{RP}^{n-1} \neq \mathbb{R}^n.$$

10.4.2 Homogeneity of the open unit ball in $\mathbb{R}^n$

Even if the action $GL(n+1,\mathbb{R})$ is not stable when operating on $\mathbb{R}^n$, its subgroup $SO(n,1)$ acts in a stable way on the unit ball

$$B := B_{\mathbb{R}}(0,1) = \{x \in \mathbb{R}^n \ :\ \langle x,x\rangle = \|x\|^2 < 1\} \subset \mathbb{R}^n.$$

Even more, the action of $SO(n,1)$ is transitive on B. To prove this result, it is useful to show that we can find a copy of the unit ball in $\mathbb{RP}^n$.

[3] $cu \in \mathbb{R}$ because it is the matrix product of $c \in M(1\times n,\mathbb{R})$ and u interpreted as an element of $M(n\times 1,\mathbb{R})$, so cu is nothing but the Euclidean scalar product $\langle c,u\rangle$ if we interpret both c and u as column vectors of $\mathbb{R}^n$.

Theorem 10.4.1 *The unit ball B in $\mathbb{R}^n$ can be characterized as follows:*

$$B \cong B' := \{[u] \in \mathbb{RP}^n \;:\; \langle u, u\rangle_L < 0\} \subset \mathbb{RP}^n.$$

Proof. We start by noting that the constraint that defines B', i.e. $\langle u, u\rangle_L < 0$, is well-defined in $\mathbb{RP}^n$, in fact, for each $[u] \in \mathbb{RP}^n$ and $\lambda \in \mathbb{R}^\times$, $\langle u, u\rangle_L < 0 \iff \langle \lambda u, \lambda u\rangle_L = \lambda^2\langle u, u\rangle_L < 0$.

Now, let $v \in B \subset \mathbb{R}^n$, i.e. $\langle v, v\rangle < 1$, and let $\begin{bmatrix} v \\ 1 \end{bmatrix}$ be its copy in $\mathbb{RP}^n$, then

$$\langle \begin{bmatrix} v \\ 1 \end{bmatrix}, \begin{bmatrix} v \\ 1 \end{bmatrix} \rangle_L := \langle v, v\rangle - 1 < 0.$$

Conversely, let $[u] = \begin{bmatrix} \bar{u} \\ u_{n+1} \end{bmatrix} \in B'$, i.e. $\langle u, u\rangle < 0$, with $\bar{u} \in \mathbb{R}^n$, then

$$\langle \begin{pmatrix} \bar{u} \\ u_{n+1} \end{pmatrix}, \begin{pmatrix} \bar{u} \\ u_{n+1} \end{pmatrix} \rangle_L = \langle \bar{u}, \bar{u}\rangle - u_{n+1}^2 < 0, \tag{10.9}$$

which implies that u_{n+1} must be different than 0, given that $\langle \bar{u}, \bar{u}\rangle = \|\bar{u}\| \geqslant 0$. So,

$$\begin{bmatrix} \bar{u} \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} \frac{\bar{u}}{u_{n+1}} \\ 1 \end{bmatrix} \underset{(10.7)}{\cong} \frac{\bar{u}}{u_{n+1}} \in \mathbb{R}^n.$$

To verify that $\frac{\bar{u}}{u_{n+1}} \in B$ we notice that (10.9) implies that $\|\bar{u}\|^2 < u_{n+1}^2$, i.e. $\|\bar{u}\| < |u_{n+1}|$ so $\left\|\frac{\bar{u}}{u_{n+1}}\right\| = \frac{\|\bar{u}\|}{|u_{n+1}|} < 1$, thus $\frac{\bar{u}}{u_{n+1}} \in B$. $\square$

Theorem 10.4.1 implies that **the unit ball in $\mathbb{R}^n$ can be identified with the (double) cone in $\mathbb{R}^{n+1}$ obtained as the set of straight lines passing through the origin of $\mathbb{R}^{n+1}$ and with slope strictly smaller than** 1. Figure 10.2 gives a pictorial illustration of this cone.

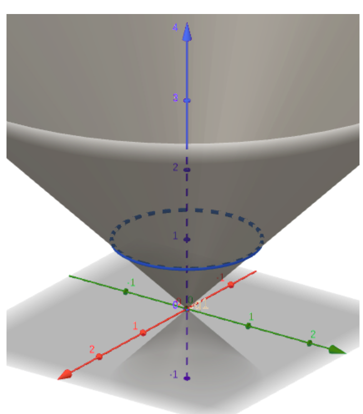


Figure 10.2: The double cone in $\mathbb{R}^{n+1}$ in bijection with the unit ball in $\mathbb{R}^n$.

Until the end of this section, the open unit ball B in $\mathbb{R}^n$ will be identified with its copy in $\mathbb{RP}^n$ as defined by the previous theorem.

The **action of $O(n,1)$ on B** is:

$$\begin{array}{ccc} O(n,1) \times B & \longrightarrow & B \\ (g,[u]) & \mapsto & g \cdot [u] := [gu], \end{array}$$

well-defined because the action of $O(n,1)$ is stable on the elements of B since the matrices belonging to $O(n,1)$ preserve the Lorentzian product, so, for all $u \in B$ and $g \in O(n,1)$,

$$\langle gu, gu \rangle_L = \langle u, u \rangle_L < 0.$$

It turns out that the subgroup $SO(n,1)$ is enough to guarantee a transitive action on B. A couple of preliminary results will help us prove this result quite easily.

Lemma 10.4.1 *Let $a \in SO(n)$, then $\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \in SO(n,1)$.*

Proof. From eq. (10.5) we know that, given $\eta = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \in SO(n,1)$ if and only if its determinant is 1, which is true, and if

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}^t \eta \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \eta \iff \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}^t \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a^t a & 0 \\ 0 & -1 \end{pmatrix} \underset{a^t a = I_n}{=} \eta.$$

Therefore, $\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \in SO(n,1)$. □

Lemma 10.4.2 *Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SO(1,1)$, then $\begin{pmatrix} a & 0 & b \\ 0 & I_{n-1} & 0 \\ c & 0 & d \end{pmatrix} \in SO(n,1)$.*

Proof. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SO(1,1)$, then by eq. (10.5) it holds that

$$g^t \eta g = \eta \iff \begin{pmatrix} a^2 - c^2 & ab - cd \\ ab - cd & b^2 - d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{10.10}$$

and so, if we set $h = \begin{pmatrix} a & 0 & b \\ 0 & I_{n-1} & 0 \\ c & 0 & d \end{pmatrix}$, then $\det(h) = \det(g) = 1$ and

$$h^t \eta h = \begin{pmatrix} a^2 - c^2 & 0 & ab - cd \\ 0 & I_{n-1} & 0 \\ ab - cd & 0 & b^2 - d^2 \end{pmatrix} \underset{(10.10)}{=} \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix},$$

so $h \in SO(n,1)$. □

Theorem 10.4.2 *Let $n \geqslant 2$.*

1. *The action of $SO(n,1)$ on B is transitive*
2. $\mathrm{Stab}(0) = \left\{ \begin{pmatrix} a & 0 \\ 0 & \det(a) \end{pmatrix},\ a \in O(n) \right\} \cong O(n).$

Therefore, the stabilizer-orbit theorem implies:

$$B \cong SO(n,1) \diagup_{O(n)} \iff B \cong \left\{ \left\{ g \begin{pmatrix} a & 0 \\ 0 & \det(a) \end{pmatrix},\ a \in O(n) \right\},\ g \in SO(n,1) \right\}.$$

Proof.

1. Let $x \in B$ arbitrary, $x \neq 0$. We wish to show that there is a $g \in SO(n,1)$ such that $g \cdot x = 0$. We will do this following this path:

- we search for $g_1 \in SO(n,1)$ such that $g_1 \cdot x = \begin{pmatrix} |x| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$
- we search for $g_2 \in SO(n,1)$ such that $g_2 \cdot \begin{pmatrix} |x| \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0$
- finally, we set $g = g_2 g_1$ to get the wanted result : $g \cdot x = 0$.

Let $r = |x|$. Since $SO(n)$ is transitive on the sphere S_r^{n-1}, there exists $a \in SO(n)$ such that $a \cdot x = \begin{pmatrix} r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. We then define $g_1 = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$, which belongs to $SO(n,1)$ thanks to Lemma 10.4.1, then:

$$g_1 \cdot x = \left[\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{pmatrix} \right] = \begin{bmatrix} r \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \begin{pmatrix} r \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Next, it can be verified with straightforward computations that the matrix

$$\tilde{g}_2 = \frac{1}{\sqrt{1-r^2}} \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix} \in SO(1,1)$$

verifies

$$\tilde{g}_2 \cdot \begin{bmatrix} r \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{ar+b}{cr+b} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0.$$

We now use Lemma 10.4.2 to extend $\tilde{g}_2$ to $SO(n,1)$ with

$$g_2 = \frac{1}{\sqrt{1-r^2}} \begin{pmatrix} 1 & 0 & -r \\ 0 & I_{n-1} & 0 \\ -r & 0 & 1 \end{pmatrix},$$

with g_2 having the desired property: $g_2 \cdot \begin{pmatrix} r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0$. Finally, if $g = g_2 g_1$, we get $g \cdot x = 0$.

2. Let $g = \begin{pmatrix} A & b \\ c & d \end{pmatrix} \in SO(n,1)$, with A a $n \times n$ matrix such that $g \cdot 0 = 0$. Thus,

$$g \cdot 0 = 0 \iff \begin{pmatrix} A & b \\ c & d \end{pmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix} = \frac{b}{d} = 0 \iff b = 0,$$

$d \neq 0$ since this would go against the already established stability of $SO(n,1)$ on B.

Moreover, $g \in O(n,1)$, thus:

$$\begin{aligned} g^t \eta g = \eta \quad &\iff \quad \begin{pmatrix} A^tA - c^tc & -c^td \\ -dc & -d^2 \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix} \\ &\iff \quad c = 0\,, \quad A^tA = I_n \quad \text{and} \quad d^2 = 1, \end{aligned}$$

therefore $A \in O(n)$. Finally, since $g \in SO(n,1)$, $\det(g) = d\det(A) = 1$, thus $d = \det(A)^{-1} = \det(A^{-1}) \underset{A \in O(n)}{=} \det(A^t) = \det(A)$. This concludes the proof since we have proven that:

$$\mathrm{Stab}(0) = \left\{ \begin{pmatrix} A & 0 \\ 0 & \det(A) \end{pmatrix} : A \in O(n) \right\} \cong O(n).$$

□

10.4.3 Homogeneity of the open unit ball in $\mathbb{C}^n$

To extend the previous results to the open unit ball in $\mathbb{C}^n$ we just need replace $O(n)$ by $U(n)$, the proofs are practically identical to the real case.

Once again, we can identify the complex open unit ball $B \subset \mathbb{C}^n$ with elements of $[u] \in \mathbb{CP}^n$ such that $\langle u, u \rangle_L < 0$ and the action of $U(n,1)$ on B is stable.

Theorem 10.4.3 *Let $n \geqslant 2$.*

1. *The action of $SU(n,1)$ is transitive on $B \subset \mathbb{C}^n$*
2. $\mathrm{Stab}(0) = \left\{ \begin{pmatrix} A & 0 \\ 0 & \overline{\det(A)} \end{pmatrix},\ A \in U(n) \right\} \cong U(n).$

Therefore, by the stabilizer-orbit theorem:

$$B \cong SU(n,1) \diagup_{U(n)} \iff B \cong \left\{ \left\{ g \begin{pmatrix} A & 0 \\ 0 & \overline{\det(A)} \end{pmatrix},\ A \in U(n) \right\},\ g \in SU(n,1) \right\}.$$

10.5 Homogeneity of the upper-half plane H

The upper half plane H is another very important example of homogeneous space. To prove this result we first need to introduce and discuss the Möbius transformations. The most general definition of **Möbius transformations** in two dimensions is given in the context of the action of $GL(2,\mathbb{C})$ on $\mathbb{CP}^1$, that is convenient to write through the splitting

$$\mathbb{CP}^1 \cong \mathbb{C} \sqcup \{\infty\} = \left\{ \begin{bmatrix} z \\ 1 \end{bmatrix} \ : \ z \in \mathbb{C} \right\} \sqcup \left\{ \begin{bmatrix} z \\ 0 \end{bmatrix} \ : \ z \in \mathbb{C}\backslash\{0\} \right\} \tag{10.11}$$

called the **Riemann sphere**. For all $z, w \in \mathbb{C}$, the action

$$\begin{array}{ccc} GL(2,\mathbb{C}) \times \mathbb{C} \sqcup \{\infty\} & \longrightarrow & \mathbb{C} \sqcup \{\infty\} \\ \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{bmatrix} z \\ w \end{bmatrix} \right) & \longmapsto & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{bmatrix} z \\ w \end{bmatrix}, \end{array}$$

is defined by considering two cases corresponding to $w \neq 0$ and $w = 0$, respectively.
In the first case, i.e. for all $z \in \mathbb{C}$ and $w \neq 0$, we have:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{bmatrix} z \\ 1 \end{bmatrix} = \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} az+b \\ cz+d \end{pmatrix} \right] = \begin{bmatrix} az+b \\ cz+d \end{bmatrix} = \begin{cases} \begin{bmatrix} \frac{az+b}{cz+d} \\ 1 \end{bmatrix} & \text{if} \quad cz+d \neq 0 \\ \begin{bmatrix} az+b \\ 0 \end{bmatrix} & \text{if} \quad cz+d = 0 \end{cases}$$

$$\cong \begin{cases} \frac{az+b}{cz+d} \in \mathbb{C} & \text{if} \quad cz+d \neq 0 \\ \infty \in \{\infty\} & \text{if} \quad cz+d = 0 \end{cases}.$$

In the second case, i.e. for all $z \in \mathbb{C}$ and $w \neq 0$, we have:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{bmatrix} z \\ 0 \end{bmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right] = \begin{bmatrix} a \\ c \end{bmatrix} = \begin{cases} \frac{a}{c} \in \mathbb{C} & \text{if } c \neq 0 \\ \infty \in \{\infty\} & \text{if } c = 0 \end{cases}.$$

This very general definition of Möbius transformations is not needed to show that H is a homogeneous space, in fact, we can restrict our attention to the much simpler action of the group $\mathrm{SL}(2,\mathbb{R})$ on H to obtain this result, as we discuss in the next subsection.

10.5.1 Möbius transformations on the upper-half plane H

In this section, $H = \{z \in \mathbb{C} \ : \ \mathfrak{Im}(z) > 0\} = \{(x+iy) \in \mathbb{C} \ : \ y > 0\}$ will denote the upper half plane in $\mathbb{C}$. When we consider H, the Möbius transformations acquire a much simpler form as it is stated in the following result.

Lemma 10.5.1 *The* ***Möbius action***

$$\begin{array}{rccc} \mathscr{M}: & GL^+(2,\mathbb{R}) \times H & \longrightarrow & H \\ & \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, z \right) & \longmapsto & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az+b}{cz+d} \end{array},$$

is an actual group action on H.

Proof. First of all, we show that the operation is stable. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL^+(2, \mathbb{R})$, so that $\det(g) = ad - bc > 0$, and $z = x + iy \in H$, so that $y > 0$. Then,

- $cz + d = 0 \iff c(x + iy) = -d \iff cx = -d$ and $y = 0$, which cannot happen because $y > 0$, thus the denominator of the Möbius transformations is always different than 0 in the whole H.
- $\mathfrak{Im}(g \cdot z) = \mathfrak{Im}\left(\frac{az+b}{cz+d}\right) = \frac{1}{|cz+d|^2}\mathfrak{Im}((az+b)(c\overline{z}+d)) = \frac{1}{|cz+d|^2}\mathfrak{Im}(iy(ad-bc)) = \frac{y(ad-bc)}{|cz+d|^2} = \frac{y\det(g)}{|cz+d|^2} > 0.$

Hence, the Möbius transformation defined above is stable on H. We now need to verify the properties of group action.

1. If $g = I_n$, then $a = d = 1$, $b = c = 0$, so $I_n \cdot z = \frac{z+0}{0+1} = z$.
2. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL^+(2, \mathbb{R})$ and $h = \begin{pmatrix} k & l \\ m & n \end{pmatrix} \in GL^+(2, \mathbb{R})$, then $gh = \begin{pmatrix} ak + bm & al + bn \\ ck + dm & cl + dn \end{pmatrix}$ and:
   - $(gh) \cdot z = \frac{(ak+bm)z+(al+bn)}{(ck+dm)z+(cl+dn)}$
   - $g \cdot (h \cdot z) = \frac{a\frac{kz+l}{mz+n}+b}{c\frac{kz+l}{mz+n}+d} = \frac{a(kz+l)+b(mz+n)}{c(kz+l)+d(mz+n)} = \frac{(ak+bm)z+(al+bn)}{(ck+dm)z+(cl+dn)}.$

   Hence, $(gh) \cdot z = g \cdot (h \cdot z)$ for all $z \in H$. □

$\mathrm{GL}^+(2, \mathbb{R})$ is the maximal stability group for the Möbius action on H, since $\mathrm{SL}(2, \mathbb{R})$ is a subgroup of $\mathrm{GL}^+(2, \mathbb{R})$, we get that also the Möbius action on H restricted to the matrices of $\mathrm{SL}(2, \mathbb{R})$ is an actual group action.

It turns out that the $\mathrm{SL}(2, \mathbb{R})$ Möbius action on H is enough to guarantee transitivity. Compared to the proof of other homogeneous spaces that we have discussed so far, the $\mathrm{SL}(2, \mathbb{R})$-homogeneity of H is relatively easy to demonstrate.

Theorem 10.5.1 *The following statements hold:*

1. *the upper-half plane H is $SL(2, \mathbb{R})$-homogeneous*
2. $\mathrm{Stab}(i) = SO(2)$.

Thus, the stabilizer-orbit theorem implies:

$$H \cong SL(2, \mathbb{R}) \diagup_{SO(2)} \iff H \cong \{\{gh,\ h \in SO(2)\},\ g \in SL(2, \mathbb{R})\}.$$

Proof.

1. Let $z = x + iy \in H$ arbitrary, so $y > 0$ and the matrix $g = \begin{pmatrix} \sqrt{y} & \frac{x}{\sqrt{y}} \\ 0 & \frac{1}{\sqrt{y}} \end{pmatrix}$ belongs to $SL(2, \mathbb{R})$.

Then,

$$g \cdot i = \frac{\sqrt{y}i + x/\sqrt{y}}{0 \cdot i + 1/\sqrt{y}} = \sqrt{y}\left(i\sqrt{y} + \frac{x}{\sqrt{y}}\right) = x + iy = z.$$

Thus, the action of $SL(2,\mathbb{R})$ on H is transitive because the whole H can reached by i via a Möbius transformations.

2. Given $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{R})$, let us explicitly write the stabilization condition on i:

$$g \cdot i = i \iff \frac{ai+b}{ci+d} = i \iff ai + b = di - c \iff a = d \text{ and } b = -c.$$

Furthermore, if we fuse these equalities with the fact that $\det(g) = ad - bc = 1$, we get the constraint $a^2 + b^2 = 1$. Consequently, every matrix of $\mathrm{Stab}(i)$ is written as $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, with $a^2 + b^2 = 1$, which is the parameterization of a generic $SO(2,\mathbb{R})$ matrix, thus $\mathrm{Stab}(i) = SO(2)$. $\square$

Let $[g] = gSO(2,\mathbb{R})$ for a generic equivalence classes in $SL(2,\mathbb{R}) \diagup_{SO(2,\mathbb{R})}$ and let

$$\begin{array}{rccc} \varphi : & SL(2,\mathbb{R}) \diagup_{SO(2,\mathbb{R})} & \longrightarrow & H \\ & [g] & \longmapsto & g \cdot i \end{array}$$

be the bijection generated by the orbit-stabilizer theorem and $\mathscr{M}_g : H \to H$, $z \mapsto \mathscr{M}(g,z)$, for all $g \in SL(2,\mathbb{R})$, then the following diagram

$$\begin{array}{ccc} SL(2,\mathbb{R}) \diagup_{SO(2,\mathbb{R})} & \xrightarrow{\varphi} & H \\ \downarrow{\scriptstyle Id} & & \downarrow{\scriptstyle \mathscr{M}_g} \\ SL(2,\mathbb{R}) \diagup_{SO(2,\mathbb{R})} & \xrightarrow{\varphi} & H \end{array}$$

is commutative.

10.5.2 The isomorphism $H \cong \mathbf{Sym}_1^+(2,\mathbb{R})$

A very useful characterization of the upper half plane H is represented by the set (which is not a group) of matrices

$$\mathrm{Sym}_1^+(2,\mathbb{R}) = \{g \in SL(2,\mathbb{R}) : g = g^t,\ g \text{ positive definite: } u^t g u \geqslant 0\ \forall u \in \mathbb{R}^2\}.$$

To show this fact, we first need to recall a handy representation of the elements in $\mathrm{Sym}_1^+(2,\mathbb{R})$. In the proof of the theorem, (e_1, e_2) will denote the canonical basis of $\mathbb{R}^2$.

Lemma 10.5.2 $g = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \in \mathrm{Sym}_1^+(2,\mathbb{R})$ *if and only if* $\alpha > 0$ *and* $\det(g) = 1$.

Proof.

$\boxed{\Longrightarrow}$: we assume $g \in \mathrm{Sym}_1^+(2,\mathbb{R})$. Then, $\det(g) = 1$ by definition and, since g is positive-definite, $\alpha = \langle ge_1, e_1 \rangle > 0$, the inequality is strict because, if $\alpha = 0$, then $\det(g) = -\beta^2 \leqslant 0$, which contradicts the fact that $\det(g) = 1$.

$\boxed{\Longleftarrow}$: now we assume $g = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$ such that $\alpha > 0$ and $\det(g) = 1$. First of all we notice that $\det(g) = 1 \iff \alpha\gamma = 1 + \beta^2$, which also implies that $\gamma > 0$.
If we write $u = xe_1 + ye_2 = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, then:

$$\begin{aligned} \langle gu, u \rangle &= \langle g(xe_1 + ye_2), xe_1 + ye_2 \rangle = \alpha x^2 + \gamma y^2 + 2xy\beta \\ &= X^2 + Y^2 + 2XY \frac{\beta}{\sqrt{\alpha\gamma}} \quad \text{with} \quad X = x\sqrt{\alpha}\,,\ Y = y\sqrt{\gamma}. \end{aligned}$$

We remark that $\frac{\beta}{\sqrt{\alpha\gamma}} = \frac{\beta}{\sqrt{1+\beta^2}} \in (-1, 1)$ for all $\beta \in \mathbb{R}$. Therefore, if $\underline{XY \geqslant 0}$,

$$X^2 + Y^2 + 2XY \frac{\beta}{\sqrt{\alpha\gamma}} \geqslant X^2 + Y^2 - 2XY = (X - Y)^2 \geqslant 0$$

and in the other case, if $\underline{XY \leqslant 0}$,

$$X^2 + Y^2 + 2XY \frac{\beta}{\sqrt{\alpha\gamma}} \geqslant X^2 + Y^2 + 2XY = (X + Y)^2 \geqslant 0$$

Therefore, g is definite positive and $g \in \mathrm{Sym}_1^+(2, \mathbb{R})$. □

By writing the determinant of g explicitly we get $\alpha\gamma - \beta^2 = 1$, solving w.r.t. γ we obtain $\gamma = \frac{1+\beta^2}{\alpha}$, so that the generic parameterization of a matrix in $\mathrm{Sym}_1^+(2, \mathbb{R})$ is:

$$\mathrm{Sym}_1^+(2, \mathbb{R}) = \left\{ \begin{pmatrix} \alpha & \beta \\ \beta & \frac{1+\beta^2}{\alpha} \end{pmatrix},\ \alpha > 0,\ \beta \in \mathbb{R} \right\}.$$

Theorem 10.5.2 *The following assertions hold.*

1. *The function*

$$\begin{array}{rccl} F: & H & \xrightarrow{\sim} & \mathrm{Sym}_1^+(2, \mathbb{R}) \\ & z = x + iy & \longmapsto & \frac{1}{y} \begin{pmatrix} 1 & -x \\ -x & x^2 + y^2 \end{pmatrix}, \end{array}$$

   is bijective with inverse given by:

$$\begin{array}{rccl} \omega: & \mathrm{Sym}_1^+(2, \mathbb{R}) & \xrightarrow{\sim} & H \\ & \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} & \longmapsto & \frac{1}{\alpha}(-\beta + i). \end{array}$$

2. *The map*

$$\begin{array}{rccl} *: & SL(2, \mathbb{R}) \times \mathrm{Sym}_1^+(2, \mathbb{R}) & \longrightarrow & \mathrm{Sym}_1^+(2, \mathbb{R}) \\ & (m, g) & \longmapsto & m * g = (m^{-1})^t g m^{-1} \end{array}$$

   is a group action.

Proof.

1. Both maps F and ω are well-defined thanks to the previous lemma. First we check that $\omega(F(z)) = z$ for all $z = x + iy \in H$:

$$F(z) = \frac{1}{y}\begin{pmatrix} 1 & -x \\ -x & x^2 + y^2 \end{pmatrix} \quad \Longrightarrow \quad \omega(F(z)) = y\left(\frac{x}{y} + i\right) = x + iy = z.$$

Now we check that $F(\omega(g)) = g$ for all $g \in \mathrm{Sym}_1^+(2, \mathbb{R})$. To this aim, let $g = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \in \mathrm{Sym}_1^+(2, \mathbb{R})$, then:

$$\omega(g) = \frac{-\beta}{\alpha} + i\frac{1}{\alpha} \quad \Longrightarrow \quad F(\omega(g)) = \alpha\begin{pmatrix} 1 & \frac{\beta}{\alpha} \\ \frac{\beta}{\alpha} & \frac{1+\beta^2}{\alpha^2} \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \beta & \frac{1+\beta^2}{\alpha} \end{pmatrix} \underset{1=\det(g)=\alpha\gamma-\beta^2}{=} \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}.$$

2. The axioms of group actions follow from direct computations. Let $g \in \mathrm{Sym}_1^+(2, \mathbb{R})$ and $m, n \in SL(2, \mathbb{R})$.

- $Id * g = Id\, g\, Id = g$
- $(mn) * g = ((mn)^{-1})^t\ g\ (mn)^{-1} = (m^{-1})^t(n^{-1})^t g\, n^{-1}m^{-1} = m * (n * g)$.

All that is left to prove is then the stability of the action on $Sym_1^+(2, \mathbb{R})$, i.e. that $m * g$ has unitary determinant, is symmetric and positive-definite for all $m \in SL(2, \mathbb{R})$ and $g \in Sym_1^+(2, \mathbb{R})$:

- $\det(m * g) = \det((m^{-1})^t)\det(g)\det(m) = \det(m)\det(g)\det(m) = 1 \implies m * g \in SL(2, \mathbb{R})$
- $\left((m^{-1})^t g m^{-1}\right)^t = (m^{-1})^t g m^{-1}$, thus $m * g$ is symmetric (which explains why we must consider $(m^{-1})^t$) in the definition of the action $*$ and not m^{-1})
- for all $u \in \mathbb{R}^2$, $\langle m * gu, u\rangle = \langle (m^{-1})^t g\, m^{-1}u, u\rangle = \langle g(m^{-1}u), (m^{-1}u)\rangle \geqslant 0$ since $g \in Sym_1^+(2, \mathbb{R})$.
  Moreover, $m^{-1}u = 0 \iff u = 0$ because m is invertible. Therefore, $m * g$ is positive-definite. $\square$

To resume, we have determined the following isomorphisms:

$$\boxed{SL(2, \mathbb{R})/SO(2) \cong H \cong \mathrm{Sym}_1^+(2, \mathbb{R})}.$$

As we will see later, these are **three among the six prototypes of the hyperbolic plane** (also called hyperbolic models), the remaining three being the hyperboloid in $\mathbb{R}^3$, the Poincaré and the Klein disks.

10.5.3 The action of $\mathbf{SL}(2,\mathbb{R})$ on $\mathbf{Sym}_1^+(2,\mathbb{R})$

Finally, we show that the action of $SL(2,\mathbb{R})$ on $\mathrm{Sym}_1^+(2,\mathbb{R})$ is analogous to the action of $SL(2,\mathbb{R})$ on H by Möbius transformations. The proof of this result needs a lemma.

Lemma 10.5.3 *Let* $g = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \in \mathrm{Sym}_1^+(2,\mathbb{R})$. *Then,* $\omega(g) = \frac{1}{\alpha}(-\beta + i)$ *is the unique solution in* H *of the equation:*

$$\begin{pmatrix} z \\ 1 \end{pmatrix}^t g \begin{pmatrix} z \\ 1 \end{pmatrix} = \alpha z^2 + 2\beta z + \gamma = 0.$$

Proof. The two complex solutions of the equation are:

$$z_{1,2} = \frac{-2\beta \pm \sqrt{4\beta^2 - 4\alpha\gamma}}{2\alpha} = \frac{-\beta \pm \sqrt{\beta^2 - \alpha\gamma}}{\alpha},$$

but $1 = \det(g) = \alpha\gamma - \beta^2$, so $\beta^2 - \alpha\gamma = -1$, so

$$z_{1,2} = \frac{-\beta \pm i}{\alpha},$$

the only solution in H is the one corresponding to $+i$, i.e. the only solution in H is:

$$\frac{1}{\alpha}(-\beta + i) = \omega(g).$$

$\square$

Theorem 10.5.3 *Let* $m \in SL(2,\mathbb{R})$, $g \in \mathrm{Sym}_1^+(2,\mathbb{R})$ *and* $z \in H$. *Then,*

1. $\omega(m * g) = m \cdot \omega(g)$
2. $F(m \cdot z) = m * F(z)$,

i.e. the following diagram

$$\begin{array}{ccc} \mathrm{Sym}_1^+(2,\mathbb{R}) & \underset{F}{\overset{\omega}{\rightleftarrows}} & H \\ \downarrow{\scriptstyle m*} & & \downarrow{\scriptstyle m\cdot} \\ \mathrm{Sym}_1^+(2,\mathbb{R}) & \underset{F}{\overset{\omega}{\rightleftarrows}} & H \end{array}$$

is commutative. Hence, the action of $SL(2,\mathbb{R})$ *on* $\mathrm{Sym}_1^+(2,\mathbb{R})$ *is transitive.*

Proof.

1. We start by fixing the notation:

- $m = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{R})$, so that $m^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$
- $\tilde{z} = \omega(m * g) \in H$

- $z = \omega(g) \in H$
- $\tilde{\omega} = m^{-1} \cdot \tilde{z} = \mathscr{M}(m^{-1}, \tilde{z}) \in H$.

We note that:

$$m^{-1}\begin{pmatrix}\tilde{z}\\1\end{pmatrix} = \begin{pmatrix}d\tilde{z}-b\\-c\tilde{z}+a\end{pmatrix} = \underbrace{(-c\tilde{z}+a)}_{\neq 0}\begin{pmatrix}\frac{d\tilde{z}-b}{-c\tilde{z}+a}\\1\end{pmatrix} = (-c\tilde{z}+a)\begin{pmatrix}m^{-1}\cdot\tilde{z}\\1\end{pmatrix} = (c\tilde{z}+a)\begin{pmatrix}\tilde{\omega}\\1\end{pmatrix}, \tag{10.12}$$

where $-c\tilde{z}+a \neq 0$ because, if we write $\tilde{z} \in H$ as $\tilde{z} = x+iy$, then $-c\tilde{z}+a = 0$ would be equivalent to $-cx+a-icy = 0$, but, since $y > 0$, this would imply $a = c = 0$, that cannot be because m^{-1} would have a null row and would not be invertible.

Thanks to Lemma 10.5.3, $\tilde{z} = \omega(m * g) = \omega((m^{-1})^t g m^{-1})$ verifies:

$$\begin{pmatrix}\tilde{z}\\1\end{pmatrix}^t (m^{-1})^t g\, m^{-1}\begin{pmatrix}\tilde{z}\\1\end{pmatrix} = 0, \tag{10.13}$$

but

$$\begin{pmatrix}\tilde{z}\\1\end{pmatrix}^t (m^{-1})^t = \left(m^{-1}\begin{pmatrix}\tilde{z}\\1\end{pmatrix}\right)^t \underset{(10.12)}{=} (-c\tilde{z}+a)\begin{pmatrix}\tilde{\omega}\\1\end{pmatrix}^t,$$

so:

$$\begin{aligned}\begin{pmatrix}\tilde{z}\\1\end{pmatrix}^t (m^{-1})^t g\, m^{-1}\begin{pmatrix}\tilde{z}\\1\end{pmatrix} = 0 \quad &\iff \quad (-c\tilde{z}+a)^2\begin{pmatrix}\tilde{\omega}\\1\end{pmatrix}^t g\begin{pmatrix}\tilde{\omega}\\1\end{pmatrix} = 0\\ &\iff \quad \begin{pmatrix}\tilde{\omega}\\1\end{pmatrix}^t g\begin{pmatrix}\tilde{\omega}\\1\end{pmatrix} = 0.\end{aligned}$$

Therefore, thanks to Lemma 10.5.3, $\tilde{\omega} = \omega(g)$ and so $\omega(m * g) = m \cdot \omega(g)$.

2. Having proven 1., i.e. $m \cdot \omega(g) = \omega(m * g)$, the proof of 2. is very easy. In fact, thanks to theorem 10.5.2, $z = \omega(g)$ so $F(z) = F(\omega(g)) = g$, thus:

$$F(m \cdot z) = F(m \cdot \omega(g)) = F(\omega(m * g)) = m * g = m * F(z).$$

$\square$

Chapter 11

Geometry of the Lorentz space and Lorentz transformations

Minkowski discovered that the fundamental facts of Einstein's special relativity could be expressed via the pseudo-Euclidean geometry of the Lorentz space. Klein praised this fact as a triumph of his Erlangen program because Minkowski's discovery showed that the kinematics of the special theory of relativity coincides with the geometry of the Lorentz space.

This geometry and the related Lorentz transformations are two of the three basic tools that will allow us to rigorously describe the different realizations (called models) of hyperbolic geometry, the third tool being represented by Möbius transformations, that will be analyzed in the following chapter.

11.1 A quick recap about the Euclidean scalar product

In this section, we recall very quickly just the basic facts of Euclidean geometry that will use in the rest of this chapter.

Def. 11.1.1 *The Euclidean scalar product on* $\mathbb{R}^n$ *is defined as:*

$$\langle x, y\rangle = x_1 y_1 + \cdots + x_n y_n,$$

its associated norm[1] *is:*

$$|x| = \|x\|_E = \sqrt{\langle x, x\rangle} = (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}},$$

and its associated metric is:

$$d_E(x, y) = |x - y| = \sqrt{\langle x - y, x - y\rangle}.$$

Lemma 11.1.1 (Cauchy-Schwarz inequality) *Let* $x, y \in \mathbb{R}^n$*. Then,*

$$|\langle x, y\rangle| \leqslant |x|\,|y|$$

and the equality holds if and only if x *and* y *are linearly dependent.*

[1] In this chapter, exceptionally, it is notationally more convenient to use the symbol $|\ |$ for the Euclidean norm and reserve $\|\ \|$ to the Lorentzian norm.

Proof. Suppose x and y are linearly dependent. Then, it exists $t \neq 0$ such that $y = tx$, so

$$|\langle x, y\rangle| = |\langle x, tx\rangle| = |t\langle x, x\rangle| = |t||x|^2 = |x||t||x| = |x||y|.$$

Conversely, assume x and y are linearly independent. Then for all $t \in \mathbb{R}$, $tx - y \neq 0$ and so

$$0 < |tx - y|^2 = |x|^2 t^2 - 2\langle x, y\rangle t + |y|^2 = f(t),$$

which implies that the discriminant of f is negative, i.e. $\Delta = 4\langle x, y\rangle^2 - 4|x|^2|y|^2 < 0$, hence

$$\langle x, y\rangle^2 < |x|^2|y|^2 \iff |\langle x, y\rangle| < |x||y|.$$

□

We will denote by $E^n = (\mathbb{R}^n, d_E)$ the Euclidean metric n-space considering it as an affine space so that we can perform translations in E^n.

Def. 11.1.2 (Isometries and similarities in E^n) *The isometries of E^n are transformations that preserve distances:*

$$\mathcal{I}(E^n) = \{\phi : \mathbb{R}^n \to \mathbb{R}^n \ : \ d_E(\phi(x), \phi(y)) = d_E(x, y) \ \ \forall x, y \in \mathbb{R}^n\}.$$

The similarities of E^n are transformations that preserve shapes:

$$\mathcal{S}(E^n) = \{\phi : \mathbb{R}^n \to \mathbb{R}^n \ : \ \exists k > 0 \ : \ d_E(\phi(x), \phi(y)) = k d_E(x, y) \ \ \forall x, y \in \mathbb{R}^n\}.$$

The sets of isometries and similarities form a group under composition.

The most important set of transformations in Euclidean geometry are the orthogonal ones, which form the group $\mathrm{O}(n)$, defined as maps that preserve the scalar product:

$$\langle \phi(x), \phi(y)\rangle = \langle x, y\rangle \qquad \forall x, y \in \mathbb{R}^n.$$

The following lemma allows us to characterize the orthogonal transformations.

Lemma 11.1.2 *$\phi : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation if and only if it is linear and, given an orthonormal basis $(u_1, \ldots, u_n)$ of $\mathbb{R}^n$, $(\phi(u_1), \ldots, \phi(u_n))$ is an orthonormal basis of $\mathbb{R}^n$.*

Proof. Suppose ϕ is an orthogonal transformation and $(u_1, \ldots, u_n)$ is any orthonormal basis of $\mathbb{R}^n$. Then,

$$\langle \phi(u_i), \phi(u_j)\rangle = \langle u_i, u_j\rangle = \delta_{ij},$$

hence $(\phi(u_1), \cdots, \phi(u_n))$ is, by definition, an orthonormal basis of $\mathbb{R}^n$ and so, for all $x \in \mathbb{R}^n$,

$$\phi(x) = \sum_{i=1}^{n} \langle \phi(x), \phi(u_i)\rangle \phi(u_i) \underset{(\phi \text{ orthogonal})}{=} \sum_{i=1}^{n} \langle x, u_i\rangle \phi(u_i),$$

but we also have $x = \sum\limits_{i=1}^{n} \langle x, u_i\rangle u_i$, so, by writing $x_i = \langle x, u_i\rangle$, we get:

$$\phi(\sum_{i=1}^{n} x_i u_i) = \sum_{i=1}^{n} x_i \phi(u_i), \qquad \forall x \in \mathbb{R}^n. \tag{11.1}$$

Now, if we consider $\lambda x = \sum_{i=1}^{n} \lambda x_i u_i$, we have:

$$\phi(\lambda x) = \sum_{i=1}^{n} \lambda x_i \phi(u_i) = \lambda \sum_{i=1}^{n} x_i \phi(u_i) = \lambda \phi(x), \qquad \forall x \in \mathbb{R}^n.$$

Moreover, if we consider another vector $y = \sum_{i=1}^{n} y_i u_i \in \mathbb{R}^n$, thanks to (11.1) we get

$$\phi(\sum_{i=1}^{n} y_i u_i) = \sum_{i=1}^{n} y_i \phi(u_i),$$

thus

$$\begin{aligned}\phi(x+y) &= \phi(\sum_{i=1}^{n} x_i u_i + \sum_{j=1}^{n} y_j u_j) = \phi(\sum_{k=1}^{n}(x_k + y_k)u_k) \underset{(11.1)}{=} \sum_{k=1}^{n}(x_k + y_k)\phi(u_k) \\ &= \sum_{i=1}^{n} x_i \phi(u_i) + \sum_{i=1}^{n} y_i \phi(u_i) \\ &= \phi(x) + \phi(y),\end{aligned}$$

hence the linearity of ϕ.

Conversely, suppose that ϕ is linear and that, for any orthonormal basis $(u_1, \dots, u_n)$ of $\mathbb{R}^n$, $(\phi(u_1), \dots, \phi(u_n))$ is again an orthonormal basis of $\mathbb{R}^n$. Then,

$$\phi(x) = \phi(\sum_{i=1}^{n} x_i u_i) \underset{(\phi \text{ linear})}{=} \sum_{i=1}^{n} x_i \phi(u_i) \qquad \forall x \in \mathbb{R}^n,$$

thus

$$\begin{aligned}\langle \phi(x), \phi(y) \rangle &= \left\langle \sum_{i=1}^{n} x_i \phi(u_i), \sum_{j=1}^{n} y_j \phi(u_j) \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \langle \phi(u_i), \phi(u_j) \rangle \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \delta_{i,j} = \sum_{i=1}^{n} x_i y_i \\ &= \langle x, y \rangle .\end{aligned}$$

□

Def. 11.1.3 *The function*

$$\begin{aligned} q: \ \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto q(x) := \langle x, x \rangle = \|x\|_E^2 \end{aligned}$$

is called the ***quadratic form*** *associated to the Euclidean scalar product.*

The following result shows how an orthogonal transformation $\phi : \mathbb{R}^n \to \mathbb{R}^n$ can be completely characterized via the quadratic form q.

Lemma 11.1.3 *$\phi : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation if and only if it preserves the quadratic form q, i.e. $q(\phi(x)) = q(x)$ for all $x \in \mathbb{R}^n$.*

Proof. By direct computation we get

$$q(x-y) = q(x) - 2\langle x, y\rangle + q(y) \iff \langle x, y\rangle = \frac{q(x)+q(y)-q(x-y)}{2}, \qquad \forall x, y \in \mathbb{R}^n,$$

so, it also holds that

$$\langle \phi(x), \phi(y)\rangle = \frac{q(\phi(x)) + q(\phi(y)) - q(\phi(x)-\phi(y))}{2}, \qquad \forall x, y \in \mathbb{R}^n.$$

So, if $q(\phi(x)) = q(x)$ for all $x \in \mathbb{R}^n$, then:

$$\langle \phi(x), \phi(y)\rangle = \frac{q(\phi(x)) + q(\phi(y)) - q(\phi(x)-\phi(y))}{2} = \frac{q(x)+q(y)-q(x-y)}{2} = \langle x, y\rangle,$$

i.e. ϕ is orthogonal.

Vice-versa, if ϕ is orthogonal, then $q(\phi(x))+q(\phi(y))-q(\phi(x)-\phi(y)) = q(x)+q(y)-q(x-y)$ for all $x, y \in \mathbb{R}^n$, but this equality holds true no matter how x and y are chose only when $q(\phi(x)) = q(x)$ for all $x \in \mathbb{R}^n$. □

Since $q(\phi(x)) = \|\psi(x)\|^2$, the condition $q(\phi(x)) = q(x)$ means $\|\psi(x)\|^2 = \|x\|^2$ or, since the Euclidean norm is non-negative, $\|\psi(x)\|^2 = \|x\|$, hence Lemma 11.1.3 says that requiring the preservation of the Euclidean scalar product is equivalent to requiring the preservation of the Euclidean norm.

Finally, we come to the complete characterization of $\mathcal{I}(E^n)$ and $\mathcal{S}(E^n)$.

Theorem 11.1.1 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$.*

1. *$f \in \mathcal{I}(E^n)$ if and only if f is of the form $f(x) = a + \phi(x)$, with $a \in \mathbb{R}^n$ and $\phi \in \mathrm{O}(n)$.*
2. *$f \in \mathcal{S}(E^n)$ if and only if f is of the form $f(x) = a + k\phi(x)$, with $a \in \mathbb{R}^n$, $k > 0$ and $\phi \in \mathrm{O}(n)$.*

Proof. First of all, notice that 1. is simply a special case of 2. with $k = 1$, thus we will concentrate only on the proof of 2.

$\boxed{2. \implies}$: if $f(x) = a + \phi(x)$, then, for all $x, y \in \mathbb{R}^n$,

$$\begin{aligned} d(f(x), f(y)) &= \langle a + k\phi(x) - (a + k\phi(y)), a + k\phi(x) - (a + k\phi(y))\rangle^{\frac{1}{2}} \\ &= \langle k(\phi(x) - \phi(y)), k(\phi(x) - \phi(y))\rangle^{\frac{1}{2}} \\ &\underset{(\phi \text{ linear})}{=} k\,\langle(\phi(x-y)), \phi(x-y)\rangle^{\frac{1}{2}} \\ &\underset{(\phi \text{ orthogonal})}{=} k\,\langle x-y, x-y\rangle^{\frac{1}{2}} \\ &= kd(x, y). \end{aligned}$$

$\boxed{2. \Longleftarrow}$: suppose $f \in \mathcal{S}(E^n)$ and let $a = f(0)$ and $\psi(x) = f(x) - a$. Since f is a similarity there is a $k > 0$ such that $|f(x) - f(y)| = k|x - y|$ for all $x, y \in \mathbb{R}^n$ and so

$$|\psi(x)| = |f(x) - f(0)| = k|x - 0| = k|x|.$$

Consequently, by setting $\phi = \frac{1}{k}\psi$ and using lemma 11.1.3, we have $\phi \in O(n)$ and

$$f(x) = a + k\phi(x).$$

□

Corollary 11.1.1 *An affine function $f : \mathbb{R}^n \to \mathbb{R}^n$, $f(x) = a + \lambda x$, with $a, x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R} \backslash \{0\}$ is always a Euclidean similarity and it is a Euclidean isometry if and only if $\lambda = 1$.*

Proof. The proof consists simply in remarking that λ can be identified with a one-entry matrix, which is orthogonal if and only if $\lambda = 1/\lambda$, i.e. $\lambda = 1$. □

11.2 The geometry of the Lorentz n-space

The main reference throughout this section is Ratcliffe's book [18].

Lorentzian geometry is founded on an alternative definition of the scalar product in $\mathbb{R}^n$ w.r.t. the Euclidean one for $n \geqslant 2$, when $n = 1$ the two products agree. For this reason, in this chapter we will always implicitly consider $n \geqslant 2$.

11.2.1 Rudiments about the pseudo-scalar product

The Lorentz scalar product is actually a so-called pseudo-scalar product. The formal algebraic theory that allowed the modern definition of such a concept has been developed by E. Witt in [26]. Here we collect only the definitions and results that are needed to understand Lorentz's geometry, for a more thorough discussion see, e.g., [15].

Let V be a real vector space and x, y arbitrary vectors in V.

1. A bilinear form on V is an $\mathbb{R}$-bilinear function $b : V \times V \to \mathbb{R}$.

2. The quadratic form associated to b is the linear functional $q_b : V \to \mathbb{R}$ defined by $q_b(x) := b(x, x)$. It is often simpler to work with q_b than with b and no information is lost, since we can reconstruct b from q via the well-known polarization identity:

$$b(x, y) = \frac{1}{2}(q_b(x + y) - q_b(x) - q_b(y));$$

3. b is symmetric if $b(x, y) = b(y, x)$ for all x, y. In what follows, b will always be implicitly considered symmetric.

4. b is **definite** if $x \neq 0$ implies $q_b(x) \neq 0$. b is **positive (negative) definite** if $x \neq 0$ implies $q_b(x) > 0$ (< 0).

5. b is **positive (negative) semi-definite** if $x \neq 0$ implies $q_b(x) \geqslant 0$ ($\leqslant 0$).

6. Of course, if b is positive (negative) definite, then it is also positive (negative) semi-definite.

7. If b is neither positive nor negative semi-definite, b is called **indefinite**.

8. b is **nondegenerate** if $b(x, y) = 0\ \forall y$ implies $x = 0$.

9. A **scalar product** g on V is a positive-definite nondegenerate symmetric bilinear form on V. (V, g) is called a scalar product space.

10. A **pseudo-scalar product** b on V is a nondegenerate symmetric bilinear form on V. (V, b) is called a pseudo-scalar product space. Thus, the big difference between a pseudo- and a scalar product is the *lack of definite-positiveness* for the first.

11. For any vector subspace $W \subset V$, we denote with $b|_W$ and $q_b|_W$ the restriction of b to $W \times W$ and of q_b to W, respectively. If b is a symmetric bilinear form, so is $b|_W$.

12. The **index** ν of a symmetric bilinear form b on V is the largest integer that coincides with the dimension of a subspace $W \subset V$ on which b_W is **negative definite**. Thus $0 \leqslant \nu \leqslant \dim(V)$, with $\nu = 0$ if and only if b is positive-semidefinite or positive-definite and $\nu = \dim(V)$ if and only if b is negative definite.

13. If $(u_1, \ldots, u_n)$ is a basis for V, the $n \times n$ matrix $B = (b_{ij}) = b(u_i, u_j)$ is called the **matrix of** b **relative to** $(u_1, \ldots, u_n)$. If b is symmetric, then B is a symmetric matrix. If $x = \sum_{i=1}^{n} x_i u_i$ and $y = \sum_{j=1}^{n} y_j u_j$, then, by the bilinearity and symmetry of b we have

$$\begin{aligned} b(x, y) &= b\left(\sum_{i=1}^{n} x_i u_i, \sum_{j=1}^{n} y_j u_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j b(u_i, u_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j b_{ij} \\ &= \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} b_{ij} y_j\right) = \langle x, By\rangle \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} b_{ij} x_i\right) y_j = \langle Bx, y\rangle, \end{aligned}$$

thus the action of b on the vectors of V is completely determined by B.

14. A symmetric bilinear form b on V is nondegenerate if and only if its matrix B relative to an arbitrary basis of V is invertible.

15. A vector $u \in (V, b)$ such that $q_b(u) = \pm 1$ is said to be a unit vector in (V, b). Two vectors $x, y \in (V, b)$ are orthogonal if $b(x, y) = 0$.

16. A set of $m \leqslant n$ mutually orthogonal unit vectors in (V, b) is said to be an orthonormal family. If $m = n$, then we talk about an orthonormal basis of (V, b).

17. The matrix of b associated to an orthonormal basis $(u_1, \ldots, u_n)$ is diagonal, in fact its entries are given by:
$$b(u_i, u_j) = \pm\delta_{i,j},$$
the ordered sequence of -1 and $+1$, repeated all the time they appear in the diagonal of the matrix associated to b w.r.t. any orthonormal basis, is called **signature** of b.

18. The signature appears in the orthonormal expansion of any $x \in (V, b)$ on an orthonormal basis $(u_1, \ldots, u_n)$ as follows:
$$x = \sum_{i=1}^{n} \varepsilon_i b(x, u_i) u_i \ , \tag{11.2}$$
where $\varepsilon_i = q_b(u_i) \in \{-1, +1\}$.

19. The orthogonal projection π of any $x \in (V, b)$ onto a subspace $W = \text{span}(u_1, \ldots, u_m)$, where $(u_1, \ldots, u_m)$ is an orthonormal family and $m < n$ is the following:
$$\pi(x) = \sum_{i=1}^{m} \varepsilon_i b(x, u_i) u_i \ .$$

20. For all $x \notin W$, the **residual vector**
$$y := x - \pi(x) = x - \sum_{i=1}^{m} \varepsilon_i b(x, u_i) u_i \tag{11.3}$$
is orthogonal to all vectors of W: $b(y, w) = 0 \ \forall w \in W$.

21. The number of negative signs in the signature of b is constant for any orthonormal basis $(u_1, \ldots, u_m)$ of (V, b) and it coincides with the index ν. For the proof see [**?**], lemma 26 page 51.

11.2.2 Lorentz's and Minkowski's pseudo-scalar product

Let us specify all this in the case of the Lorentz pseudo-scalar product in $\mathbb{R}^n$, which is the most important example of symmetric bilinear form of index -1.

Def. 11.2.1 (Lorentz's pseudo-scalar product) *Let $x, y \in \mathbb{R}^n$. The Lorentz (or Lorentzian) pseudo-scalar product between x and y is defined as follows:*
$$\boxed{x \circ y = -x_1 y_1 + x_2 y_2 + \cdots + x_n y_n} \ .$$
$(\mathbb{R}^n, \circ)$, *i.e.* $\mathbb{R}^n$ *interpreted as a vector space endowed with the Lorentzian pseudo-scalar product, is denoted by* $\mathbb{R}^{1,n-1}$ *and called* ***the Lorentzian*** n***-space****.*

In literature, we find several other definitions of the Lorentzian scalar product. The first alternative definition that we discuss is the following:
$$x \circ y = x_1 y_1 + \cdots + x_{n-1} y_{n-1} - x_n y_n$$
and $\mathbb{R}^n$ endowed with this last Lorentzian scalar product is denoted by $\mathbb{R}^{n-1,1}$. Results in both cases are exactly the same and the choice of $\mathbb{R}^{n-1,1}$ or $\mathbb{R}^{1,n-1}$ depends on convenience or taste.

Instead, the following alternative and perfectly valid choice:
$$x \circ y = x_1 y_1 - \cdots - x_{n-1} y_{n-1} - x_n y_n$$
corresponds to the definition given above but with an opposite sign.

The case of $n = 4$ is of particular importance in Physics, as it is the geometric setting of special relativity, with the coordinate $x_1 = t$ playing the role of time and $(x_2, x_3, x_4) = (x, y, z)$ the role of space coordinates. For this reason, it bears a special name.

Def. 11.2.2 *The Lorentz space* $\mathbb{R}^{1,3}$ *is called* ***Minkowski spacetime*** $\mathcal{M}$.

The bilinearity and symmetry of the Lorentz pseudo-scalar product is immediate to see, to prove its nondegeneracy it is enough to take as y all the vectors of the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$:

$$x \circ e_1 = -x_1 \cdot 1 + x_2 \cdot 0 + \cdots + x_n \cdot 0 = -x_1$$

$$\vdots$$

$$x \circ e_n = -x_1 \cdot 0 + x_2 \cdot 0 + \cdots + x_n \cdot 1 = x_n$$

so, $x \circ e_i = 0$ for all $i = 1, \ldots, n$ implies $-x_1 = x_2 = \cdots = x_n = 0$, i.e. $x = 0$. By linearity, this result can be extended to all $x \in \mathbb{R}^{1,n}$.

Def. 11.2.3 *The quadratic form associated to the Lorentz pseudo-scalar product is:*

$$q(x) := x \circ x = \|x\|^2 = -x_1^2 + x_2^2 + \cdots + x_n^2, \quad q(x) \in \mathbb{R},$$

where $\| \ \|$ *is the Lorentz pseudo-norm given by:*

$$\|x\| := \sqrt{q(x)} = \sqrt{-x_1^2 + x_2^2 + \cdots + x_n^2}, \quad \|x\| \in \mathbb{R}^+ \cup \{0\} \cup i\mathbb{R}^+.$$

In the Minkowski space, the Lorentz pseudo-norm is called **Minkowski pseudo-norm**.

Remark 11.2.1 The Lorentz pseudo-scalar product can be rewritten using the Euclidean scalar product:

$$x \circ y = \langle x, \eta y \rangle = \langle \eta x, y \rangle, \ \text{ with } \ \eta = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix} = \mathrm{diag}(-1, 1, \ldots).$$

Crucial properties of η, that will be used many times in future computations, are the following:

$$\eta^t = \eta^{-1} = \eta, \ \eta^2 = I_n. \tag{11.4}$$

11.2.3 Likeness and orientation

It is clear that *the Lorentz pseudo-scalar product is indefinite*, thus there are non-null vectors in $\mathbb{R}^{1,n-1}$ with zero Lorentz pseudo norm, these vectors turn out to be extremely important, in fact we will see in chapter 14, dedicated to the special theory of relativity, that they define the trajectories of massless particles traveling at the speed of light.

This fact allows us to separate the Lorentz n-space in three different subsets called time-like, light-like and space-like, a terminology taken from special relativity.

Def. 11.2.4 $x \in \mathbb{R}^{1,n-1}$ *is said to be*

- ***time-like*** *if* $q(x) = x \circ x = \|x\|^2 < 0$;
- ***light-like*** *if* $q(x) = x \circ x = \|x\|^2 = 0$;
- ***space-like*** *if* $q(x) = x \circ x = \|x\|^2 > 0$;
- ***causal*** *if it is not space-like, i.e. if it is either time-like or light-like.*

One of the three options is called ***likeness*** *of x.*
Moreover, the ***orientation*** *(or* ***parity****) of a time-like or light-like vector is* ***positive*** *if $x_1 > 0$ and* ***negative*** *if $x_1 < 0$.*

Examples: if we consider the vectors $(e_1, \ldots, e_n)$ of the canonical basis of $\mathbb{R}^n$, then, by direct computation, $e_1 \circ e_1 = -1$, $e_j \circ e_j = 1$ and $\|e_1 + e_j\|^2 = 0$ for all $2 \leqslant j \leqslant n$, so:

- e_1 is a time-like vector;
- e_j, for all $2 \leqslant j \leqslant n$, are space-like vectors;
- $e_1 + e_j$ is a light-like vector for all $2 \leqslant j \leqslant n$, e.g. $e_1 + e_2 = (1, 1, 0, \ldots, 0)^t$.

It is instructive to examine the geometric meaning of likeness in the case $n = 2$ for a generic $u \in \mathbb{R}^{1,1}$. Figure 11.1 gives a graphical representation of this simple analysis.

- $u = (t, x)$ is light-like if and only if $x^2 = x^2 \iff |t| = |x|$, i.e. light-like vectors lie on the two $\pi/4$ degrees straight lines passing through the origin.
- The time-like case is characterized by $x^2 < t^2 \iff |x| < |t|$, i.e. time-like vectors belong to the interior[2] of the upper and lower triangular regions in Figure 11.1.
- The space-like case is obviously identified by the remaining areas.

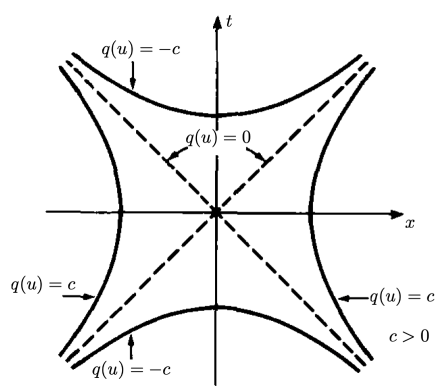


Figure 11.1: A graphical depiction of likeness regions in 2 dimensions, together with the level lines of the Lorentz pseudo-norm.

In the figure we can also see that the level lines of the quadratic form $q(x)$, i.e. the **vectors with same Lorentz pseudo-norm, lie on hyperbolas** contained in either the time-like or space-like regions with asymptotes given by the light-like straight lines. In fact, for all $c \neq 0$, $q(x) = c \iff -t^2 + x^2 = c$, which is the equation of a hyperbola. The hyperbola belongs to the time-like or space-like region, respectively, when $c < 0$ or $c > 0$.

More generally, the light-like equation $x \circ x = q(x) = \|x\|^2 = 0$ defines a hypercone $\mathcal{C}^{n-1}$ in $\mathbb{R}^n$, called **lightcone**. Time-like vectors belong to its interior, while space-like vectors belong to the external region. Figure 11.2 depicts the case $n = 3$.

[2] To see this, consider for simplicity the vector $u = (t, x)$ with $t, u > 0$, then u is time-like when $x < t$. If the t and x axes were switched w.r.t. Figure 11.1 , then of course the time-like regions would be external to the dotted triangular regions.

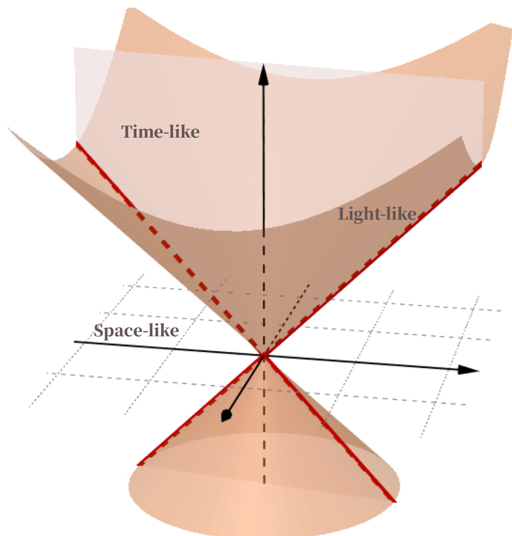


Figure 11.2: The Lorentz n-space is separated in three components: light-like vectors belong to the surface of the lightcone, time-like vectors lie inside and space-like vectors lie outside.

The name lightcone is an extension of the case $n = 4$, where $\mathcal{C}^{n-1}$ is the cone traveled by rays of light in the Minkowski spacetime of special relativity. The positive and negative time-like regions are called, respectively, **future** and **past lightcone**.

In E^n, vectors with same Euclidean norm lie on spheres, while, in the n-Lorentz space, **vectors with same Lorentz pseudo-norm lie on hyperboloids** contained in either the time-like or space-like regions. This remarkable difference should not be surprising, since, if the iso-norm surfaces would be the same, then there would be no difference between the hyperbolic and the Euclidean geometry.

These very simple considerations give also the first hint of why the Lorentzian pseudo-scalar product is related to hyperbolic geometry.

The following *notational convention* will simplify a lot future equations: whenever useful, we will write $x \in \mathbb{R}^n$ as

$$x = (x_1, x_2, \dots, x_n)^t = (x_1, \bar{x})^t, \quad \text{i.e.} \quad \bar{x} = (x_2, x_3, \dots, x_n)^t.$$

With this notation:

- the Lorentz scalar product and norm can be written as follows:

$$x \circ y = \langle \bar{x}, \bar{y} \rangle - x_1 y_1, \quad \|x\|^2 = |\bar{x}|^2 - x_1^2 \ ;$$

- $\mathbb{R}^{1,n-1} \ni x = \begin{cases} \text{light-like:} & |\bar{x}| = |x_1| \quad (x \in \mathcal{C}^{n-1}) \\ \text{time-like:} & |\bar{x}| < |x_1| \quad (x \in \text{int}(\mathcal{C}^{n-1})) \\ \text{space-like:} & |\bar{x}| > |x_1| \quad (x \in \text{ext}(\mathcal{C}^{n-1})) \end{cases}$.

11.2.4 Lorentz-orthogonality

Just like in Euclidean geometry, orthogonality plays a major role in the geometry of the Lorentz space.

Def. 11.2.5 *$x, y \in \mathbb{R}^{1,n-1}$ are called* ***Lorentz-orthogonal*** *if $x \circ y = 0$. They are* ***Lorentz-orthonormal*** *if they are Lorentz-orthogonal and if the modulus of their Lorentz pseudo-norm is 1, i.e. $\|x\| = \pm i$ ($\|x\|^2 = -1$) if x is time-like, and $\|x\| = 1$ ($\|x\|^2 = 1$) if x is space-like.*

The Euclidean scalar product allows us to fully understand the Euclidean geometry, hence it is not surprising that many information about the Lorentz n-space geometry can be gathered by studying the Lorentz scalar product.

We start by first proving a very simple fact and then a result that will have important consequences.

Lemma 11.2.1 *For all $x \in \mathbb{R}^{1,n-1}$ and all $t > 0$, the vector tx has the same likeness and orientation as x.*

Proof. For all $t > 0$, $\|tx\| = t\,\|x\|$ and $(tx)_1 = tx_1$, thus tx and x have the same likeness and orientation. □

Theorem 11.2.1 *If $x, y \in \mathbb{R}^{1,n-1}$ are non-zero, equioriented and causal, then $x \circ y \leqslant 0$ and the equality holds if and only if x and y are linearly dependent light-like vectors.*

Proof. The case of x and y being both negatively oriented can be derived from the positive case by replacing x and y with $-x$ and $-y$, respectively, thus we can assume that both x and y are positively oriented, i.e. $x_1 > 0$ and $y_1 > 0$.

By hypothesis, x is time-like or light-like, so:

$$x \circ x \leqslant 0 \iff |\bar{x}|^2 \leqslant x_1^2 \underset{x_1>0}{\iff} |\bar{x}| \leqslant x_1,$$

and, similarly, $|\bar{y}| \leqslant y_1$. These inequalities, together with the Cauchy-Schwartz inequality 11.1.1 applied on the Euclidean scalar product $|\ |$, imply

$$\langle \bar{x}, \bar{y} \rangle \leqslant |\langle \bar{x}, \bar{y} \rangle| \leqslant |\bar{x}||\bar{y}| \leqslant x_1 y_1, \tag{11.5}$$

hence we come to the conclusion that $x \circ y = \langle \bar{x}, \bar{y} \rangle - x_1 y_1 \leqslant 0$.

Let us now examine when $x \circ y = \langle \bar{x}, \bar{y} \rangle - x_1 y_1 = 0$, i.e. $x_1 y_1 = \langle \bar{x}, \bar{y} \rangle$, thus we can replace $x_1 y_1$ on the rightmost part of inequality (11.5) with $\langle \bar{x}, \bar{y} \rangle$, this implies $\langle \bar{x}, \bar{y} \rangle \leqslant |\bar{x}||\bar{y}| \leqslant \langle \bar{x}, \bar{y} \rangle$, i.e. $\langle \bar{x}, \bar{y} \rangle = |\bar{x}||\bar{y}|$. Lemma 11.1.1 guarantees that this can happen if and only if $\bar{x}$ and $\bar{y}$ are linearly dependent, we can therefore set $\bar{y} = t\bar{x}$, with $t \neq 0$, and observe that

$$x \circ y = 0 \iff x_1 y_1 = t|\bar{x}|^2 \iff y_1 = \frac{t|\bar{x}|^2}{x_1}, \tag{11.6}$$

which implies two things: firstly $t > 0$ (x_1 and y_1 are both supposed to be positive), secondly, $y \circ y = |\bar{y}|^2 - y_1^2 \underset{\bar{y}=t\bar{x}}{=} t^2|\bar{x}|^2 - \frac{t^2|\bar{x}|^4}{x_1^2}$. Recalling that y is either time-like or light-like, we must have:

$$y \circ y \leqslant 0 \iff x_1^2 \leqslant |\bar{x}|^2 \iff 0 \leqslant |\bar{x}|^2 - x_1^2 \iff 0 \leqslant x \circ x.$$

If x is time-like, then $x \circ x < 0$ and the previous inequality is not verified, thus x must be light-like (i.e. $x \circ x = 0$). Being light-like and positively oriented, x satisfies $|\bar{x}| = x_1$ and so the central equation of formula (11.6) implies $x_1 y_1 = t x_1^2$, i.e. $y_1 = t x_1$. In conclusion, $\bar{y} = t\bar{x}$ and $y_1 = t x_1$ imply that $y = tx$ and also that x and y are both light-like vectors. □

Corollary 11.2.1 *If $x, y \in \mathbb{R}^{1,n-1}$ are equioriented time-like vectors, then $x \circ y < 0$, so, in particular,* ***two equioriented time-like vectors are never Lorentz-orthogonal****.*

The following corollary reveals a crucial constraint for a Lorentz-orthogonal family of vectors in $\mathbb{R}^{1,n-1}$.

Corollary 11.2.2 *Let x, y be two non-zero Lorentz-orthogonal vectors, i.e. $x, y \neq 0$ and $x \circ y = 0$, then:*

$$\boxed{x \text{ time-like} \implies y \text{ space-like}}.$$

Proof. We can assume that y has the same orientation of x, in the opposite case it is sufficient to replace y with $-y$ to obtain the proof. If x is time-like then, by theorem 11.2.1, if y is time-like or light-like, we must have $x \circ y < 0$, where the inequality is strict because the equality can happen only with two linearly dependent light-like vectors. Since $x \circ y < 0$ is incompatible with the hypothesis $x \circ y = 0$, the only option that remains is that y must be space-like. □

Notice that *the reverse statement of the previous result is not true*: *given two non-zero Lorentz-orthogonal vectors $x, y \in \mathbb{R}^{1,n-1}$, if x is space-like then y can be both space-like and time-like.* A simple example is given by the vectors $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ of the canonical basis of $\mathbb{R}^{1,2}$: $\|e_2\|^2 = \|e_3\|^2 = 1$, so they are space-like, but, by direct computation we see that $e_2 \circ e_3 = 0$.

A significant implication of the results collected above is that **two Lorentz-orthogonal vectors of $\mathbb{R}^{1,n-1}$ are no longer characterized by the fact of having a relative angle of $\pi/2$**: if they belong to the lightcone, then they are orthogonal if and only if their relative angle is either 0 or π and, if one belongs to the interior and the other to the exterior of the lightcone, they can be orthogonal even if their relative angle is different than $\pi/2$.

Let us see an example using again the simple case of $n = 2$: let $u_1 = (t_1, x_1)$ and $u_2 = (t_2, x_2)$, then $u_1 \circ u_2 = 0 \iff t_1 t_2 = x_1 x_2$. It follows that the first two vectors of the canonical basis of $\mathbb{R}^n$, $e_1 = (1, 0)$ and $e_2 = (0, 1)$, are Lorentz-orthogonal, but also any couple of vectors of the form $u_1 = (1, a)$ and $u_2 = (a, 1)$ is Lorentz-orthogonal $\forall a \in \mathbb{R}$. Figure 11.3 depicts the case $0 < a < 1$.

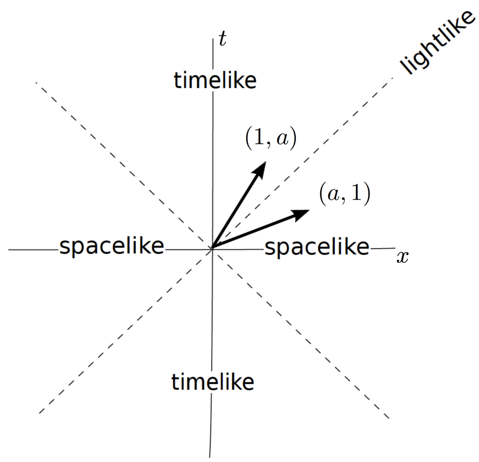


Figure 11.3: A graphical depiction of orthogonality in the 2 dimensional Lorentz space.

The relative position between the two vectors $(1, a)$ and $(a, 1)$ is not accidental, in fact, as we are going to prove, **a time-like vector u_1 and a space-like vector u_2**, both in $\mathbb{R}^{1,1}$, **are Lorentz-orthogonal if and only if the angles α and β that they form w.r.t. the horizontal axis are complementary, i.e. they sum to** $\pi/2$, modulo an integer multiple of π: $\alpha + \beta = \pi/2 + k\pi$, $k \in \mathbb{Z}$. To prove this, we just write down their polar coordinates $u_1 = (r \cos \alpha, r \sin \alpha)$ and $u_2 = (R \cos \beta, R \sin \beta)$, with $r, R > 0$ and $\alpha, \beta \in [0, 2\pi)$. Then, the Lorentz-orthogonality condition can be rewritten as

$$rR \cos \alpha \cos \beta - rR \sin \alpha \sin \beta = 0 \iff rR(\cos(\alpha + \beta)) = 0 \iff \cos(\alpha + \beta) = 0,$$

which implies $\alpha + \beta = \pi/2 + k\pi$.

The following result tell us, among other information, that the sum of two equioriented time-like vectors is still a time-like vector with the same orientation.

Corollary 11.2.3 *If x and y are non-zero, equioriented, causal vectors, then $x + y$ has the same orientation as x and y. Moreover, $x + y$ is light-like if and only if x and y are linearly dependent light-like vectors, otherwise $x + y$ is time-like.*

Proof. For the same reason given in the proof of theorem 11.2.1, we consider only the positively oriented case. In this case we have $x_1, y_1 > 0$, so $(x + y)_1 = x_1 + y_1 > 0$ and so $x + y$ is also positively oriented. Additionally, by direct computation we have:

$$\|x + y\|^2 = \|x\|^2 + 2(x \circ y) + \|y\|^2 ,$$

which is $\leqslant 0$ as the sum of three terms $\leqslant 0$. So, $x + y$ is either light-like or space-like. Finally, thanks to the previous theorem,

$$\|x + y\|^2 = 0 \iff \|x\|^2 = \|y\|^2 = x \circ y = 0$$

which is true if and only if x and y are light-like and linearly dependent. □

To fix the ideas, let us consider only vectors oriented toward the future, then the previous result can be re-written as follows:

- the sum of two linearly dependent light-like vectors is a light-like vector toward the future;
- the sum of two non-linearly dependent light-like vectors is a time-like vector toward the future;
- the sum of two time-like vectors is a time-like vector toward the future;
- the sum of a time-like vector and a light-like vector is a time-like vector toward the future.

Before stating the following corollary, we recall some definitions about cones taken from [5].

Def. 11.2.6 *Let $\mathcal{C}$ be a subset of a vector space V, then:*

- *$\mathcal{C}$ is a **cone** if, for all $t > 0$, $x \in \mathcal{C} \implies tx \in \mathcal{C}$;*
- *a cone $\mathcal{C}$ is **convex** if, for all $t \in [0, 1]$ and all couple of vectors $x, y \in \mathcal{C}$, $tx + (1-t)y \in \mathcal{C}$;*
- *a cone $\mathcal{C}$ is **proper** (or **regular**) if $\overline{\mathcal{C}} \cap -\overline{\mathcal{C}} = \{0\}$, where 0 is the zero vector of V, $\overline{\mathcal{C}}$ is the topological closure of $\mathcal{C}$ and $-\mathcal{C} := \{-x,\ x \in \mathcal{C}\}$.*

Corollary 11.2.4 *The set of all positively (respectively negatively) oriented time-like vectors forms an open connected proper convex cone in $\mathbb{R}^{1,n-1}$.*

Proof. Theorem 11.2.1 implies that the set of all positive (respectively negative) oriented time-like vectors forms a cone in $\mathbb{R}^{1,n-1}$. This cone is either the upper or the bottom part of the open set given by the interior of $\mathcal{C}^{n-1}$, thus it is connected, open and proper. Finally, let x and y two positively (resp. negatively) oriented time-like vectors.

The convexity of the proper cone they form follows from the combination of theorem 11.2.1 with theorem 11.2.1: for all $t \in (0, 1)$ set $\tilde{x} := tx$ and $\tilde{y} := (1-t)y$, then $\tilde{x}$ and $\tilde{y}$ belong to the same cone as x and y by theorem 11.2.1, thus their sum $\tilde{x} + \tilde{y} = tx + (1-t)y$ belongs to the same cone too for all $t \in (0, 1)$ by theorem 11.2.1. Since $x = (tx + (1-t)y)|_{t=1}$ and $y = (tx + (1-t)y)|_{t=0}$, we have that the convex combination $tx + (1-t)y$ belongs to the same cone as x and y for all $t \in [0, 1]$. □

Def. 11.2.7 (Time-like cone) *We call the cone of all positively (respectively, negatively) oriented time-like vectors in $\mathbb{R}^{1,n-1}$ the **future time-like cone** (respectively, the **past time-like cone**).*

These results explain why the concept of orientation is defined only for causal vectors: **the future and past lightcones and the future and past time-like cones are the connected components of two disjoint sets**, thus orientation allows us to single out which connected component we are dealing with.

Instead, **for all $n \geqslant 3$ the space-like region is connected** (but not convex because, for example, antipodal points w.r.t. the origin (which identifies a light-like vector) cannot communicate via a straight line segment), so specifying an orientation of a space-like vector does not single out any particular connected component of the space-like region. The only exception is represented by the case $n = 2$, but in the literature this special case is simply treated separately from the others without introducing a particular nomenclature.

11.2.5 Lorentz-orthonormality

As a preliminary remark, we notice that, since light-like vectors have null Lorentz pseudo-norm, **only time-like or space-like vectors can be unit vectors** in $\mathbb{R}^{1,n-1}$.

Def. 11.2.8 (Orthonormality in $\mathbb{R}^{1,n-1}$) *A set of $m \leqslant n$ mutually Lorentz-orthogonal unit vectors in $\mathbb{R}^{1,n-1}$ is said to be a Lorentz-orthonormal family. If $m = n$, then we have a Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$.*

Referring to the concepts introduced in section 11.2, we note that the matrix associated to the Lorentz pseudo-scalar product w.r.t. an orthonormal basis $u_1, \ldots, u_n$ coincides with the diagonal matrix $\eta = \text{diag}(-1, +1, +1, \ldots, +1)$. Thus, the signature of the Lorentz pseudo-scalar product is $(-, +, +, \ldots, +)$, so, thanks to property 21 of pseudo-scalar products, its index ν is 1.

By definition of time-like vector, it follows immediately that such vectors generate the subspace of $\mathbb{R}^{1,n-1}$ with highest dimension on which the Lorentz pseudo-scalar product is negative-definite. So, **the index ν of the Lorentz pseudo-scalar product is the maximal number of linearly independent time-like vectors**.

As a consequence, **in every Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$ there is exactly one time-like vector and $n-1$ space-like vectors**. By convention, the time-like vector is set to be the first basis vector. This justifies the following, more explicit, definition of Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$.

Def. 11.2.9 *A set of n vectors $\mathscr{B} = (u_1, \ldots, u_n)$ is a Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$ if*

$$u_i \circ u_j = \begin{cases} -1 & \text{if} \quad i = j = 1 \\ 1 & \text{if} \quad 2 \leqslant i, j \leqslant n, \ i = j \\ 0 & \text{if} \quad i \neq j \end{cases}.$$

Moreover, we say the basis is positive if u_1 is a positively oriented time-like vector.

By direct computation, it can be verified that **the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$ is a positive Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$**.

A Lorentz-orthonormal basis $\mathscr{B} = (u_1, \ldots, u_n)$ of $\mathbb{R}^{1,n-1}$ is an actual basis of $\mathbb{R}^n$, i.e. it is a family of n linearly independent vectors. To verify this, consider an arbitrary linear combination of the vectors of $\mathscr{B}$, i.e. $\tilde{u} = \alpha_1 u_1 + \cdots + \alpha_n u_n$, with $\alpha_i \in \mathbb{R}$ for all i. Then, by using Lorentz-orthonormality, we have that of the Lorentz pseudo-scalar products between $\tilde{u}$ and each vector of $\mathscr{B}$ is:

$$\tilde{u} \circ u_i = \alpha_i (u_i \circ u_i) = \begin{cases} -\alpha_1 & \text{if} \quad i = 1 \\ \alpha_i & \text{if} \quad 2 \leqslant i \leqslant n \end{cases},$$

$$\tilde{u} = 0 \implies 0 = 0 \circ u_i = \begin{cases} -\alpha_1 & \text{if} \quad i = 1 \\ \alpha_i & \text{if} \quad 2 \leqslant i \leqslant n \end{cases} \implies \alpha_1 = \cdots = \alpha_n = 0.$$

The existence of Lorentz-orthonormal bases is guaranteed by the following result.

Theorem 11.2.2 (Gram-Schmidt Lorentz-orthonormalization) *Let $\mathscr{B} = (u_1, \ldots, u_n)$ be a vector basis of $\mathbb{R}^{1,n-1}$ with u_1 a time-like vector. Then we can extract a basis $\mathscr{B}_L = (w_1, \cdots, w_n)$ from $\mathscr{B}$ such that:*

1. *$\mathscr{B}_L$ is a positive Lorentz-orthonormal basis.*

2. $\operatorname{span}(w_1, \ldots, w_k) = \operatorname{span}(u_1, \cdots, u_k)$ *for all* $k \in \{1, \ldots, n\}$.

Proof. This proof follows a path very similar to the proof for the usual Gram-Schmidt process, but with some minor tweaks. We start by setting

$$w_1 := \begin{cases} \frac{u_1}{|||u_1|||} & \text{if } u_1 \text{ is positive} \\ -\frac{u_1}{|||u_1|||} & \text{if } u_1 \text{ is negative} \end{cases},$$

and then we set

$$\begin{cases} v_2 := u_2 + (u_2 \circ w_1) w_1 \\ w_2 := \frac{v_2}{\|v_2\|} \end{cases}$$

and

$$\begin{cases} v_k := u_k + (u_k \circ w_1) w_1 - \sum\limits_{i=2}^{k-1} (u_k \circ w_k) w_k \\ w_k := \frac{v_k}{\|v_k\|} \end{cases} \qquad \text{for} \quad k \geqslant 3.$$

With such a construction, $\mathscr{B}_L = \{w_1, \cdots, w_n\}$ is a positive Lorentz orthonormal basis that verifies the wanted conditions. $\square$

Remark 11.2.2

We can apply a similar process when $(u_1, \ldots, u_m)$ is a set of linearly independent, space-like vectors to extract a set of space-like vectors $(w_1, \ldots, w_m)$ such that

$$w_i \circ w_j = \delta_{ij} \quad \text{and} \quad \operatorname{span}(w_1, \ldots, w_k) = \operatorname{span}(u_1, \ldots, u_k),$$

for all $k \in \{1, \ldots, m\}$.

11.2.6 Subspaces of the Lorentz n-space

Def. 11.2.10 *Let V be a vector subspace of $\mathbb{R}^n$. The Lorentz-orthogonal complement of V is:*

$$V^L = \{x \in \mathbb{R}^n : x \circ y = 0 \ \ \forall y \in V\}.$$

The properties of the Lorentz-orthogonal complement are given in the next result, where we use the symbol $V^\perp$ to denote the Euclidean orthogonal complement of V.

Lemma 11.2.2 *Let V be a vector subspace of $\mathbb{R}^n$ and write $\eta(V^\perp) := \{\eta x, \ x \in V^\perp\}$, then:*

1. *$V^L = \eta(V^\perp)$;*

2. *$(V^L)^L = V$, i.e. the Lorentz-orthogonalization is an involutive operation.*

Proof.

1. : it is a consequence of remark 11.2.1, i.e. $x \circ y = \langle x, \eta y \rangle = \langle \eta x, y \rangle$ $\forall x, y \in \mathbb{R}^n$, and of the fact that $\eta = \operatorname{diag}(-1, 1, \ldots, 1)$ verifies $\eta^2 = I_n$.

$\boxed{V^L \subseteq \eta(V^\perp)}$: let $\underline{x \in V^L}$, then $x \circ y = 0$ $\forall y \in V$ and so $\langle \eta x, y\rangle = 0$ $\forall y \in V$, i.e. $\eta x \in V^\perp$ but then $\underline{x = \eta(\eta x) \in \eta(V^\perp)}$.

$\boxed{\eta(V^\perp) \subseteq V^L}$: let $\underline{x \in \eta(V^\perp)}$, then $\eta x \in \eta(\eta(V^\perp)) = V^\perp$, so $\langle \eta x, y\rangle = 0$ $\forall y \in V$, but then $x \circ y = 0$ $\forall y \in V$, i.e. $\underline{x \in V^L}$.

2. : we have

$$x \in (V^L)^L \iff x \circ y = 0 \ \ \forall y \in V^L \underset{(1.)}{\iff} x \circ \eta z = 0 \ \ \forall z \in V^\perp,$$

which is equivalent to $\langle x, \eta(\eta z)\rangle = \langle x, z\rangle = 0$ $\forall z \in V^\perp$, i.e. $x \in (V^\perp)^\perp = V$. □

Property 1. of the previous lemma is clearly the consequence of the fact that $x \circ y = \langle x, \eta y\rangle = \langle \eta x, y\rangle$, hence Lorentz-orthogonality between two vectors x and y of $\mathbb{R}^{1,n-1}$ can be interpreted as the Euclidean orthogonality between one vector and the Euclidean orthogonal reflection of the other along its first coordinate.

We end this section with the classification of vector subspaces in $\mathbb{R}^{1,n-1}$ in 3 categories.

Def. 11.2.11 *Let V be a vector subspace of $\mathbb{R}^{1,n-1}$.*

- *V is **time-like** if it contains at least a time-like vector;*
- *V is **space-like** if every $x \in V \backslash \{0\}$ is space-like;*
- *V is **light-like** otherwise.*

$\mathbb{R}^n$, as improper subset of $\mathbb{R}^{1,n-1}$, is a time-like vector subspace, because $\mathbb{R}^n = \text{span}(e_1, \ldots, e_n)$ and e_1 is time-like.

It might be a little surprising that a time-like vector subspace V of $\mathbb{R}^{1,n-1}$ is defined just by requiring the existence of a time-like vector in V, while, on the contrary, we demand all non-null vectors of a space-like vector subspace to be space-like. The difference is justified by corollary 11.2.2 which imposes a strong constraint on the number of time-like vectors that can appear in a Lorentz-orthonormal basis of a vector subspace, either one or zero, and by the Gram-Schmidt Lorentz-orthonormalization theorem, which assures us that, even if a vector subspace V of $\mathbb{R}^{1,n-1}$ is generated by a basis with two ore more (necessarily non-orthogonal) time-like vectors, after orthonormalization, only one vector will remain time-like and both bases will generate the same subspace V. For this reason, requiring just the presence of a time-like vector in V is sufficient to characterize it as a time-like subspace.

Moreover, as a consequence of corollary 11.2.3, **only one light-like vector can appear in a basis of a light-like vector subspace and all the other basis vectors must be space-like**.

Figure 11.4 gives a graphical depiction of vector subspaces of $\mathbb{R}^{1,2}$ (which are necessarily either hyperplanes or straight lines passing through the origin).

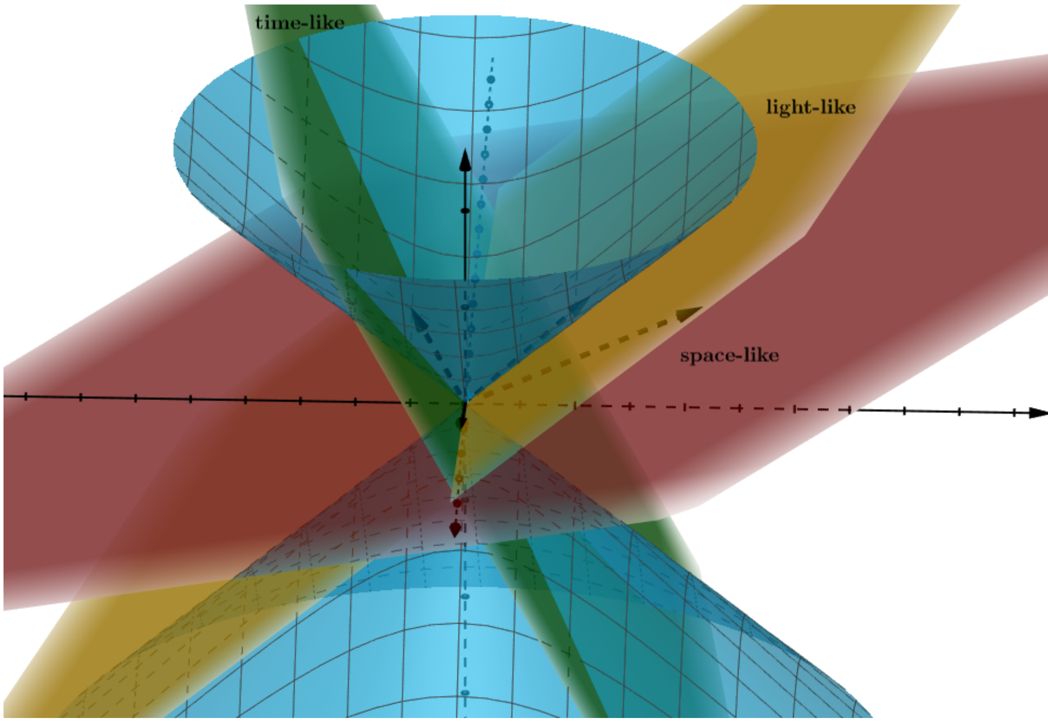


Figure 11.4: A graphical representation of vector subspaces in $\mathbb{R}^{1,2}$.

We note that:

- space-like vector subspaces intersect the double lightcone only in the origin, either perpendicularly w.r.t. the cone axis, or not, but without intersecting it in other points;
- time-like vector subspaces intersect the double lightcone not trivially;
- light-like vector subspaces are either straight lines defined by light-like vectors, or hyperplanes generated by a light-like vector and a space-like vector such that the hyperplane is tangent to the lightcone.

The relationship between time-like and space-like vector subspaces of $\mathbb{R}^{1,n-1}$ is established by the following result.

Theorem 11.2.3 *A vector subspace $V \subseteq \mathbb{R}^{1,n-1}$ is time-like if and only if V^L is space-like.*

Proof.

$\boxed{\Longrightarrow}$: suppose V is time-like and $x \in V$ a time-like vector. If $y \in V^L \backslash \{0\}$ then, by corollary 11.2.2, y must be space-like since x and y are Lorentz-orthogonal. Thus, V^L is space-like.

$\boxed{\Longleftarrow}$: we now assume V^L to be space-like. Since $(V^L)^L = V$, to prove that V is time-like it is enough to exhibit a time-like vector y Lorentz-orthogonal to V^L. A good candidate for this role is provided by the residual vector of the Lorentz-orthogonal projection on V^L of a vector $x \notin V^L$, as reported in eq. (11.3).

To this aim, we need to consider an orthonormal basis $\mathscr{B} = (u_1, \dots, u_m)$, $m = \dim(V^L) \leqslant n$, of V^L, which we know to exist thanks to remark 11.2.2.

Let $x \in \mathbb{R}^n$ be a time-like vector, then $x \notin V^L$ because V^L is space-like, so, by definition, all its vectors are space-like. The residual vector of the Lorentz-orthogonal projection of x on V^L is:

$$y = x - \sum_{i=1}^{m} (x \circ u_i) u_i,$$

notice that the coefficients ϵ_i appearing in eq. (11.3) quoted before are all $+1$ because V^L is space-like.

Eq. (11.3) assures us that y is Lorentz-orthogonal to every vector in V^L (in particular, to every u_i) and so, to finish the proof, we just have to check if y is a time-like vector. By the bilinearity of $\circ$ we have:

$$\begin{aligned} y \circ y &= \left(x - \sum_{i=1}^{m} (x \circ u_i) u_i \right) \circ y = x \circ y - \sum_{i=1}^{m} (x \circ u_i) \underbrace{(u_i \circ y)}_{=0} = x \circ y \\ &= x \circ \left(x - \sum_{i=1}^{m} (x \circ u_i) u_i \right) = x \circ x - \sum_{i=1}^{m} (x \circ u_i)(x \circ u_i) = x \circ x - \sum_{i=1}^{m} (x \circ u_i)^2 \\ &\leqslant x \circ x, \end{aligned}$$

but x is time-like, so $x \circ x < 0$ and so $y \circ y < 0$ and y is time-like. □

If we interchange the role of V with that of V^L and we use the fact that $(V^L)^L = V$, we get the following corollary.

Corollary 11.2.5 *A vector subspace $V \subseteq \mathbb{R}^{1,n-1}$ is space-like if and only if V^L is time-like.*

Finally, if $V \subseteq \mathbb{R}^{1,n-1}$ is a light-like vector subspace, since it cannot be neither time-like nor space-like due to the previous results, we get the set of light-like vector subspaces is stable w.r.t. Lorentz-orthogonalization.

Corollary 11.2.6 *A vector subspace $V \subseteq \mathbb{R}^{1,n-1}$ is light-like if and only if V^L is light-like.*

11.3 Lorentz transformations

Lorentz transformations play the role of orthogonal transformations for the Lorentz n-space, as formalized in the following definition.

Def. 11.3.1 *A Lorentz transformation on $\mathbb{R}^{1,n-1}$ is a map $\phi : \mathbb{R}^{1,n-1} \to \mathbb{R}^{1,n-1}$ that preserves the Lorentz pseudo-scalar product, i.e.*

$$\boxed{\phi(x) \circ \phi(y) = x \circ y} \qquad \forall x, y \in \mathbb{R}^{1,n-1}.$$

It is simple to prove that the set of Lorentz transformation on $\mathbb{R}^{1,n-1}$ form a group under composition.

Def. 11.3.2 (The Lorentz group) *The group of Lorentz transformations on $\mathbb{R}^{1,n-1}$ is called the Lorentz group and it is denoted with the symbol $\mathrm{O}(1, n-1)$ or $\mathscr{L}$.*

The following result is the equivalent of theorem 11.1.2 for the Lorentz n-space.

Theorem 11.3.1 *$\phi : \mathbb{R}^{1,n-1} \to \mathbb{R}^{1,n-1}$ is a Lorentz transformation if and only if it is linear and, given a Lorentz-orthonormal basis $(u_1, \dots, u_n)$ of $\mathbb{R}^{1,n-1}$, $(\phi(u_1), \dots, \phi(u_n))$ is a Lorentz-orthonormal basis.*

Proof.

$\boxed{\Longrightarrow}$: we start by assuming that $\phi : \mathbb{R}^{1,n-1} \to \mathbb{R}^{1,n-1}$ preserves the Lorentz pseudo-scalar product. Then, for all Lorentz-orthonormal basis $(u_1, \dots, u_n)$,

$$\phi(u_i) \circ \phi(u_j) = u_i \circ u_j = \begin{cases} -1 & \text{if} \quad i = j = 1 \\ \delta_{ij} & \text{otherwise} \end{cases},$$

so $(\phi(u_1), \cdots, \phi(u_n))$ is a Lorentz-orthonormal basis. Hence, for all $x \in \mathbb{R}^{1,n-1}$, thanks to eq. (11.2) we have:

$$\phi(x) = \sum_{i=1}^{n} \varepsilon_i(\phi(x) \circ \phi(u_i))\phi(u_i) \underset{(\phi \text{ Lorentz transf.})}{=} \sum_{i=1}^{n} \varepsilon_i(x \circ u_i)\phi(u_i),$$

with $\varepsilon_1 = -1$ and $\varepsilon_i = +1$ for all $2 \leqslant i \leqslant n$. Eq. (11.2) implies also that $x = \sum_{i=1}^{n} \varepsilon_i(x \circ u_i)u_i$, so, by writing $x_i = \varepsilon_i(x \circ u_i)$, we get:

$$\phi(\sum_{i=1}^{n} x_i u_i) = \sum_{i=1}^{n} x_i \phi(u_i), \qquad \forall x \in \mathbb{R}^{1,n-1}. \tag{11.7}$$

Following exactly the same line of reasoning used in the proof of theorem 11.1.2, we obtain the linearity of the Lorentz transformation.

$\boxed{\Longleftarrow}$: conversely, suppose that ϕ is linear and that, for any Lorentz-orthonormal basis $(u_1, \dots, u_n)$ of $\mathbb{R}^{1,n-1}$, $(\phi(u_1), \dots, \phi(u_n))$ is again a-Lorentz orthonormal basis of $\mathbb{R}^{1,n-1}$. Then,

$$\phi(x) = \phi(\sum_{i=1}^{n} x_i u_i) \underset{(\phi \text{ linear})}{=} \sum_{i=1}^{n} x_i \phi(u_i) \qquad \forall x \in \mathbb{R}^{1,n-1},$$

thus

$$\phi(x) \circ \phi(y) \quad = \quad \left(\sum_{i=1}^{n} x_i \phi(u_i)\right) \circ \left(\sum_{j=1}^{n} y_j \phi(u_j)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \phi(u_i) \circ \phi(u_j),$$

but $\phi(u_1) \circ \phi(u_1) = -1$ and $\phi(u_i) \circ \phi(u_j) = \delta_{ij}$ otherwise, so

$$\phi(x) \circ \phi(y) = \sum_{i=1}^{n} x_i y_i - x_1 y_1 = x \circ y,$$

i.e. ϕ is a Lorentz transformation. $\square$

Being a linear transformation in $\mathbb{R}^n$, a Lorentz transformation ϕ can be written as a matrix, usually denoted with $\Lambda \in \mathrm{M}(n, \mathbb{R})$. More precisely, we have the following definition.

Def. 11.3.3 (Lorentzian matrix) *A matrix $\Lambda \in \mathrm{M}(n, \mathbb{R})$ is called Lorentzian if the function $\phi_\Lambda : \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto \phi_\Lambda(x) := \Lambda x$ is a Lorentz transformation.*

Remark 11.3.1 *Λ can be interpreted as the matrix associated to ϕ_Λ w.r.t. the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$.*

As it happens for orthogonal transformations, Lorentzian matrices and Lorentz transformations can be identified, moreover, the algebraic properties of a Lorentzian matrix characterize completely the associated Lorentz transformation, as stated in the following theorem.

Theorem 11.3.2 *Let Λ be a $n \times n$ real matrix and $\eta = \mathrm{diag}(-1, 1, \ldots, 1)$. The following statements are equivalent.*

1. *Λ is a Lorentzian matrix*
2. *Λ^t is a Lorentzian matrix*
3. *The columns of Λ form a Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$*
4. *The rows of Λ form a Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$*
5. *Λ verifies $\Lambda^t \eta \Lambda = \eta$*
6. *Λ admits an inverse given by $\Lambda^{-1} = \eta \Lambda^t \eta$, which is a Lorentzian matrix*
7. *Λ verifies $\Lambda \eta \Lambda^t = \eta$*
8. *Λ preserves the quadratic form $q(x) = -x_1^2 + x_2^2 + \cdots + x_n^2 = \|x\|^2$ associated to the Lorentz pseudo-scalar product, i.e. $q(\Lambda x) = q(x)$, or, equivalently, $\|\Lambda x\|^2 = \|x\|^2$.*

Proof.

$\boxed{1. \iff 3.}$: thanks to remark 11.3.1, Λ has on the columns the vectors $(\phi_\Lambda(e_1), \ldots, \phi_\Lambda(e_n))$ of $\mathbb{R}^n$, which is a Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$. Lemma 11.3.1 can then be used to guarantee the equivalence between 1. and 3.

$\boxed{1. \iff 8.}$: the proof is exactly the same as the one of lemma 11.1.3, the only difference being that the Euclidean scalar product $\langle \; , \; \rangle$ as to be replaced by the Lorentz pseudo-scalar product $\circ$.

$\boxed{1. \iff 5.}$: by remark 11.2.1 we get, for all $x, y \in \mathbb{R}^n$,

$$x \circ y = \langle x, \eta y \rangle \quad \text{and} \quad \Lambda x \circ \Lambda y = \langle \Lambda x, \eta \Lambda y \rangle = \langle x, \Lambda^t \eta \Lambda y \rangle.$$

Λ is Lorentzian if and only if $\Lambda x \circ \Lambda y = x \circ y$ for all $x, y \in \mathbb{R}^n$, i.e.

$$\langle x, \Lambda^t \eta \Lambda y \rangle = \langle x, \eta y \rangle, \qquad \forall x, y \in \mathbb{R}^n,$$

but this is equivalent to $\Lambda^t \eta \Lambda = \eta$, which shows that 1. is equivalent to 5.

$\boxed{5. \iff 6.}$: since $\eta^2 = I_n$, property 5. can be restated by saying that Λ is Lorentzian if and only if $\eta \Lambda^t \eta \Lambda = I_n$, or that Λ is invertible with inverse $\Lambda^{-1} = \eta \Lambda^t \eta$. To show that Λ^{-1} is

Lorentzian we consider the associated transformation $\phi_\Lambda^{-1}(x) = \Lambda^{-1}x$ for all $x \in \mathbb{R}^n$ and we write

$$x \circ y = ((\phi_\Lambda \phi_{\Lambda^{-1}})x) \circ ((\phi_\Lambda \phi_{\Lambda^{-1}})y) = (\phi_\Lambda(\phi_{\Lambda^{-1}}(x)) \circ (\phi_\Lambda(\phi_{\Lambda^{-1}}(y)) \qquad \forall x, y \in \mathbb{R}^{1,n-1},$$

but ϕ_Λ is a Lorentz transformation, so

$$x \circ y = (\phi_{\Lambda^{-1}}(x)) \circ (\phi_{\Lambda^{-1}}(y)) \qquad \forall x, y \in \mathbb{R}^{1,n-1}.$$

$\boxed{1. \iff 7.}$: since $\Lambda^{-1} = \eta\Lambda^t\eta$ is Lorentzian, we can use property 5. on it and write

$$(\eta\Lambda^t\eta)^t\eta(\eta\Lambda^t\eta) = \eta \iff \eta\Lambda\eta\eta^2\Lambda^t\eta = \eta \iff \eta^{-1}\eta\Lambda\eta\Lambda^t\eta\eta = \eta^{-1}\eta\eta \iff \Lambda\eta\Lambda^t = \eta,$$

having used the fact that $\eta^2 = I_n$ and $\eta = \eta^t = \eta^{-1}$.

$\boxed{1. \iff 2.}$: immediate consequence of the equivalence between 1. and 5. and 1. and 6. In fact, by the first equivalence we have that Λ is Lorentizan if and only if $\Lambda^t\eta\Lambda = \eta$, by the latter this is equivalent to $\Lambda\eta\Lambda^t = \eta$, which is nothing by the first equivalence written for Λ^t, thus implying that Λ is Lorentzian if and only if Λ^t is.

$\boxed{1. \iff 4.}$: Λ is Lorentzian if and only if Λ^t is, if and only if (by 1. $\iff$ 3.) Λ^t has a Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$ on its columns, which is equivalent to say that Λ has a Lorentz-orthonormal basis of $\mathbb{R}^{1,n-1}$ on its rows. □

This theorem has important consequences. The first one follows immediately from property 8.

Corollary 11.3.1 *A Lorentz transformation preserves the likeness of a vector $x \in \mathbb{R}^{1,n-1}$, i.e. even if vectors can be modified by a Lorentz transformation,*

- *Lorentz-transformed time-like vectors still belong to the interior of the lightcone;*
- *Lorentz-transformed space-like vectors still belong to the exterior of the lightcone;*
- *Lorentz-transformed light-like vectors still belong to the lightcone.*

Thus, ***as a whole, the lightcone, the time-like and the space-like regions remain unaltered after a Lorentz transformation****.*

The lightcone is characterized by the equation $q(x) = 0$, i.e. it is the $0-$level set of the quadratic form q associated to the Lorentz pseudo-scalar product. Of course, there is nothing special about the value 0, as underlined by the next corollary.

Corollary 11.3.2 *The level sets of the quadratic form q associated to the Lorentz pseudo-scalar product are preserved by a Lorentz transformation, i.e. if $x \in \mathbb{R}^{1,n-1}$ belongs to the hyperboloid defined by $q(x) = c$, then also its Lorentz transformed (which is, in general, another vector) belongs to the same hyperboloid. In other words,* ***as a whole, the hyperboloid*** *$q(x) = c$, $c \in \mathbb{R}$,* ***remains unaltered after a Lorentz transformation****.*

Thanks to property 6. a Lorentzian matrix Λ is invertible with inverse $\Lambda^{-1} = \eta\Lambda^t\eta$ which is a Lorentzian matrix too, thus, if we identify Lorentz transformations ϕ with their Lorentzian matrices Λ, we can identify $\mathrm{O}(1, n-1)$ with a subgroup of $\mathrm{GL}(n, \mathbb{R})$ as follows:

$$\begin{aligned}\mathrm{O}(1, n-1) &= \{\phi : \mathbb{R}^n \to \mathbb{R}^n \ : \ \phi(x) \circ \phi(y) = x \circ y, \ \ \forall x, y \in \mathbb{R}^{1,n-1}\} \\ &= \{\Lambda \in \mathrm{GL}(n, \mathbb{R}) \ : \ \Lambda^t\eta\Lambda = \eta = \Lambda\eta\Lambda^t, \ \ \eta = \mathrm{diag}(-1, 1, \ldots, 1)\}.\end{aligned}$$

Notice that if $\circ$ is replaced by the Euclidean scalar product $\langle\ ,\ \rangle$ and η by I_n, then $\mathrm{O}(1, n-1)$ becomes the group $\mathrm{O}(n)$:

$$\begin{aligned}\mathrm{O}(n) &= \{\phi : \mathbb{R}^n \to \mathbb{R}^n \ : \ \langle\phi(x), \phi(y)\rangle = \langle x, y\rangle, \ \ \forall x, y \in \mathbb{R}^n\} \\ &= \{A \in \mathrm{GL}(n, \mathbb{R}) \ : \ A^t I_n A = I_n = A I_n A^t\}.\end{aligned}$$

By Binet's theorem, $\det(\eta) = \det(\Lambda^t\eta\Lambda) = \det(\Lambda^t)\det(\eta)\det(\Lambda)$, i.e. $\det(\Lambda^t)\det(\Lambda) = 1$, but $\det(\Lambda^t) = \det(\Lambda)$, thus

$$\boxed{\det(\Lambda) = \pm 1},$$

as it happens for an orthogonal matrix. As usual, we denote with

$$\mathrm{SO}(1, n-1) \equiv \mathscr{L}_+ := \{\Lambda \in \mathrm{O}(1, n-1) \ : \ \det(\Lambda) = 1\},$$

the **special** (or **proper**) **Lorentz group**.

Another important subgroup of $\mathrm{O}(1, n-1)$ is defined below.

Def. 11.3.4 (Positive Lorentz transformations) *We say that a Lorentz transformation $\Lambda \in \mathrm{O}(1, n-1)$ is positive (or positively oriented) if it preserves the orientation of the lightcone, i.e. if, for all $x \in \mathbb{R}^{1,n-1}$, x time-like,*

$$x_1 > 0 \implies (\Lambda x)_1 > 0.$$

The subgroup of $\mathrm{O}(1, n-1)$ given by positive Lorentz transformations is denoted as follows:

$$\mathrm{PO}(1, n-1) \equiv \mathscr{L}^\uparrow := \{\Lambda \in \mathrm{O}(1, n-1) \ : \ x \text{ time-like}, \ \ x_1 > 0 \implies (\Lambda x)_1 > 0\},$$

and called either ***positive*** *or* ***orthochronous*** *Lorentz group.*

The definition is exhaustive, in fact, if $\Lambda \in \mathrm{PO}(1, n-1)$ and $x \in \mathbb{R}^{1,n-1}$ is a time-like vector such that $x_1 < 0$, then $-x_1 > 0$ and $-(\Lambda x)_1 = (\Lambda(-x))_1 > 0$, hence $(\Lambda x)_1 < 0$. In other words, a positive Lorentz transformation preserves the orientation of time-like vectors in both the future and past time-like cones.

The subgroup

$$\mathrm{SPO}(1, n-1) \equiv \mathrm{SO}^+(1, n-1) \equiv \mathscr{L}_+^\uparrow := \mathrm{SO}(1, n-1) \cap \mathrm{PO}(1, n-1),$$

is called **special positive** or **proper orthochronous** or **restricted Lorentz group**. It can be proven to be the *connected component to the identity* of the Lorentz group.

More insights about the structure of the subgroups of the Lorentz group just defined are provided by the next theorems. In particular, it is quite remarkable that *the positivity of a Lorentz transformation is fully encoded in the first entry of the Lorentzian matrix Λ.*

Theorem 11.3.3 *Let $\Lambda = (\Lambda_{ij})_{1\leqslant i,j\leqslant n} \in \mathrm{O}(1, n-1)$.*

1. *$\Lambda \in \mathrm{PO}(1, n-1)$ if and only if $\Lambda_{11} \geqslant 1$.*
2. *$\Lambda \in \mathrm{O}(n) \cap \mathrm{PO}(1, n-1)$ if and only if $\Lambda_{11} = 1$.*

Proof.

1. : we start by showing that Λ is positive if and only if $\Lambda_{11} > 0$. Since Λ has a Lorentz-orthonormal basis on its rows and since the first vector of this basis is time-like, the entries of the first row of Λ form a the time-like vector that we denote with v_1.

Fixed a generic positive time-like vector $x \in \mathbb{R}^n$, Λ is positive if and only if Λx is a positive time-like vector, i.e. if the first entry of Λx is positive, but this is nothing but $\langle v_1, x\rangle$, so Λ is positive if and only if $\langle v_1, x\rangle > 0$.

However $\langle v_1, x\rangle = \langle v_1, \eta^2 x\rangle = \langle v_1, \eta(\eta x)\rangle = v_1 \circ \eta x$, thus Λ is positive if and only if $v_1 \circ \eta x > 0$. Notice that ηx is a negative vector because x is positive, so $x_1 > 0$, but $\eta = \mathrm{diag}(-1, 1, \dots, 1)$, so $(\eta x)_1 < 0$. By theorem 11.2.1, if v_1 is a negative vector too, then $v_1 \circ \eta x \leqslant 0$, so the fact that $v_1 \circ \eta x > 0$ implies that v_1 must be positive, i.e. $\Lambda_{11} > 0$.

Moreover, since v_1 a unit norm time-like vector, we have $\|v_1\|^2 = -1$, but $\|v_1\|^2 = -{\Lambda_{11}}^2 + |\bar{v}_1|^2$, thus

$$\Lambda_{11}^2 = 1 + |\bar{v}_1|^2 \geqslant 1,$$

which implies $\Lambda_{11} \geqslant 1$ since we have shown that Λ_{11} is positive.

2. : first of all we remark that 1. implies the following equivalence $\Lambda^t \in \mathrm{PO}(1, n-1) \iff \Lambda_{11} \geqslant 1$, in fact Λ^t is a Lorentzian matrix too an it shares the first entry, Λ_{11}, with Λ.

Now, if $\Lambda_{11} = 1$ then $\Lambda_{11}^2 = 1 + |\bar{v}_1|^2$ implies $|\bar{v}_1|^2 = 0$, so $v_1 = (1, 0, \dots, 0) \equiv {e_1}^t$ (the transposed of the first vector of the canonical basis of $\mathbb{R}^n$, recalling that we always work under the assumption that vectors in $\mathbb{R}^n$ are column vectors). It follows that the first column c_1 of Λ is $c_1 = v_1^t = e_1$. Therefore, Λ has the form

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix},$$

$A \in \mathrm{GL}(n-1, \mathbb{R})$ and, by theorem 11.3.2,

$$\Lambda^t \eta \Lambda = \eta \iff \begin{pmatrix} 1 & 0 \\ 0 & A^t \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix} \implies A^t A = I_{n-1},$$

but then $\Lambda^t \Lambda = I_n$ and so $\Lambda \in \mathrm{O}(n)$.

It remains only to prove that if $\Lambda \in \mathrm{O}(n) \cap \mathrm{PO}(1, n-1)$, then, necessarily, its first entry is equal to 1. To this aim, notice that, in this case, every column of Λ is an orthonormal vector w.r.t. both the Euclidean and the Lorentz scalar product, this implies that c_1, the first column of Λ, satisfies $\|c_1\|^2 = -1$ and $|c_1|^2 = 1$, so $\|c_1\|^2 + |c_1|^2 = 0$. But, since the first element of c_1 is actually Λ_{11}, we also have $\|c_1\|^2 = -{\Lambda_{11}}^2 + |\bar{c}_1|^2$ and $|c_1|^2 = {\Lambda_{11}}^2 + |\bar{c}_1|^2$, so

$$0 = \|c_1\|^2 + |c_1|^2 = -{\Lambda_{11}}^2 + |\bar{c}_1|^2 + {\Lambda_{11}}^2 + |\bar{c}_1|^2 = 2|\bar{c}_1|^2,$$

i.e. $\bar{c}_1 = 0$. Since $c_1 = (\Lambda_{11}, \bar{c}_1)$, this implies $c_1 = (\Lambda_{11}, 0, \dots, 0)$ and the only what that c_1 has unit Euclidean norm is that $\Lambda_{11} = \pm 1$, but $\Lambda_{11} \geqslant 1$ so only the option $\Lambda_{11} = 1$ is valid. $\square$

During the proof of the last theorem, we have also proven this result, which says, geometrically, that the only orthogonal transformations that preserve the orientation of the lightcone are the orthogonal transformations that leave the first axis invariant.

Corollary 11.3.3

$$\begin{aligned} \mathrm{O}(n) \cap \mathrm{PO}(1, n-1) &= \left\{ \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix} : A \in \mathrm{O}(n-1) \right\} \\ &\cong \mathrm{O}(n-1). \end{aligned}$$

We end this section by showing that the group of Lorentz transformations acts transitively on the vector subspace of $\mathbb{R}^{1,n-1}$ of different likeness, which become, then, homogeneous spaces w.r.t. the action of $\mathscr{L}$. In fact, we can prove an even stronger result: transitivity is guaranteed already by the subgroup of positive Lorentz transformations.

Theorem 11.3.4 *Let $\mathcal{V}_m^T$, $\mathcal{V}_m^S$ and $\mathcal{V}_m^L$ be the set of m-dimensional time-like, space-like and light-like vector subspace of $\mathbb{R}^{1,n-1}$, respectively. Then, the action of $\mathrm{PO}(1, n-1)$ is transitive on each of them.*

Proof.

Transitivity of $\mathrm{PO}(1, n-1)$ on $\mathcal{V}_m^T$. We start by observing that $\overline{V}_m := \mathrm{span}(e_1, \ldots, e_m) \cong \mathbb{R}^m$ is a time-like vector subspace of $\mathbb{R}^{1,n-1}$ because e_1 is time-like. We will prove the transitivity of $\mathrm{PO}(1, n-1)$ on $\mathcal{V}_m^T$ by showing that for a given time-like m-dimensional vector subspace V belonging to the set $\mathcal{V}_m^T$ there is a $\Lambda \in \mathrm{PO}(1, n-1)$ such that $\Lambda(V_m) = V$.

By theorem 11.2.2 we can guarantee the existence of a positive Lorentz-orthonormal basis $\mathscr{B} = (w_1, \ldots, w_m)$ of V. We use the vectors of $\mathscr{B}$ to define the matrix

$$\Lambda := \begin{pmatrix} | & \cdots & | \\ w_1 & \cdots & w_n \\ | & \cdots & | \end{pmatrix},$$

which, thanks to theorem 11.3.2, belongs to $\mathrm{PO}(1, n-1)$ because its columns form a positive Lorentz-orthonormal basis. By direct computation we have that $\Lambda(e_i) = w_i \ \forall i = 1, \ldots, m$, so, by linearity, $\Lambda(V_m) = V$.

Transitivity of $\mathrm{PO}(1, n-1)$ on $\mathcal{V}_m^S$. This time, we set $\overline{W}_m := \mathrm{span}(e_2, \ldots, e_{m+1})$ to be our m-dimensional space-time vector subspace of reference (note that the dimension m of a space-like vector subspace must be strictly less than n since it cannot contain any time-like vector by definition). Let $W \in \mathcal{V}_m^S$ and $u \in W^L$, where W^L is the Lorentz-orthogonal of W, a positive time-like vector such that $\|u\|^2 = -1$. Then, $\tilde{V} := \mathrm{span}(u, W)$ is a time-like vector subspace of $\mathbb{R}^{1,n-1}$ and so, by what we have just proven, it is connected to $\overline{V}_{m+1}$ by a positive Lorentz transformation, that we indicate again with $\Lambda \in \mathrm{PO}(1, n-1)$ for simplicity. So: $\Lambda(\tilde{V}) = \overline{V}_{m+1}$ and $\Lambda(u) = e_1$ (the time-like vectors are connected by Λ). If we prove that $\Lambda(W) = \overline{W}_m$, then, being Λ invertible, we have $\Lambda^{-1}(\overline{W}_m) = W$, thus proving the transitivity. To this aim, let $w \in W \subset \tilde{V}$. Then $\Lambda(w) \in \mathrm{span}(e_1)^L$ because

$$\Lambda(w) \circ e_1 = \Lambda(w) \circ \Lambda(u) = w \circ u = 0,$$

and so, since the Lorentz-orthogonal of span(e_1) is span$(e_2, \dots, e_n) = \overline{W}_{n-1}$,

$$\Lambda(w) \in \text{span}(e_1)^L \cap \Lambda(\tilde{V}) = \overline{W}_{n-1} \cap \overline{V}_{m+1} = \overline{W}_m,$$

which allows us to conclude that $\Lambda(W) \subset \overline{W}_m$. Finally, if $\mathscr{B} = (w_1, \dots, w_m)$ is a Lorentz-orthonormal basis of W, then $w_i \circ w_j = \delta_{ij}$ and so $\Lambda(w_i) \circ \Lambda(w_j) = \delta_{ij}$, which means that also $(\Lambda(w_i))_{1 \leqslant i \leqslant m}$ is a basis of $\overline{W_m}$. Hence, by linearity, we have $\Lambda(W) = \overline{W_m}$.

We leave the transitivity of $\text{PO}(1, n-1)$ on $\mathcal{V}_m^L$ as an exercise. □

The previous theorem has an important consequence: the transitivity of Lorentz transformations on the hyperboloids in $\mathbb{R}^{1,n-1}$ defined as set-level surfaces of the quadratic form associated to the Lorentz pseudo-scalar product. To understand how this is possible, it is sufficient to consider the particular case of $m = 1$: the elements of $\mathcal{V}_1^T$ and $\mathcal{V}_1^S$ are straight lines passing through the origin and belonging to the interior or the exterior of the lightcone, respectively.

Each one of these straight lines intersects the hyperboloid defined by the equation $\|x\|^2 = \alpha$, $\alpha \in \mathbb{R}\backslash\{0\}$, in two antipodal points w.r.t. the origin and belonging to the two disconnected hyperboloid sheets. Thus, the transitivity of positive Lorentz transformations on $\mathcal{V}_1^T$ and $\mathcal{V}_1^S$ implies that every couple of vectors belonging to same sheet of the hyperboloid can be connected by a positive Lorentz transformation.

The following result summarizes the previous arguments.

Corollary 11.3.4 *Let $\alpha \in \mathbb{R}\backslash\{0\}$, fixed.* $\text{O}(1, n-1)$ *acts transitively on the hyperboloid*

$$\mathcal{H}_\alpha^{n-1} = \{x \in \mathbb{R}^{1,n-1} : \|x\|^2 = \alpha\}.$$

Proof. We start by remarking that given $x \in \mathcal{H}_\alpha^{n-1}$, we have

$$V_x \cap \mathcal{H}_\alpha^{n-1} = \{x, -x\} \ , \quad \text{where} \quad V_x = \text{span}\{x\}.$$

Let $x, y \in \mathcal{H}_\alpha^{n-1}$. Then by theorem 11.3.4, there is a transformation $\Lambda \in PO(1, n-1)$ such that $\Lambda(V_x) = V_y$. Because Λ preserves the Lorentz pseudo-scalar,

$$\Lambda(V_x \cap \mathcal{H}_\alpha^{n-1}) = V_y \cap \mathcal{H}_\alpha^{n-1}$$

and so we have either

$$\Lambda(x) = y \quad \text{or} \quad \Lambda(x) = -y.$$

In the second case, it suffices to take $-\Lambda \in O(1, n-1)$ as the transitive action from x to y. □

We end this section with a nice exercise which shows that **the Frobenius norm of a Lorentzian matrix Λ depends only on its dimension and on the first matrix element**. The square Frobenius norm of $\Lambda \in \text{O}(1, n-1)$ is defined by

$$\|\Lambda\|^2 := \sum_{i=1}^n \sum_{j=1}^n \Lambda_{ij}^2,$$

we will compute it explicitly by using the fact that the rows and columns of Λ are a Lorentz-orthonormal basis. For that, we must distinguish between the first row and column, which are unit time-like vectors, and the remaining ones, which are unit space-like vectors. Let us consider $i = 1$, i.e. the first row:

$$\sum_{j=1}^{n} \Lambda_{1j}^2 = \Lambda_{11}^2 + \sum_{j=2}^{n} \Lambda_{1j}^2,$$

but, using the fact that the square norm of the first (time-like) row satisfies $-\Lambda_{11}^2 + \sum\limits_{j=2}^{n} \Lambda_{1j}^2 = -1$, we get $\sum\limits_{j=2}^{n} \Lambda_{1j}^2 = \Lambda_{11}^2 - 1$, which implies:

$$\sum_{j=1}^{n} \Lambda_{1j}^2 = 2\Lambda_{11}^2 - 1. \tag{11.8}$$

Analogously, for the first column we find the formula

$$\sum_{i=1}^{n} \Lambda_{i1}^2 = 2\Lambda_{11}^2 - 1. \tag{11.9}$$

The following rows, determined by an index i between 2 and n, are unit space-like vectors, hence they satisfy $-\Lambda_{i1}^2 + \sum\limits_{j=2}^{n} \Lambda_{ij}^2 = 1$, so $\sum\limits_{j=2}^{n} \Lambda_{ij}^2 = \Lambda_{i1}^2 + 1$ and

$$\sum_{j=1}^{n} \Lambda_{ij}^2 = 2\Lambda_{i1}^2 + 1, \quad i = 2, \dots, n.$$

Thus,

$$\begin{aligned}
\|\Lambda\|^2 &= \sum_{j=1}^{n} \Lambda_{1j}^2 + \sum_{i=2}^{n} \sum_{j=1}^{n} \Lambda_{ij}^2 \underset{(11.8)}{=} 2\Lambda_{11}^2 - 1 + \sum_{i=2}^{n} (2\Lambda_{i1}^2 + 1) = 2\Lambda_{11}^2 - 1 + n - 1 + \sum_{i=2}^{n} 2\Lambda_{i1}^2 \\
&= n - 2 + 2\sum_{i=1}^{n} \Lambda_{i1}^2 \underset{(11.9)}{=} n - 2 + 4\Lambda_{11}^2 - 2 \\
&= n - 4 + 4\Lambda_{11}^2.
\end{aligned}$$

Chapter 12

Möbius transformations

Möbius transformations are the main toolbox for the conformal model of hyperbolic geometry. Before discussing them rigorously, we give an intuitive introduction.

12.1 Introduction to Möbius transformations

The most natural setting for Möbius transformations is that of sphere, where a Möbius transformation is defined as a finite composition of basic geometric transformations called *inversions*, whose basic idea is depicted in Figure 12.1: by sliding continuously an elastic band on a ball, we can transform it into the equator, thus transforming the surface of the ball contained in the interior of the elastic band to half the surface of ball; in this situation the ball surface contained in the elastic band and the remaining one are isomorphic.

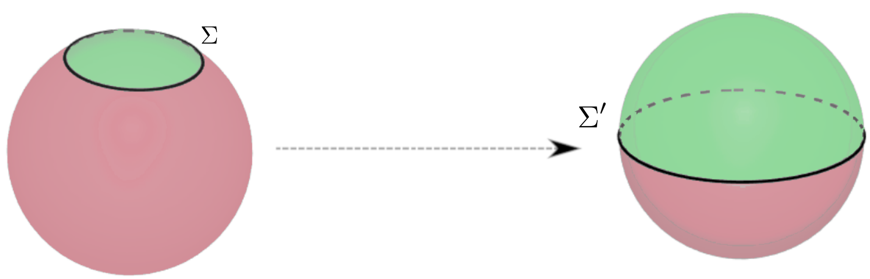


Figure 12.1: Inversion on a sphere: as the radius of the circle increased until reaching the diameter of the sphere, the 'interior' of the circle (in green) is diffeomorphic to the 'outside' (in pink).

Once written in mathematical terms, this continuous transformation that maps the spherical surface contained in a circle to the one left outside is called *inversion on a sphere*. By noticing that a circle can be obtained by cutting a sphere in $\mathbb{R}^3$ with a plane, it should not be surprising that, in the definition of inversion on a sphere of an arbitrary (finite) dimension n, a circle is replaced by a hypersphere, i.e. the intersection between a n-hyperplane and a n-sphere.

The group of Möbius transformations on the sphere, denoted by $\mathcal{M}(S^n)$ is the subgroup of $\mathrm{Aut}(S^n) = \{f : S^n \to S^n,\ f \text{ bijective}\}$ generated by inversions w.r.t. hyperspheres.

While it is true that handling Möbius transformations on the sphere is the most economical way to do it in terms of transformations involved (i.e. inversions), it is also true that it is more intuitive to analyze Möbius transformations in the Euclidean space. This can be done thanks to the stereographic projection introduced in chapter 1, which allows us to set up a bijection between the n-sphere minus the north pole and the hyperplane in $\mathbb{R}^{n+1}$ defined by $\{x \in \mathbb{R}^{n+1} : x_{n+1} = 0\}$, which can be identified with $\mathbb{R}^n$. To obtain a complete bijection we will need to introduce an artificial element, an abstract point denoted by ∞ and called the *point at infinity*.

Dealing with Möbius transformations in the Euclidean domain enlarged with the point at infinity comes with a price: we will see (corollary 12.4.1) that we no longer need only inversions to characterize them, but also other geometric operations, called reflections w.r.t. hyperplanes.

The importance of reflections and inversions motivates why we start the formal analysis of Möbius transformations by defining them in the next subsection.

12.2 Reflections and inversions

Consider a unit vector $a \in \mathbb{R}^n$, $|a| = 1$, then, by the orthogonal projection theorem, we have $\mathbb{R}^n = \text{span}(a) \oplus \text{span}(a)^\perp$ and so $\text{span}(a)^\perp := \{x \in \mathbb{R}^n : \langle x, a\rangle = 0\}$ is a $(n-1)$-dimensional vector subspace of $\mathbb{R}^n$, i.e. a hyperplane in $\mathbb{R}^n$ passing through the origin. If we consider the affine structure of $\mathbb{R}^n$, the vectors of $\text{span}(a)^\perp$ can be rigidly translated away from 0 by a real quantity t via the transformation $x \mapsto x - ta$. This operation identifies an **affine space of dimension** $n-1$ whose algebraic expression can be obtained by replacing x with $x - ta$ in the equation $\langle x, a\rangle = 0$, i.e. $\langle x - ta, a\rangle = 0 \iff \langle x, a\rangle - t|a|^2 = 0 \underset{|a|=1}{\iff} \langle x, a\rangle = t$.

These considerations justify the following definition.

Def. 12.2.1 (Hyperplane in $\mathbb{R}^n$) *Given $a \in \mathbb{R}^n$, $|a| = 1$, and $t \geqslant 0$, the hyperplane associated to a and t is the set*

$$P(a, t) := \{x \in \mathbb{R}^n,\ \langle x, a\rangle = t\}.$$

Thus:

- a is the **normal vector** to $P(a, t)$
- t is the **distance** between $P(a, t)$ and 0, which can be taken non-negative because its possible negative sign can be incorporated in the vector a without changing its unit norm by redefining t and a as follows:

$$t \mapsto |t| \geqslant 0 \text{ and } a \mapsto \text{signum}(t)\, a.$$

Geometrically, the reflection w.r.t. $P(a, t)$ is the map ρ that takes any point $x \in \mathbb{R}^n$ at a distance d from $P(a, t)$ to a point $\rho(x)$ which lies specularly on the other side of the hyperplane at the same distance d. The 2D version of this operation is depicted in Figure 12.2.

To understand how to analytically define $\rho(x)$ notice that, if we perform the sum $x + \lambda a$, then we move x perpendicularly w.r.t. $P(a, t)$ and by a magnitude λ. Let λ^* be such that $x + \lambda^* a \in P(a, t)$, then clearly $\rho(x) = x + 2\lambda^* a$. To make λ^* explicit we have to write $\langle x + \lambda^* a, a\rangle = t \iff \langle x, a\rangle + \lambda^* \cancelto{1}{\|a\|^2} = t \iff \lambda^* = t - \langle x, a\rangle$.

We formalize this concept in the following definition.

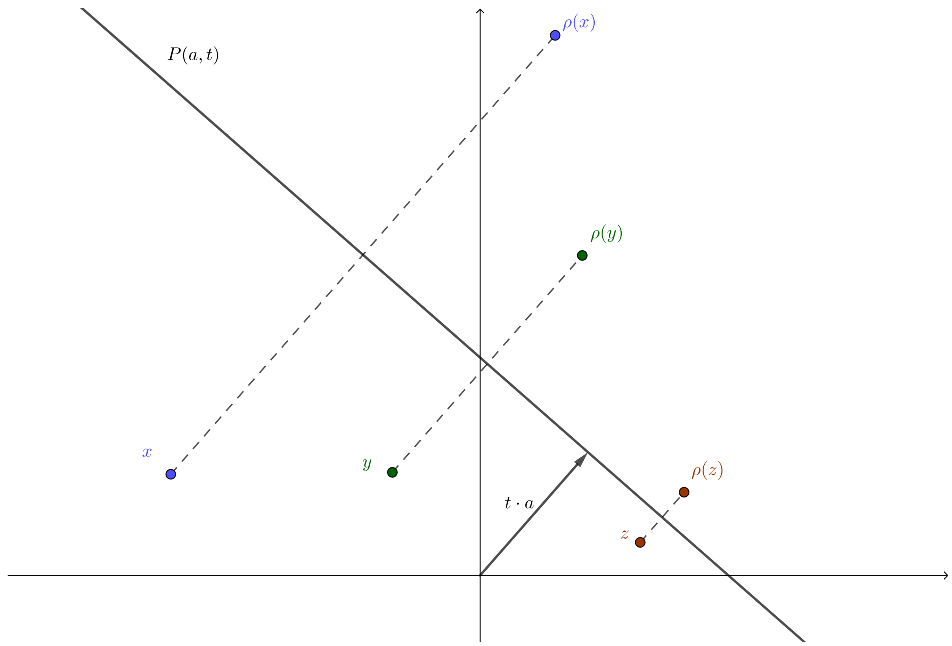


Figure 12.2: 2D graphical representation of a reflection w.r.t. a hyperplane, which is a straight line in two dimensions.

Def. 12.2.2 *A* ***reflection*** *in $\mathbb{R}^n$ w.r.t. the hyperplane $P(a,t)$ is the affine function:*

$$\begin{array}{rccl} \rho_{a,t}: & \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ & x & \mapsto & \rho_{a,t}(x) := x + 2(t - \langle x, a \rangle)a. \end{array} \tag{12.1}$$

$\rho_{a,t}(x)$ is said to be the reflection of x w.r.t. to the hyperplane $P(a,t)$.

If the dependence of ρ on the parameters a and t of the hyperplane $P(a,t)$ is not significant, we will simplify the notation and write ρ instead of $\rho_{a,t}$.

Figure 12.2 suggests some geometrical properties of $\rho_{a,t}$, e.g. the vectors belonging to $P(a,t)$ are unaffected by the action of $\rho_{a,t}$, if we apply it two times we come back to the original vector, so that the inverse of $\rho_{a,t}$ is itself and the Euclidean distance between any two reflected vectors is the same as the original distance.

These properties, and one more, are rigorously stated in the following theorem.

Theorem 12.2.1 *$\rho_{a,t}$ satisfies the following properties for all $x, y \in \mathbb{R}^n$:*

1. *$\rho_{a,t}(x) = x$ if and only if $x \in P(a,t)$*

2. *$\rho^2_{a,t}(x) = x$, i.e. $\rho_{a,t}$ is an involution, and so $\rho^2_{a,t} = id_{\mathbb{R}^n}$, i.e. $\rho_{a,t}$ is a bijection with ${\rho_{a,t}}^{-1} = \rho_{a,t}$*

3. *$\rho_{a,t}$ is a Euclidean isometry: $|\rho_{a,t}(x) - \rho_{a,t}(y)| = |x - y|$*

4. *$\rho_{a,t} \in \mathrm{O}(n) \iff t = 0$.*

Proof. The proofs can be obtained by direct computation.

1. :

$$\begin{aligned}
\rho_{a,t}(x) = x \quad &\Longleftrightarrow \quad x = x + 2(t - \langle x, a\rangle)a \quad (a \neq 0)\\
&\Longleftrightarrow \quad \langle x, a\rangle = t\\
&\Longleftrightarrow \quad x \in P(a,t).
\end{aligned}$$

2. :

$$\begin{aligned}
\rho_{a,t}^2(x) = \rho_{a,t}(\rho_{a,t}(x)) \quad &= \quad \rho_{a,t}(x) + 2(t - \langle \rho_{a,t}(x), a\rangle)a\\
&\qquad (\text{letting } s \equiv 2(t - \langle x, a\rangle))\\
&= \quad x + sa + 2(t - \langle x + sa, a\rangle)a\\
&= \quad x + sa + 2(t - \langle x, a\rangle)a - 2sa\\
&= \quad x + sa + sa - 2sa\\
&= \quad x.
\end{aligned}$$

3. : first of all, we note that

$$\rho_{a,t}(x) - \rho_{a,t}(y) = x - y - 2\langle y - x, a\rangle a,$$

so

$$\begin{aligned}
|\rho_{a,t}(x) - \rho_{a,t}(y)|^2 \quad &= \quad |(x-y) - 2\langle y - x, a\rangle a|^2\\
&= \quad |x-y|^2 - 4\langle y - x, a\rangle^2 + 4\langle y - x, a\rangle^2\\
&= \quad |x-y|^2.
\end{aligned}$$

4. : if $t \neq 0$, then $\rho_{a,t}(0) = 2ta \neq 0$ since $a \in S^n$, thus $\rho_{a,t}$ is not linear and thus it cannot belong to $\mathrm{O}(n)$. If $t = 0$ then, for all $x, y \in \mathbb{R}^n$,

$$\begin{aligned}
\langle \rho_{a,0}(x), \rho_{a,t}(y)\rangle &= \langle x - 2\langle x, a\rangle a, y - 2\langle y, a\rangle a\rangle\\
&= \langle x, y\rangle - 2\langle x, a\rangle\langle a, y\rangle - 2\langle y, a\rangle\langle x, a\rangle + 4\langle x, a\rangle\langle a, y\rangle \cancel{|a|^2}\ 1\\
&= \langle x, y\rangle .
\end{aligned}$$

$\square$

As a consequence of property **2.**, $\rho_{a,t}$ is bijective, thus a reflection w.r.t. a hyperplane in $\mathbb{R}^n$ maps bicontinuously any point in $\mathbb{R}^n$ that lies on one side of the hyperplane to a unique point that lies on the other side.

The concept of inversion deals with the same problem, with one (major) difference: the hypersurface w.r.t. the inversion is performed is not a hyperplane but a (hyper)sphere. While a hyperplane extends towards the infinite, a sphere is bounded, this fact implies that it is impossible to continuously fill the whole outer space to the spherical surface simply by reflecting its interior points w.r.t. the tangent hyperplane to the sphere at a point, a different geometrical operation is needed.

As proven by G. Bellavitis in his 1836 paper [2], this operation consists in mapping any point x inside the sphere to the unique point $\sigma(x)$ outside the sphere characterized by the following two properties: firstly, $\sigma(x)$ lies on the same line joining x with the center of the sphere; secondly, the norm of $\sigma(x)$ is inverted w.r.t. that of x.

The easiest way to formalize this idea is by first considering the unit sphere S^{n-1} in $\mathbb{R}^n$ centered in 0: if $x \in \mathbb{R}^n$ is such that $|x| < 1$, then $\sigma_{0,1}(x) := \frac{1}{|x|}\frac{x}{|x|} = \frac{1}{|x|^2}x$ is the desired inverted point outside S^{n-1}.

If, instead of S^{n-1}, we consider S_r^{n-1}, $r > 0$, then we can turn back to the previous case by applying $\sigma_{0,1}$ to $\frac{x}{r}$ and then by restoring the correct radius via a multiplication by r, denoted with m_r. Mathematically, this corresponds to the composed function $\sigma_{0,r} := m_r \circ \sigma_{0,1} \circ m_{1/r}$, hence, given any $x \in \mathbb{R}^n$ such that $|x| < r \iff |x/r| < 1$ we have

$$\sigma_{0,r}(x) = (m_r \circ \sigma_{0,1} \circ m_{1/r})(x) = m_r(\sigma_{0,1}(x/r)) = m_r\left(\frac{r^{\cancel{2}}}{|x|^2}\frac{x}{\cancel{r}}\right) = m_r\left(\frac{r}{|x|^2}x\right) = \left(\frac{r}{|x|}\right)^2 x.$$

The most general case is that of $S_{a,r}^{n-1}$, the $(n-1)$-sphere centered in $a \in \mathbb{R}^n$ with radius $r > 0$, i.e.

$$S_{a,r}^{n-1} := \{x \in \mathbb{R}^n \ : \ |x - a| = r\}.$$

Following the same argument used above, the inversion $\sigma_{a,r}$ will be given by the composition $\tau_a \circ \sigma_{0,r} \circ \tau_{-a}$, τ being the translation operator. Thus, for all $x \in \mathbb{R}^n \backslash \{a\}$ satisfying $|x - a| < r \iff |x - a|/r < 1$:

$$\sigma_{a,r}(x) = \tau_a(\sigma_{0,r}(x-a)) = \tau_a\left(\left(\frac{r}{|x-a|}\right)^2 (x-a)\right) = a + \frac{r^2}{|x-a|^2}(x-a).$$

Def. 12.2.3 *Let $a \in \mathbb{R}^n$ and $r > 0$, then the **inversion** in $\mathbb{R}^n$ w.r.t. the sphere $S_{a,r}^{n-1}$ is the non-linear function*

$$\begin{array}{rccl} \sigma_{a,r}: & \mathbb{R}^n\backslash\{a\} & \longrightarrow & \mathbb{R}^n\backslash\{a\} \\ & x & \mapsto & \sigma_{a,r}(x) := a + \frac{r^2}{|x-a|^2}(x-a). \end{array}$$

*$\sigma_{a,r}(x)$ is said to be the **inverse of** x w.r.t. to the sphere $S_{a,r}^{n-1}$.*

If the specification of the parameters a and r is not significant, we will simply write σ instead of $\sigma_{a,r}$.

The following result shows that the conjugation that is needed to define a generic inversion starting from the inversion w.r.t. to unit sphere can be operated by fusing into an affine function the multiplication by r and the translation by a.

Lemma 12.2.1 *Let:*

- *$\sigma_{0,1}$ the inversion w.r.t. S^{n-1}, the unit sphere centered in 0 in $\mathbb{R}^n$*
- *$\sigma_{a,r}$ the inversion w.r.t. the sphere $S_{a,r}^{n-1}$, $a \in \mathbb{R}^n$, $r > 0$*
- *for all $x \in \mathbb{R}^n$, $\phi(x) = a + rx$, with a and r as above.*

Then,

$$\boxed{\sigma_{a,r} = \phi \circ \sigma_{0,1} \circ \phi^{-1}}.$$

Proof. For all $x \in \mathbb{R}^n$ we have, by definition, $\sigma_{a,r}(x) = a + \frac{r^2}{|x-a|^2}(x-a)$, but, if we replace x by $r\frac{(x-a)}{|x-a|^2}$ as argument of ϕ we get

$$\phi\left(r\frac{x-a}{|x-a|^2}\right) = a + \frac{r^2}{|x-a|^2}(x-a),$$

so $\sigma_{a,r}(x) = \phi\left(r\frac{x-a}{|x-a|^2}\right)$. Since $\sigma_{0,1}(x) = \frac{x}{|x|^2}$,

$$\sigma_{0,1}\left(\frac{x-a}{r}\right) = \frac{x-a}{r}\frac{r^2}{|x-a|^2} = r\frac{x-a}{|x-a|^2},$$

hence $\sigma_{a,r}(x) = \phi \circ \sigma_{0,1}\left(\frac{x-a}{r}\right)$ for all $x \in \mathbb{R}^n$.

Finally, by solving $\phi(x) = a + rx$ w.r.t. x we obtain $\phi^{-1}(x) = \frac{x-a}{r}$, so that $\sigma_{a,r}(x) = \phi \circ \sigma_{0,1} \circ \phi^{-1}(x)$ for all $x \in \mathbb{R}^n$. □

Remark 12.2.1 Both **reflection** w.r.t. a hyperplane **and inversion** w.r.t. a sphere **are**, essentially, **one-dimensional operations**, in the sense that all the points belonging to the same straight line orthogonal to the hyperplane involved in a reflection are left on this straight line; in the same way, all the points belonging to the straight line passing through the origin of the sphere involved in an inversion are left on that line.

Contrarily to a reflection w.r.t. a hyperplane, which is defined on the whole $\mathbb{R}^n$, an inversions w.r.t. a sphere $S^{n-1}_{a,r}$ is defined on $\mathbb{R}^n$ deprived of the sphere center a.

Notice also that $\sigma_{a,r}(x)$ is the only point verifying

$$|\sigma_{a,r}(x) - a||x-a| = r^2, \tag{12.2}$$

thus, **the closer** x **is to the center of the sphere** a**, the further apart** $\sigma_{a,r}(x)$ **is sent on the straight line connecting** x **to** a. Figure 12.3 gives a graphical representation of this phenomenon in two dimensions.

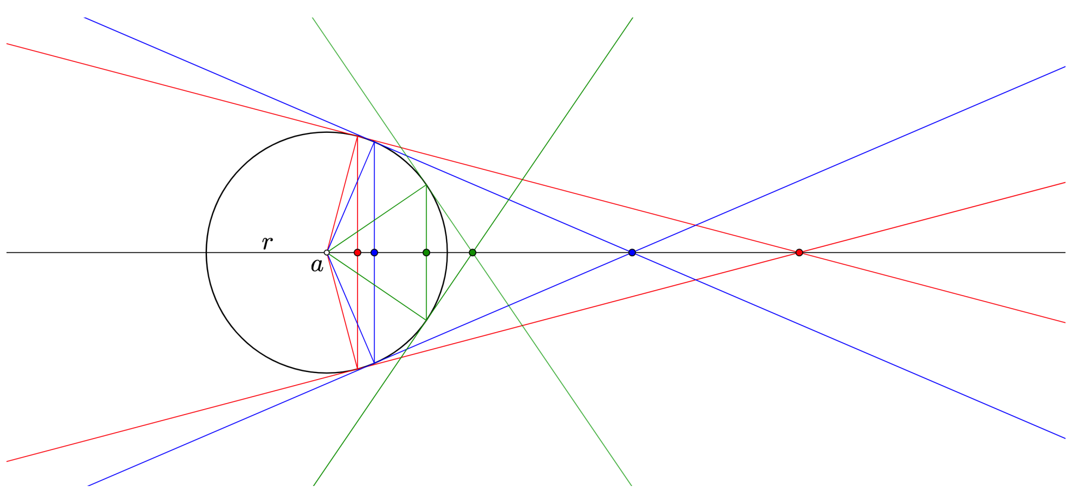


Figure 12.3: 2D graphical representation of an w.r.t the circle $S^1_{a,r}$.

As it can be seen, as we approach the center of the circle, the inverted point goes farther and farther. It is not difficult to imagine that, if we want to extend the concept of inversion to contemplate also the center of the sphere a, then we have to associate it to a point at infinite, that we will rigorously define later.

In the following theorem we prove the properties of inversions analogous to those of reflections. We put the accent on the fact that, by its own definition, an inversion cannot be an isometry, except for the points belonging to the sphere w.r.t. the inversion is performed.

Theorem 12.2.2 *Let $a \in \mathbb{R}^n$, $r > 0$ and σ be the inversion w.r.t. $S^{n-1}_{a,r}$. Then, for all $x, y \in \mathbb{R}^n \backslash \{a\}$:*

1. $\sigma_{a,r}(x) = x$ *if and only if* $x \in S^{n-1}_{a,r}$

2. $\sigma^2_{a,r}(x) = x$, *i.e.* $\sigma_{a,r}$ *is an involution, and so* $\sigma_{a,r}$ *is invertible with* ${\sigma_{a,r}}^{-1} = \sigma_{a,r}$

3. $|\sigma_{a,r}(x) - \sigma_{a,r}(y)| = \frac{r^2}{|x-a||y-a|}|x-y|$.

Proof. Let $x, y \in \mathbb{R}^n \backslash \{a\}$.

1. : the relationship $|\sigma_{a,r}(x) - a||x-a| = r^2$ always holds for $\sigma_{a,r}$, thus $\sigma_{a,r}(x) = x$ if and only if $|x-a|^2 = r^2$, i.e. $x \in S^n_{a,r}$.

2. :

$$\begin{aligned}
\sigma^2_{a,r}(x) = \sigma_{a,r}(\sigma_{a,r}(x)) &= a + \frac{r^2}{|\sigma_{a,r}(x) - a|^2}(\sigma_{a,r}(x) - a) \\
&= a + \frac{|x-a|^2}{r^2}\left(\frac{r^2}{|x-a|^2}(x-a)\right) = x.
\end{aligned}$$

3. :

$$\begin{aligned}
|\sigma_{a,r}(x) - \sigma_{a,r}(y)| &= \left|\frac{r^2}{|x-a|^2}(x-a) - \frac{r^2}{|y-a|^2}(y-a)\right| = r^2\left|\frac{x-a}{|x-a|^2} - \frac{y-a}{|y-a|^2}\right| \\
&= r^2\left\langle \frac{x-a}{|x-a|^2} - \frac{y-a}{|y-a|^2}, \frac{x-a}{|x-a|^2} - \frac{y-a}{|y-a|^2}\right\rangle^{\frac{1}{2}} \\
&= r^2\left|\frac{|x-a|^2}{|x-a|^4} - 2\frac{\langle x-a, y-a\rangle}{|x-a|^2|y-a|^2} + \frac{|y-a|^2}{|y-a|^4}\right|^{\frac{1}{2}} \\
&= r^2\left|\frac{1}{|x-a|^2} - 2\frac{\langle x-a, y-a\rangle}{|x-a|^2|y-a|^2} + \frac{1}{|y-a|^2}\right|^{\frac{1}{2}} \\
&= r^2\left|\frac{|y-a|^2 - 2\langle x-a, y-a\rangle + |x-a|^2}{|x-a|^2|y-a|^2}\right|^{\frac{1}{2}} \\
&= r^2\left|\frac{\langle (x-a)-(y-a), (x-a)-(y-a)\rangle}{|x-a|^2|y-a|^2}\right|^{\frac{1}{2}} \\
&= r^2\left|\frac{\langle x-y, x-y\rangle}{|x-a|^2|y-a|^2}\right|^{\frac{1}{2}} = r^2\frac{|x-y|}{|x-a||y-a|}.
\end{aligned}$$

□

Property **3.** of theorem 12.2.2 states that, if x and y do not belong to the sphere $S^{n-1}_{a,r}$, then their Euclidean distance after the application of $\sigma_{a,r}$ will be proportional to their original Euclidean distance inside the sphere, with a non-linear proportionality coefficient that depends on both x and y through the formula $r^2/|x-a||y-a|$.

This property is crucial to understand the profound link between Möbius transformations and the so-called *cross ratio*.

The following theorems underline the importance of reflections and inversions, by relating them to the Euclidean isometries and similarities.

Theorem 12.2.3 *Every Euclidean isometry of* $\mathbb{R}^n$ *is a composition of at most* $n+1$ *reflections.*

Proof. As a preliminary observation, we recall that, by theorem 11.1.1, all Euclidean isometry $f : \mathbb{R}^n \to \mathbb{R}^n$ can be written as $f(x) = a + \phi(x)$ with $a \in \mathbb{R}^n$ and $\phi \in \mathrm{O}(n)$, for all $x \in \mathbb{R}^n$. Hence, an isometry is an orthogonal transformation if and only is it leaves 0 fixed.

The proof is constructive and it is based on the following strategy:

- we start by building the first reflection ρ_0 such that $\phi_0 := \rho_0 \circ f$ belongs to $\mathrm{O}(n)$;
- then we build *by induction* the other n reflections $\rho_1, \ldots, \rho_n$ such that, for all $k \in \{1, \ldots, n\}$, the transformation $\phi_k := \rho_k \circ \rho_{k-1} \circ \cdots \circ \rho_0 \circ f$ belongs to $\mathrm{O}(n)$ and leaves all the first k vectors of the canonical basis $e_1, \ldots, e_k$ of $\mathbb{R}^n$ fixed;
- when we arrive to $k = n$ we obtain an orthogonal (hence linear) transformation $\phi_n = \rho_n \circ \cdots \circ \rho_0 \circ f$ which leaves all the vectors of the canonical basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$ fixed. The matrix associated to ϕ_n w.r.t. the canonical basis of $\mathbb{R}^n$ is of course I_n, so $\phi_n \equiv id_{\mathbb{R}^n}$;
- finally, we observe that

  $$\phi_n = \rho_n \circ \cdots \circ \rho_0 \circ f = id_{\mathbb{R}^n} \iff (\rho_0 \circ \cdots \circ \rho_n) \circ (\rho_n \circ \cdots \circ \rho_0) \circ f = \rho_0 \circ \cdots \circ \rho_n$$

  but the reflections ρ_i are involutions, i.e. $\rho_i^2 = id_{\mathbb{R}^n}$ for all $i = 0, 1, \ldots, n$, so that $f = \rho_0 \circ \cdots \circ \rho_n$, which proves that we need at most $n+1$ reflections to represent any arbitrary isometry f.

Let us start by building the reflection ρ_0 such that $\phi_0 = \rho_0 \circ f \in \mathrm{O}(n)$. We write $x_0 := f(0) = a$ and we set

$$\rho_0 := \begin{cases} id_{\mathbb{R}^n} & \text{if } x_0 = 0 \\ \rho_{\frac{x_0}{|x_0|}, \frac{|x_0|}{2}} & \text{otherwise} \end{cases},$$

ϕ_0 is clearly an isometry as composition of two isometries, f and the reflection ρ_0. Let us verify if it leaves 0 fixed: by using definition (12.1) we have

$$\rho_0(x_0) = \begin{cases} id_{\mathbb{R}^n}(0) = 0 & \text{if } x_0 = 0 \\ \rho_{\frac{x_0}{|x_0|}, \frac{|x_0|}{2}}(x_0) = x_0 + 2\left(\frac{|x_0|}{2} - \left\langle x_0, \frac{x_0}{|x_0|} \right\rangle\right) \frac{x_0}{|x_0|} = x_0 - x_0 = 0 & \text{otherwise} \end{cases},$$

so $\phi_0(0) = \rho_0(f(0)) = \rho_0(x_0) = 0$, hence ϕ_0 is indeed an orthogonal transformation.

Let us pass to the construction of the remaining reflections $\rho_1, \ldots, \rho_n$. As we have previously declared, we will use the induction technique, so we need to start by proving that there exists a reflection ρ_1 such that $\phi_1 := \rho_1 \circ \phi_0 = \rho_1 \circ \rho_0 \circ f$ is an orthogonal transformation that leaves e_1 fixed. We set $x_1 := \phi_0(e_1) - e_1$ and we define such a reflection as follows

$$\rho_1 := \begin{cases} id_{\mathbb{R}^n} & \text{if } \phi_0(e_1) = e_1 \\ \rho_{\frac{x_1}{|x_1|},0} & \text{otherwise} \end{cases},$$

ϕ_1 is either ϕ_0, which is an orthogonal transformation, or the composition of ϕ_0 with the reflection $\rho_{\frac{x_1}{|x_1|},0}$, which is orthogonal thanks to property **4.** of theorem 12.2.1, in both cases $\phi_1 \in \mathrm{O}(n)$.

We also observe that $\phi_1(e_1) = \rho_1(\phi_0(e_1))$, which is equal to $id_{\mathbb{R}^n}(e_1) = e_1$ if $\phi_0(e_1) = e_1$, otherwise:

$$\phi_1(e_1) = \rho_1 \circ \phi_0(e_1) = \phi_0(e_1) - 2\langle \phi_0(e_1), x_1 \rangle \frac{x_1}{|x_1|^2},$$

but $|\phi_0(e_1) - x_1|^2 = |\phi_0(e_1)|^2 - 2\langle \phi_0(e_1), x_1 \rangle + |x_1|^2$, so $-2\langle \phi_0(e_1), x_1 \rangle = |\phi_0(e_1) - x_1|^2 - |\phi_0(e_1)|^2 - |x_1|^2$, thus

$$\begin{aligned} \phi_1(e_1) &= \phi_0(e_1) + \big(\underbrace{|\phi_0(e_1) - x_1|^2}_{=1} - \underbrace{|\phi_0(e_1)|^2}_{=1} - |x_1|^2 \big) \frac{x_1}{|x_1|^2} \\ &= \phi_0(e_1) - |x_1|^2 \frac{x_1}{|x_1|^2} = \phi_0(e_1) - x_1 \\ &= e_1, \end{aligned}$$

where we have used the fact that $|\phi_0(e_1) - x_1| = |e_1| = 1$ and $|\phi_0(e_1)| = 1$ because $\phi_0 \in \mathrm{O}(n)$ and $|e_1| = 1$. To resume, $\phi_1 \in \mathrm{O}(n)$ and it leaves e_1 fixed, thus the first induction step is fulfilled.

We now assume that, for all[1] $k \in \{3, \ldots, n\}$ there exists $\phi_{k-1} \in \mathrm{O}(n)$ that fixes $e_1, \ldots, e_{k-1}$. Let $x_k := \phi_{k-1}(e_k) - e_k$ and define

$$\rho_k := \begin{cases} id_{\mathbb{R}^n} & \text{if } \phi_{k-1}(e_k) = e_k \\ \rho_{\frac{x_k}{|x_k|},0} & \text{otherwise} \end{cases}.$$

By repeating exactly the same computations performed in the case of ϕ_1, it can be verified that $\phi_k := \rho_k \circ \phi_{k-1} \in \mathrm{O}(n)$, and that ϕ_k leaves e_k fixed. To verify that ϕ_k leaves also $e_1, \ldots, e_{k-1}$ fixed we write

$$\phi_k(e_i) = \rho_k(\phi_{k-1}(e_i)) = \rho_{\frac{x_k}{|x_k|},0}(\phi_{k-1}(e_i)) = \phi_{k-1}(e_i) - 2\langle x_k, \phi_{k-1}(e_i) \rangle \frac{x_k}{|x_k|^2},$$

but, for $1 \leqslant i < k \leqslant n-1$,

$$\phi_{k-1}(e_i) = e_i \tag{12.3}$$

by hypothesis of induction and so

$$\begin{aligned} \langle x_k, \phi_{k-1}(e_i) \rangle &\underset{(12.3)}{=} \langle \phi_{k-1}(e_k) - e_k, e_i \rangle = \langle \phi_{k-1}(e_k), e_i \rangle - \cancelto{0}{\langle e_k, e_i \rangle} = \langle \phi_{k-1}(e_k), e_i \rangle \\ &\underset{(12.3)}{=} \langle \phi_{k-1}(e_k), \phi_{k-1}(e_i) \rangle \underset{\phi_{k-1} \in \mathrm{O}(n)}{=} \cancelto{0}{\langle e_k, e_i \rangle} \\ &= 0, \end{aligned}$$

[1]notice that for $k = 1, 2$ we have already built ϕ_0 and ϕ_1, so we do not need to *assume* their existence.

which implies that

$$\phi_k(e_i) = \underbrace{\phi_{k-1}(e_i)}_{=e_i} - 2\underbrace{\langle x_k, \phi_{k-1}(e_i)\rangle}_{=0} \frac{x_k}{|x_k|^2} = e_i, \quad \forall i = 1 \leqslant i < k \leqslant n-1.$$

Hence, $\phi_k \in \mathrm{O}(n)$ and fixes $e_1, \ldots, e_k$, which is what we had to verify in order to conclude the proof. □

We can easily extend the previous result to Euclidean similarities.

Corollary 12.2.1 *Every Euclidean similarity is a composition of at most $n+3$ reflections and inversions.*

Proof. First we treat the special case of the similarity $g(x) = kx$, $k > 0$. Let $\sigma_1 := \sigma_{0,1}$ and $\sigma_2 := \sigma_{0,\sqrt{k}}$. Then, by direct computation, we get

$$\sigma_2 \circ \sigma_1(x) = \sigma_2\left(\frac{x}{|x|^2}\right) = kx = g(x).$$

More generally, a similarity $f \in \mathcal{S}(\mathbb{R}^n)$, by theorem 11.1.1, can be written as $f(x) = a + k\phi(x)$ with $a \in \mathbb{R}^n$, $k > 0$ and $\phi \in \mathrm{O}(n)$. As we have seen in the proof of the previous theorem, by letting $x_0 = f(0) = a$ and $\rho_0 = \rho_{\frac{x_0}{|x_0|}, \frac{|x_0|}{2}}$, we have

$$\begin{aligned} f(x) &= \rho_0 \circ f(x) = k\phi(x) \\ &= \sigma_2 \circ \sigma_1 \circ \phi(x), \end{aligned}$$

and, as seen in proof of the previous result, ϕ can be decomposed into n reflections and so the corollary is proven. □

12.3 The stereographic projection as an inversion and the one point compactification of $\mathbb{R}^n$

In chapter 1 we have shown that the stereographic projection allows us to identify the n-sphere without the north (or south) pole with the hyperplane in $\mathbb{R}^{n+1}$ defined by:

$$P(e_{n+1}, 0) = \{x \in \mathbb{R}^{n+1} \ : \ \langle x, e_{n+1} \rangle = 0\} = \{(x_1, \dots, x_n, 0), \ x_1, \dots, x_n \in \mathbb{R}\} \equiv \mathbb{R}^n \times \{0\} \cong \mathbb{R}^n.$$

In order to keep the analysis as simple as possible, we will implicitly identify $\mathbb{R}^n$ with the hyperplane in $\mathbb{R}^{n+1}$ passing through the origin and orthogonal to e_{n+1} whenever needed.

For the scopes of this section, the roles of $\mathbb{R}^n$ and the n-sphere must be reversed, as specified in the following definition.

Def. 12.3.1 *The map*

$$\begin{array}{rccl} \pi : & \mathbb{R}^n & \xrightarrow{\sim} & S^n \backslash \{e_{n+1}\} \\ & x & \longmapsto & \pi(x) := \left(\frac{2x_1}{1+|x|^2}, \dots, \frac{2x_n}{1+|x|^2}, \frac{|x|^2-1}{1+|x|^2} \right), \end{array}$$

is called the stereographic projection from $\mathbb{R}^n$ to $S^n \backslash \{e_{n+1}\}$.

π coincides with the map φ_N^{-1} defined in eq. (1.6) with $R = 1$, i.e. the inverse stereographic chart relative to the north pole of S^n, that is a bijection between $\mathbb{R}^n$ and $S^n \backslash \{e_{n+1}\}$.

Theorem 12.3.1 *For all $x = (x_1, \dots, x_n, 0) \in \mathbb{R}^n$, the stereographic projection $\pi(x)$ can be written as follows:*

$$\pi(x) = x + \frac{|x|^2 - 1}{1 + |x|^2}(e_{n+1} - x). \tag{12.4}$$

The expression $r(s) = x + s(e_{n+1} - x)$, with $s \in \mathbb{R}$ clearly represents a straight line such that $r(0) = x$ and $r(1) = e_{n+1}$, the coefficients $s = \frac{|x|^2-1}{1+|x|^2} \neq 1$ is the only value which guarantees that $\pi(x)$ lies on the unit sphere, i.e. that $|\pi(x)| = 1$.

Proof. Let us write $\pi(x) = x + s(e_{n+1} - x) = (1 - s)x + se_{n+1}$, then, by direct computation,

$$\begin{aligned} |\pi(x)| = 1 \quad &\iff \quad |x|^2(1-s)^2 + 2s(1-s)\underbrace{\langle x, e_{n+1}\rangle}_{=0} + s^2 \cancel{|e_{n+1}|^2}^{\,1} = 1 \\ &\iff \quad |x|^2 = \frac{1-s^2}{(1-s)^2} = \frac{1+s}{1-s} \iff s = \frac{|x|^2-1}{1+|x|^2}, \end{aligned}$$

if we introduce this expression of s in $\pi(x) = (1 - s)x + se_{n+1}$ we get

$$\begin{aligned} \pi(x) &= \left(1 - \frac{|x|^2-1}{1+|x|^2}\right)(x_1, \dots, x_n, 0) + \frac{|x|^2-1}{1+|x|^2}(0, \dots, 0, 1) \\ &= \left(\frac{2}{1+|x|^2}\right)(x_1, \dots, x_n, 0) + \frac{|x|^2-1}{1+|x|^2}(0, \dots, 0, 1) = \left(\frac{2x_1}{1+|x|^2}, \dots, \frac{2x_n}{1+|x|^2}, \frac{|x|^2-1}{1+|x|^2}\right), \end{aligned}$$

which coincides with the definition of stereographic projection given in (12.3.1). $\square$

Let us now connect the stereographic projection with inversions. The first hint about the fact that π acts as an inversion is provided by the its 3D depiction in Figure 12.4.

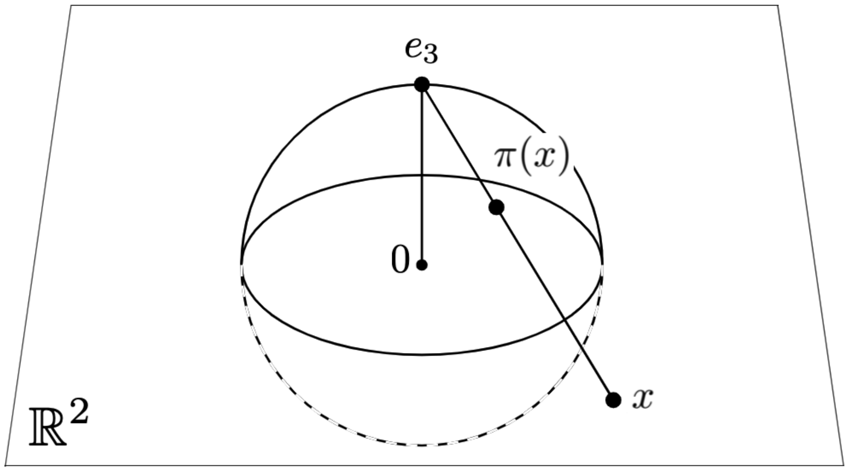


Figure 12.4: Stereographic projection in 3D.

In order to formalize the link between stereographic projection and inversions, consider $S^n_{e_{n+1},\sqrt{2}}$, the sphere with radius $\sqrt{2}$ centered in e_{n+1}, then the inversion $\sigma_{e_{n+1},\sqrt{2}} : \mathbb{R}^{n+1} \backslash \{e_{n+1}\} \xrightarrow{\sim} \mathbb{R}^{n+1} \backslash \{e_{n+1}\}$, $\sigma_{e_{n+1},\sqrt{2}}(x) = e_{n+1} + \frac{2}{|x-e_{n+1}|^2}(x - e_{n+1})$, is a bijection. If we restrict this bijection to the hyperplane $P(e_{n+1}, 0) \cong \mathbb{R}^n$ we still get a bijection with its codomain. It turns out that the codomain of $\sigma_{e_{n+1},\sqrt{2}}\big|_{\mathbb{R}^n}$ is $S^n \backslash \{e_{n+1}\}$ and that its analytical form coincides with the one of the stereographic projection.

Theorem 12.3.2 *It holds that*

$$\boxed{\pi = \sigma_{e_{n+1},\sqrt{2}}\big|_{\mathbb{R}^n}}. \tag{12.5}$$

Proof. We simply have to apply $\sigma_{e_{n+1},\sqrt{2}}$ to $(x, 0)$, with $x \in \mathbb{R}^n$, to verify that we get $\pi(x)$. To this aim, we first remark that: $|x - e_{n+1}|^2 = |x|^2 - 2\cancelto{0}{\langle x, e_{n+1}\rangle} + \cancelto{1}{|e_{n+1}|^2} = 1 + |x|^2$, so

$$\begin{aligned} \sigma_{e_{n+1},\sqrt{2}}(x) &:= e_{n+1} + \frac{2}{|x - e_{n+1}|^2}(x - e_{n+1}) = e_{n+1} + \frac{2}{1 + |x|^2}(x - e_{n+1}) \\ &= (0, \dots, 0, 1) + \frac{2}{1 + |x|^2}(x_1, \dots, x_n, -1) \\ &= \left(\frac{2x_1}{1 + |x|^2}, \dots, \frac{2x_n}{1 + |x|^2}, \frac{|x|^2 - 1}{1 + |x|^2} \right) = \pi(x). \end{aligned}$$

□

The so-called one point compactification of $\mathbb{R}^n$ is obtained by extending the stereographic projection $\pi : \mathbb{R}^n \xrightarrow{\sim} S^n \backslash \{e_{n+1}\}$ to a larger space, denoted with $\hat{\mathbb{R}}^n$, in such a way that the extended map $\hat{\pi} : \hat{\mathbb{R}}^n \to S^n$ is a bijection that encompasses also e_{n+1}. This is done by adding one single abstract point to $\mathbb{R}^n$ which is called *point at the infinity* and denoted with ∞ because no finite value of $\mathbb{R}^n$ can be mapped to e_{n+1} via π.

Def. 12.3.2 (One point compactification of $\mathbb{R}^n$) *Let ∞ be an abstract point not belonging to $\mathbb{R}^n$ and called the* ***point at infinity****. Let $\hat{\mathbb{R}}^n := \mathbb{R}^n \sqcup \{\infty\}$ be the disjoint union of $\mathbb{R}^n$ with the singleton set containing the point at infinity. Finally, let $\hat{\pi} : \hat{\mathbb{R}}^n \to S^n$ be the extension of the stereographic projection given by the bijection:*

$$\begin{array}{rccc} \hat{\pi}: & \hat{\mathbb{R}}^n & \xrightarrow{\sim} & S^n \\ & x & \longmapsto & \hat{\pi}(x) := \begin{cases} \pi(x) & \text{if } x \neq \infty \\ e_{n+1} & \text{if } x = \infty \end{cases} \end{array} .$$

$\hat{\mathbb{R}}^n$ is called the one point compactification of $\mathbb{R}^n$.

Since S^n is compact, the map $\hat{\pi} : \hat{\mathbb{R}}^n \xrightarrow{\sim} S^n$ creates a bijection between $\hat{\mathbb{R}}^n$ and a compact space, which motivates the name 'compactification'.

The point at infinity has a peculiar nature: while it must be considered as an abstract point from the perspective of $\mathbb{R}^n$, it is a very concrete object from the perspective of $\mathbb{R}^{n+1}$, i.e. the north pole of S^n:

$$\hat{\pi}(\infty) = e_{n+1} \iff \infty = \hat{\pi}^{-1}(e_{n+1}). \tag{12.6}$$

Notice that the interplay between ∞ and $\hat{\pi}$ is a sort of chicken and egg situation: we need to assume the existence of ∞ as an abstract element to define the extended map $\hat{\pi}$ and, thanks to this latter, we give a concrete identification of ∞ as the inverse image of the north pole of S^n via $\hat{\pi}$! These circumstances seem appropriate to quote the famous von Neumann's sentence: '*In mathematics, you never understand things; you just get used to them*'.

Following von Neumann's hint, let us try to get used to the concept of one-point compatification by considering the cases $n = 1, 2$.

- The one-point compactification of $\mathbb{R}$ is

  $$\hat{\mathbb{R}} = \mathbb{R} \sqcup \{\infty\} \cong S^1 \underset{(10.8)}{=} \mathbb{RP}^1,$$

  i.e. either the unit circle in $\mathbb{R}^2$ or the real projective line. The point at infinity of the real line, interpreted as a hyperplane in $\mathbb{R}^2$, can be identified with e_2.

- The one-point compactification of $\mathbb{R}^2$ is

  $$\hat{\mathbb{R}}^2 = \mathbb{R}^2 \sqcup \{\infty\} \cong S^2 \underset{(10.8)}{=} \mathbb{RP}^2,$$

  i.e. either the unit sphere in $\mathbb{R}^3$ or the real projective plane. The point at infinity of the real plane $\mathbb{R}^2$, interpreted as a hyperplane in $\mathbb{R}^3$, can be identified with e_3.

The one-point compactification of $\mathbb{C}$ has a special name.

Def. 12.3.3 (Riemann sphere) *The one-point compactification of the complex plane $\mathbb{C}$ is called Riemann sphere*

$$\hat{\mathbb{C}} := \mathbb{C} \sqcup \{\infty\} \underset{(10.11)}{=} \mathbb{CP}^1.$$

Thanks to the bijection provided by $\hat{\pi}$, it is possible to endow $\hat{\mathbb{R}}^n$ with a metric.

Def. 12.3.4 (Chordal metric) *The chordal metric d_C on $\hat{\mathbb{R}}^n$ is:*

$$d_C(x,y) := |\hat{\pi}(x) - \hat{\pi}(y)|, \qquad \forall x, y \in \hat{\mathbb{R}}^n.$$

So, to compute the chordal metric, we stereographically project $x, y \in \hat{\mathbb{R}}^n$ on the sphere S^n and then we compute the Euclidean norm of the difference between the two projections, interpreted as points of $\mathbb{R}^{n+1}$. Of course, if $x = y = \infty$, $d_C(\infty, \infty) = |e_{n+1} - e_{n+1}| = 0$. The following result shows what are the values taken by the chordal metric in all the other cases.

Theorem 12.3.3 *Let $x, y \in \mathbb{R}^n$. Then, for all $x, y \in \mathbb{R}^n$*

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

2. $d_C(x,y) = \frac{2}{\sqrt{1+|x|^2}\sqrt{1+|y|^2}}|x-y|$.

Proof. First we remind that $\pi = \sigma_{e_{n+1},\sqrt{2}}\Big|_{\mathbb{R}^n}$ and we observe that, if $x \in \mathbb{R}^n$, then $\langle x, e_{n+1}\rangle = \sum_{i=1}^{n} x_i \cdot 0 + 0 \cdot 1 = 0$, so, by Carnot's theorem, $|x - e_{n+1}|^2 = |x|^2 - 2\cancelto{0}{\langle x, e_{n+1}\rangle} + \cancelto{1}{|e_{n+1}|^2} = 1 + |x|^2$.

1.

$$\begin{aligned} d_C(x,\infty) &= |\pi(x) - \pi(\infty)| = \left|\cancel{e_{n+1}} + \frac{2}{1+|x|^2}(x - e_{n+1}) - \cancel{e_{n+1}}\right| \\ &= \frac{2}{1+|x|^2}|x - e_{n+1}| = \frac{2}{1+|x|^2}\left(|x-e_{n+1}|^2\right)^{\frac{1}{2}} = \frac{2}{1+|x|^2}\sqrt{1+|x|^2} \\ &= \frac{2}{\sqrt{1+|x|^2}}. \end{aligned}$$

2. Since here $x, y \in \mathbb{R}^n$, we can write $d_C(x,y) = |\pi(x) - \pi(y)| = |\sigma_{e_{n+1},\sqrt{2}}(x) - \sigma_{e_{n+1},\sqrt{2}}(y)|$. Using property **3.** of theorem 12.2.2 we find

$$\begin{aligned} d_C(x,y) &= \frac{2|x-y|}{|x-e_{n+1}||y-e_{n+1}|} \\ &= \frac{2|x-y|}{\sqrt{1+|x|^2}\sqrt{1+|y|^2}}. \end{aligned}$$

□

Property **1.** says that the chordal distance between any point $x \in \mathbb{R}^n$ with the point at infinity is finite. Property **2.** of this last theorem implies that the chordal metric and the Euclidean one are multiple of each other by a positive constant that depends on the points x and y. This guarantees that the metric intuition that we have in Euclidean spaces can be transferred to $\hat{\mathbb{R}}^n$ for all the points different than ∞.

Corollary 12.3.1 *$f : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n$ is continuous in a point $x_0 \in \mathbb{R}^n$ w.r.t. the chordal metric if and only if f is continuous in x_0 w.r.t. the Euclidean metric.*

The following definitions formalize the quite intuitive extension of reflections and inversions to $\hat{\mathbb{R}}^n$ (for simplicity we keep the same symbols). In particular, note that the center of a sphere is mapped to the point at infinite by the corresponding inversion, and vice-versa.

Def. 12.3.5 *Let $\rho_{a,t}$ be a reflection and $\sigma_{a,r}$ an inversion in $\mathbb{R}^n$. The extension of $\rho_{a,t}$ in ∞ and of $\sigma_{a,r}$ in ∞ and a are defined as follows:*

$$\rho_{a,t}(\infty) := \infty \qquad \textit{and} \qquad \begin{cases} \sigma_{a,r}(\infty) := a \\ \sigma_{a,r}(a) := \infty \end{cases} .$$

The properties listed in theorems 12.2.1 and 12.2.2 are valid also for their extended versions. The great advantage of considering the point at infinity is that **both reflections and inversions become bijections on $\hat{\mathbb{R}}^n$**.

We also extend isometries and similarities to $\hat{\mathbb{R}}^n$ as follows.

Def. 12.3.6 *The sets of isometries and similarities on $\hat{\mathbb{R}}^n$ are:*

$$\mathcal{I}(\hat{\mathbb{R}}^n) := \{\phi : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n,\ \phi|_{\mathbb{R}^n} \text{ is an isometry and } \phi(\infty) = \infty\}$$

$$\mathcal{S}(\hat{\mathbb{R}}^n) := \{\phi : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n,\ \phi|_{\mathbb{R}^n} \text{ is a similarity and } \phi(\infty) = \infty\}.$$

The request $\phi(\infty) = \infty$ is fully justified by theorem 12.2.3 for isometries: they are compositions of reflections, which fix ∞. Similarities instead are compositions of reflections and inversions, so the request to fix ∞ does not seems well-motivated. In fact, as we will see in corollary 12.2.1 and theorem 12.4.2, the action on ∞ of the inversions involved in the creation of a similarity cancel out, remaining with a map that fixes ∞ also in that case.

The final information that we need before passing to the definition and analysis of Möbius transformations is the concept of sphere in the one point compactification of $\mathbb{R}^n$.

In the same way as we can identify the hyperplane $P(e_{n+1}, 0) \cong \mathbb{R}^n$ united with $\{\infty\}$ with the sphere S^n by means of $\hat{\pi}$, we can identify the union of a hyperplane with the point at the infinity with a sphere. This consideration justifies the following definition.

Def. 12.3.7 *A sphere Σ in $\hat{\mathbb{R}}^n$ is either a Euclidean sphere $S^{n-1}_{a,r}$ or the union of a hyperplane with the point at infinity $\hat{P}(a,t) := P(a,t) \cup \{\infty\}$.*

12.4 Möbius transformations in the Euclidean space

Möbius transformations arise from the combinations of inversions and reflections of $\hat{\mathbb{R}}^n$, one of the main interest in combining them is that, when they are fused together, they form a group. Notice that this is not a trivial statement because neither the set of reflections nor the set inversions form a group: we do not have a identity element or any stability. However, theorem 12.2.3 tells us that by combining reflections and inversions we can obtain the identity function and the group of similarities.

Def. 12.4.1 *A **Möbius transformation** $\phi : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n$ is a finite composition of reflections w.r.t. a hyperplane and inversions w.r.t. a sphere in $\hat{\mathbb{R}}^n$. The group of Möbius transformations is:*

$$\mathcal{M}(\hat{\mathbb{R}}^n) = \big\{\phi = \mu_1 \circ \cdots \circ \mu_m \ \ m \in \mathbb{N},\ \mu_i \text{ reflections or inversions of } \hat{\mathbb{R}}^n,\ i \in \{1, \ldots, m\}\big\}.$$

It can be verified that $\mathcal{M}(\hat{\mathbb{R}}^n)$ is a group under composition. We underline that, by definition, **a Möbius transformation is a bijection in** $\hat{\mathbb{R}}^n$.

Since every Euclidean isometry and similarity can be decomposed into a combination of reflections and inversions thanks to theorem 12.2.3 and corollary 12.2.1, we have the following chain of inclusions among groups:

$$\mathcal{I}(\hat{\mathbb{R}}^n) \subset \mathcal{S}(\hat{\mathbb{R}}^n) \subset \mathcal{M}(\hat{\mathbb{R}}^n).$$

We shall see that Möbius transformation can be equivalently characterized in three different ways:

1. they are the only transformation that preserve the cross-ratio (see below)
2. they are the only transformations that map spheres of $\hat{\mathbb{R}}^n$ into other spheres of $\hat{\mathbb{R}}^n$
3. when $n \geqslant 3$ and we restrict them to $\mathbb{R}^n$ they are the only $\mathscr{C}^1$ transformations of $\mathbb{R}^n$ that preserve angles, i.e they are the only $\mathscr{C}^1$ conformal maps of $\mathbb{R}^n$.

Let us start by analyzing the relationship between Möbius transformations and the cross ratio.

12.4.1 Möbius transformations and the cross ratio

The cross ratio is the fundamental invariant of Möbius transformations.

Def. 12.4.2 *Let $u, v, x, y \in \hat{\mathbb{R}}^n$ such that $u \neq y, v \neq x$. The cross-ratio of (u, v, x, y) is the (continuous) function:*

$$\begin{array}{rccl} [\cdot,\cdot,\cdot,\cdot]: & \hat{\mathbb{R}}^n \times \hat{\mathbb{R}}^n \times \hat{\mathbb{R}}^n \times \hat{\mathbb{R}}^n & \longrightarrow & [0, +\infty) \\ & (u, v, x, y) & \longmapsto & [u, v, x, y] := \frac{d_C(u,x) d_C(v,y)}{d_C(u,y) d_C(v,x)}. \end{array}$$

In the special case that u, v, x, y belong to $\mathbb{R}^n$, then, thanks to property **2.** of theorem 12.3.3, we can re-write their cross ratio as follows:

$$[u, v, x, y] = \frac{|u - x||v - y|}{|u - y||v - x|}. \tag{12.7}$$

We remark that if one of the four points, say u, is ∞, that point can simply be 'dropped out' of the computation, in the sense that the factor in which it appears can be simply set to 1. The reason underlying this rule relies on theorem 12.3.3, in fact

$$[\infty, v, x, y] = \frac{d_C(\infty, x) d_C(v, y)}{d_C(\infty, x) d_C(v, y)} = \frac{\frac{2}{\sqrt{1+|x|^2}} \frac{2|v-y|}{\sqrt{1+|v|^2}\sqrt{1+|y|^2}}}{\frac{2}{\sqrt{1+|y|^2}} \frac{2|v-x|}{\sqrt{1+|v|^2}\sqrt{1+|x|^2}}} = \frac{|v - y|}{|v - x|}$$

and similarly if ∞ has any other place in the cross ratio. From now on, we will use the following formulae as definitions of cross ratio when ∞ is one of the points involved in its computation:

$$\begin{cases} [\infty, v, x, y] = \frac{|v-y|}{|v-x|} \\ [u, \infty, x, y] = \frac{|u-x|}{|u-y|} \\ [u, v, \infty, y] = \frac{|v-y|}{|u-y|} \\ [u, v, x, \infty] = \frac{|u-x|}{|v-x|} \end{cases}. \tag{12.8}$$

It must be stressed that there are several definitions of cross ratio in the literature, most of the time with little consequences since we can switch u, v, x, y in the cross-ratio in many different ways without changing the overall result. In particular, the definition that we gave is different than the one given by Ratcliffe in [**?**]. We chose the definition above because it will be the handier when we will deal with the conformal hyperbolic model.

Theorem 12.4.1 *A map $\phi : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n$ is a Möbius transformation if and only if ϕ preserves the cross ratio.*

Proof.

$\boxed{\implies}$: suppose ϕ is a Möbius transformation, then it is enough to show that any generic inversion and reflection preserves the cross ratio since, by definition, every Möbius transformation is a combination of inversions and reflections and thus the composition of cross ratio preserving functions will be overall cross ratio preserving.

First of all we suppose that the values taken by ϕ are finite, we will deal with the ∞ later. In this case, we can use formula (12.7) to compute the cross ratio.

If ϕ is a reflection $\rho_{a,t}$, then the cross ratio is preserved by the fact that reflections are Euclidean isometries.

This argument cannot be used if ϕ is an inversion $\sigma_{a,r}$, because inversions are not isometries. If we remove a form the possible values that the points u, v, x, y can take, then, by property **3.** of theorem 12.2.2, i.e.

$$|\sigma_{a,r}(x) - \sigma_{a,r}(y)| = \frac{r^2|x-y|}{|x-a||y-a|},$$

we have:

$$\begin{aligned}
[\sigma_{a,r}(u), \sigma_{a,r}(v), \sigma_{a,r}(x), \sigma_{a,r}(y)] &= \frac{|\sigma_{a,r}(u) - \sigma_{a,r}(x)||\sigma_{a,r}(v) - \sigma_{a,r}(y)|}{|\sigma_{a,r}(u) - \sigma_{a,r}(y)||\sigma_{a,r}(v) - \sigma_{a,r}(x)|} \\
&= \frac{r^2|u-x|}{r^2|u-y|}\frac{r^2|v-y|}{r^2|v-x|}\underbrace{\frac{|x-a||y-a||u-a||v-a|}{|x-a||y-a||u-a||v-a|}}_{=1} \\
&= [u, v, x, y].
\end{aligned}$$

Suppose now that[2] $\phi(u) = \infty$. By definition 12.3.5, if $\phi = \rho_{a,t}$, this can happen only if $u = \infty$ since $\rho_{a,t}(\infty) = \infty$, i.e. we must prove that

$$[\infty, \rho_{a,t}(v), \rho_{a,t}(x), \rho_{a,t}(y)] = [\infty, v, x, y] \underset{(12.8)}{=} \frac{|v-y|}{|v-x|},$$

which is very simple:

$$[\infty, \rho_{a,t}(v), \rho_{a,t}(x), \rho_{a,t}(y)] \underset{(12.8)}{=} \frac{|\rho_{a,t}(v) - \rho_{a,t}(y)|}{|\rho_{a,t}(v) - \rho_{a,t}(x)|} \underset{\rho_{a,t} \text{ isometry}}{=} \frac{|v-y|}{|v-x|} = [\infty, v, x, y].$$

Instead, if $\phi = \sigma_{a,r}$, then we know that $\phi(u) = \infty$ only if $u = a$. So, we must prove that

$$[\infty, \sigma_{a,r}(v), \sigma_{a,r}(x), \sigma_{a,r}(y)] = [a, v, x, y],$$

[2] This argument can be extended verbatim to v, x, y, so we will consider only u.

on the left-hand side we have

$$[\infty, \sigma_{a,r}(v), \sigma_{a,r}(x), \sigma_{a,r}(y)] \underset{(12.8)}{=} \frac{|\sigma_{a,r}(v) - \sigma_{a,r}(y)|}{|\sigma_{a,r}(v) - \sigma_{a,r}(x)|} \underset{(\mathbf{3.}\text{ of th. } 12.2.2)}{=} \frac{r^2 \frac{|v-y|}{|v-a||y-a|}}{r^2 \frac{|v-x|}{|v-a||x-a|}} = \frac{|v-y||x-a|}{|y-a||v-x|},$$

on the right-hand side we have

$$[a, v, x, y] = \frac{|v-y||x-a|}{|y-a||v-x|},$$

which verifies the preservation of the cross ratio also in this case. To resume, all Möbius transformations preserve the cross ratio.

$\boxed{\Longleftarrow}$: conversely, we assume that ϕ preserves the cross ratio. We analyze first the case when ϕ fixes ∞, i.e. $\phi(\infty) = \infty$, we will deal with the other option later. Let $u, v, x, y \in \mathbb{R}^n$ such that $u \neq y,\ v \neq x$ and $(u, v) \neq (x, y)$. If $u \neq x$, then

$$\begin{aligned} i)\ [\phi(u), \infty, \phi(x), \phi(y)] = [u, \infty, x, y] \quad &\underset{(12.8)}{\Longleftrightarrow} \quad \frac{|\phi(u) - \phi(x)|}{|\phi(u) - \phi(y)|} = \frac{|u-x|}{|u-y|} \\ &\Longleftrightarrow \quad \frac{|\phi(u) - \phi(x)|}{|u-x|} = \frac{|\phi(u) - \phi(y)|}{|u-y|} \end{aligned}$$

$$\begin{aligned} ii)\ [\phi(u), \phi(v), \phi(x), \infty] = [u, v, x, \infty] \quad &\underset{(12.8)}{\Longleftrightarrow} \quad \frac{|\phi(u) - \phi(x)|}{|\phi(v) - \phi(x)|} = \frac{|u-x|}{|v-x|} \\ &\Longleftrightarrow \quad \frac{|\phi(u) - \phi(x)|}{|u-x|} = \frac{|\phi(v) - \phi(x)|}{|v-x|}. \end{aligned}$$

Similarly, if $v \neq y$,

$$\begin{aligned} i)\ [\infty, \phi(v), \phi(x), \phi(y)] = [\infty, v, x, y] \quad &\underset{(12.8)}{\Longleftrightarrow} \quad \frac{|\phi(v) - \phi(y)|}{|\phi(v) - \phi(x)|} = \frac{|v-y|}{|v-x|} \\ &\Longleftrightarrow \quad \frac{|\phi(v) - \phi(y)|}{|v-y|} = \frac{|\phi(v) - \phi(x)|}{|v-x|} \end{aligned}$$

$$\begin{aligned} ii)\ [\phi(u), \phi(v), \infty, \phi(y)] = [u, v, \infty, y] \quad &\underset{(12.8)}{\Longleftrightarrow} \quad \frac{|\phi(v) - \phi(y)|}{|\phi(u) - \phi(y)|} = \frac{|v-y|}{|u-y|} \\ &\Longleftrightarrow \quad \frac{|\phi(v) - \phi(y)|}{|v-y|} = \frac{|\phi(u) - \phi(y)|}{|u-y|}. \end{aligned}$$

Hence, by combining *i)* and *ii)* in both cases we obtain that, for all $u, v, x, y \in \mathbb{R}^n$ such that $u \neq y$ and $v \neq x$,

$$\frac{|\phi(u) - \phi(y)|}{|u-y|} = \frac{|\phi(v) - \phi(x)|}{|v-x|},$$

if we set $k = |\phi(v) - \phi(x)|/|v - x|$, then $k > 0$ and it does not depend on u and y, which are two generic distinct elements of $\mathbb{R}^n$, so that we can write $|\phi(u) - \phi(y)| = k|u - y|$, which shows that ϕ is a Euclidean similarity and, hence a Möbius transformation.

Finally, if $a \neq \infty$ and $\phi(\infty) = a$, then we can combine ϕ with any inversion of the type $\sigma_{a,r}$, $r > 0$, obtaining $(\sigma_{a,r} \circ \phi)(\infty) = \infty$. Using the result obtained above, we have that $\sigma_{a,r} \circ \phi$ is a Möbius transformations, and so ϕ is also Möbius transformation by definition. □

The following result gives important stuctural information about Möbius transformations.

Theorem 12.4.2 *Let $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$. Then:*

1. *$\phi(\infty) = \infty$ if and only if ϕ is a similarity of $\hat{\mathbb{R}}^n$*
2. *if $\phi(\infty) \neq \infty$, then, there exist:*
   - *a unique sphere Σ in $\mathbb{R}^n$ on which ϕ acts as a Euclidean isometry, i.e. for all $x, y \in \Sigma$, $|\phi(x) - \phi(y)| = |x - y|$*
   - *a unique inversion σ w.r.t. Σ and a unique Euclidean isometry $\psi \in \mathcal{I}(\hat{\mathbb{R}}^n)$ such that ϕ can be decomposed as follows $\phi = \psi \circ \sigma$.*

Proof.

1. : the previous theorem implies directly that if $\phi(\infty) = \infty$ and ϕ is a Möbius transformation, then ϕ is a similarity on $\hat{\mathbb{R}}^n$. Vice-versa, a similarity on $\mathbb{R}^n$ is a Möbius transformation on $\mathbb{R}^n$; now, thanks to the proof of corollary 12.2.1, every similarity is the composition of at most $n + 1$ reflections and two inversions w.r.t. the same center. This implies that, the only possible extension of ϕ to the point at infinity is the one that fixes ∞, in fact, reflections fix ∞ and also the composition of the two inversions will globally leave ∞ fixed. As previously said, this argument provides a full justification of the definition given in 12.3.6.

2. : first we prove the existence of the decomposition $\phi = \psi \circ \sigma$ and then its uniqueness. Notice that this automatically implies that we also have to exhibit the sphere Σ w.r.t. the inversion σ is defined.

Existence: since ϕ is a Möbius transformation that modifies the point at infinity, it is natural to set the center of the sphere Σ that we are looking for as $a := \phi^{-1}(\infty)$. Regarding the ray of the sphere, let us preliminarly set it to 1, i.e. let us consider the sphere $S^{n-1}_{a,1}$ and the inversion $\bar{\sigma}$ w.r.t. to it.

Clearly, $\phi \circ \bar{\sigma}$ fixes ∞, in fact $(\phi \circ \bar{\sigma})(\infty) = \phi(\bar{\sigma}(\infty)) = \phi(a) = \phi(\phi^{-1}(\infty)) = \infty$. So $\phi \circ \bar{\sigma}$ is a Euclidean similarity thanks to point 1. Hence, it exists $k > 0$ such that, for all $x, y \in \mathbb{R}^n$, we have

$$|(\phi \circ \bar{\sigma})(x) - (\phi \circ \bar{\sigma})(y)| = k|x - y|. \tag{12.9}$$

Furthermore, $\bar{\sigma}$ is an inversion and so it is also an involution, $\bar{\sigma}^2 = id_{\mathbb{R}^n}$, so

$$\begin{aligned} |\phi(x) - \phi(y)| &= |(\phi \circ id_{\mathbb{R}^n})(x) - (\phi \circ id_{\mathbb{R}^n})(y)| = |(\phi \circ \bar{\sigma}^2)(x) - (\phi \circ \bar{\sigma}^2)(y)| \\ &= |(\phi \circ \bar{\sigma})(\bar{\sigma}(x)) - (\phi \circ \bar{\sigma})(\bar{\sigma}(y))| \\ &\underset{(12.9)}{=} k|\bar{\sigma}(x) - \bar{\sigma}(y)| \underset{\mathbf{3.}\text{ of th.12.2.2}}{=} k\frac{|x-y|}{|x-a||y-a|}. \end{aligned} \tag{12.10}$$

Observe now that $x, y \in S^n_{a,r}$, then $|x - a| = |y - a| = r$, so, if we set $r := \sqrt{k}$, then

$$|\phi(x) - \phi(y)| = k\frac{|x-y|}{|x-a||y-a|} = k\frac{|x-y|}{k} = |x - y|,$$

and so ϕ is a Euclidean isometry on $\Sigma := S^{n-1}_{a,r}$ if and only if $r = \sqrt{k}$ is the radius of Σ, which implies its uniqueness. Finally, we can conclude by setting $\sigma = \sigma_{a,r}$ and $\psi = \phi \circ \sigma$. The following computations show that $\psi \in \mathcal{I}(\hat{\mathbb{R}}^n)$. Clearly $\psi(\infty) = \phi(\sigma(\infty)) = \phi(a) = \infty$. Moreover, if $x, y \in \mathbb{R}^n$, then the following chain of equalities holds

$$
\begin{aligned}
|\psi(x) - \psi(y)| = |\phi(\sigma(x)) - \phi(\sigma(y))| &\underset{(12.10)}{=} k\frac{|\sigma(x) - \sigma(y)|}{|\sigma(x) - a||\sigma(y) - a|} \\
&\underset{\textbf{3. } \text{of th.12.2.2}}{=} k\frac{k\frac{|x-y|}{|x-a||y-a|}}{|\sigma(x) - a||\sigma(y) - a|} = k^2\frac{|x-y|}{|x-a||y-a|}\frac{1}{|\sigma(x) - a||\sigma(y) - a|} \\
&\underset{(12.2)}{=} k^2\frac{|x-y|}{|x-a||y-a|}\frac{|x-a||y-a|}{k^2} = |x-y|.
\end{aligned}
$$

Uniqueness: suppose that we also have $\phi = \psi_0 \circ \sigma_0$ with ψ_0 a Euclidean isometry and σ_0 an inversion w.r.t. a sphere $S^{n-1}_{a_0,r_0}$. We start by proving that Σ and Σ_0 share their center: the decomposition $\phi = \psi \circ \sigma$ gives $\phi(a) = \infty$, while the decomposition $\phi = \psi_0 \circ \sigma_0$ gives $\phi(a_0) = \infty$, so $a = \phi^{-1}(\infty) = a_0$.

As proven above, ϕ is an isometry on both Σ and Σ_0 if and only if their radius has a specific, fixed, value, thus not only Σ and Σ_0 are concentric, but they also share their radius, i.e. $\Sigma = \Sigma_0$. This implies that also σ and σ_0 coincide, which, in turn, implies that $\psi = \psi_0$.

This shows that uniqueness of Σ and of the decomposition of ϕ.

Finally, let us ask if there exists another sphere S^{n-1} on which ϕ acts isometrically. For sure, this sphere S^{n-1} must have a center different than a, otherwise, as we have just proven, we fall back to the previous sphere Σ. Thus, let us suppose that $S^{n-1} = S^{n-1}_{b,s}$, with $b \in \mathbb{R}^n$, $b \neq a$, and $s > 0$. The idea to prove the uniqueness of Σ is to show that there exist two points $\bar{x}, \bar{y} \in S^{n-1}_{b,s}$ such that $|\phi(\bar{x}) - \phi(\bar{y})| \neq |\bar{x} - \bar{y}|$, and so ϕ does not act as an isometry on $S_{b,s}$.

To this aim, we observe that it exists $\alpha > 0$, $\alpha \neq 1$, such that $S^{n-1}_{a,\alpha r}$ and $S^{n-1}_{b,s}$ intersect in two points, that will constitute the two points $\bar{x}$ and $\bar{y}$ that we are searching for, in fact, recalling that $r^2 = k$ we have, by eq. (12.10) :

$$
|\phi(\bar{x}) - \phi(\bar{y})| = \frac{r^2|\bar{x} - \bar{y}|}{|\bar{x} - a||\bar{y} - a|} = \frac{r^2}{\alpha^2 r^2}|\bar{x} - \bar{y}| \neq |\bar{x} - \bar{y}|,
$$

hence ϕ cannot be an isometry on $S^{n-1}_{b,s}$. □

Remark that the same arguments used in the proof above can be used to assure it does not exist any hyperplane $P(a, t)$ on which ϕ acts as an isometry. Hence, Σ is the not only the unique sphere in $\mathbb{R}^n$ on which ϕ is isometric, but also on $\hat{\mathbb{R}}^n$. For this reason, the following definition is completely justified.

Def. 12.4.3 *For any Möbius transformation $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$ such that $\phi(\infty) \neq \infty$, the unique sphere Σ on which ϕ acts as an isometry is called the **isometric sphere** of ϕ.*

We now arrive to the analogous of corollary 12.2.1 for Möbius transformations.

Corollary 12.4.1 *Every Möbius transformation on $\hat{\mathbb{R}}^n$ is at most the composition of $n + 3$ reflections or inversions.*

Proof. Let $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$. If $\phi(\infty) = \infty$ then ϕ is a similarity and so it can be decomposed into $n+3$ reflections or inversions by corollary 12.2.1. Instead, if $\phi(\infty) \neq \infty$ then, by theorem 12.4.2, we have the following decomposition $\phi = \psi \circ \sigma$ with $\psi \in \mathcal{I}(\mathbb{R}^n)$ and σ an inversion. Since every isometry is the product of at most $n+1$ reflections, ϕ is the composition of at most $(n+2)$ reflections or inversions. □

So, the upper bound of the number of reflections or inversions that make up a Möbius transformation in $\hat{\mathbb{R}}^n$ is always 3 units bigger than n: this is a quite fine information that was not at all intuitive from the original definition of Möbius transformations.

A very important consequence of what we have just proven is the possibility to connect the Möbius transformations geometrically defined as compositions of reflections and inversions, with the analytical formula used in $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{C}$ that makes use of fractional linear transformations.

Corollary 12.4.2 *The Möbius transformations on $\mathbb{R}$, $\mathbb{R}^2$ or H are fractional linear transformations, i.e.*

$$\phi(x) = \frac{ax+b}{cx+d}, \quad x \in R$$

for $\mathbb{R}^2$ and $\mathbb{C}$ see ch 8 section 8.5...

Proof. In progress... □

12.4.2 The action of Möbius transformations on the set spheres in $\hat{\mathbb{R}}^n$

In the previous subsection we have seen how Möbius transformations and spheres of $\hat{\mathbb{R}}^n$ are linked. We now show a very powerful result: Möbius transformations acts transitively on the set of spheres of $\hat{\mathbb{R}}^n$.

We start by proving the *stability of the set of spheres in $\hat{\mathbb{R}}^n$ w.r.t. Möbius transformations.* We will do this by using several results that we underline in separated lemmas because of their stand-alone interest.

Lemma 12.4.1 *The following assertions hold.*

1. *The set of hyperplanes and that of Euclidean spheres in $\mathbb{R}^n$ is stable w.r.t. isometries and similarities in $\mathbb{R}^n$, i.e. they map hyperplanes into hyperplanes and Euclidean spheres into Euclidean spheres.*

2. *The group of Euclidean isometries $\mathcal{I}(\mathbb{R}^n)$ acts transitively on the set of hyperplanes in $\mathbb{R}^n$ and the group of Euclidean similarities $\mathcal{S}(\mathbb{R}^n)$ acts transitively on the set of spheres in $\mathbb{R}^n$.*

Since the group of Euclidean similarities contains the group of isometries, it follows that $\mathcal{S}(\mathbb{R}^n)$ acts transitively also on the set of hyperplanes of $\mathbb{R}^n$.

Proof.

1. : since isometries are similarities, we will prove this result directly on the set of similarities. From theorem 11.1.1 we know that $f \in \mathcal{S}(\mathbb{R}^n)$ if and only if it can be expressed in the form

$$f(x) = b + k\phi(x) \iff \phi(x) = \frac{f(x)-b}{k}, \qquad \forall x \in \mathbb{R}^n, \tag{12.11}$$

with $b \in \mathbb{R}^n$, $k > 0$ and $\phi \in \mathrm{O}(n)$. If $k = 1$, $f \in \mathcal{I}(\mathbb{R}^n)$.

Hyperplanes: given a hyperplane $P(a,t) = \{x \in \mathbb{R}^n \ : \ \langle x, a\rangle = t\}$, with $a \in \mathbb{R}^n$ such that $|a| = 1$ and $t \geqslant 0$, our aim is to show that $f(P(a,t))$ is a hyperplane.

We recall that, since ϕ belongs to $\mathrm{O}(n)$, it is linear, invertible and $\langle \phi(x), \phi(y)\rangle = \langle x, y\rangle$, $\forall x, y \in \mathbb{R}^n$. Hence $\forall x \in P(a,t)$ the following chain of equalities holds:

$$\begin{aligned}\langle \phi(x), \phi(a)\rangle = \langle x, a\rangle = t \iff \langle k\phi(x), k\phi(a)\rangle = k^2 t &\underset{(12.11)}{\iff} \langle f(x) - b, k\phi(a)\rangle = k^2 t \\ &\iff \langle f(x), k\phi(a)\rangle - \langle b, k\phi(a)\rangle = k^2 t \\ &\iff \langle f(x), k\phi(a)\rangle = k^2 t + \langle b, k\phi(a)\rangle \\ &\iff \langle f(x), \frac{k\phi(a)}{|k\phi(a)|}\rangle = \frac{k^2 t + \langle b, k\phi(a)\rangle}{|k\phi(a)|},\end{aligned}$$

note that we are allowed to divide by $k|\phi(a)|$ because $k > 0$ and $a \neq 0$, so $|\phi(a)| = |a| \neq 0$. Moreover $|k\phi(a)| = k|\phi(a)| = k|a| = k$, hence $\langle f(x), \phi(a)\rangle = kt + \langle b, \phi(a)\rangle$ $\forall x \in P(a,t)$.

This means that $f(x) \in P(\phi(a), kt + \langle b, \phi(a)\rangle)$, which is well defined as a hyperplane because $|\phi(a)| = 1$. As noticed in the definition of hyperplane, the positivity of $kt + \langle b, \phi(a)\rangle$ is not an issue because its possible negative sign we can integrated in the vector $\phi(a)$ without changing its norm. To avoid a cumbersome notation, we consider this as implicitly performed.

This proves that

$$\boxed{f(P(a,t)) = P(\phi(a), kt + \langle b, \phi(a)\rangle)} \qquad \forall f \in \mathcal{S}(\mathbb{R}^n), \tag{12.12}$$

i.e. the image of a hyperplane orthogonal to a and distant t from 0 is still a hyperplane orthogonal to $\phi(a)$ and distant $\langle b, \phi(a)\rangle$ from 0.

Euclidean spheres: Let $S^{n-1}_{a,r} = \{x \in \mathbb{R}^n \ : \ |x - a| = r\}$, with $a \in \mathbb{R}^n$ and $r > 0$. We want to show that $f(S^{n-1}_{a,r})$ is a sphere in $\mathbb{R}^n$. Let x be a point of $S^{n-1}_{a,r}$, then

$$r^2 = |x-a|^2 = \langle x-a, x-a\rangle = \langle \phi(x-a), \phi(x-a)\rangle = \langle \phi(x)-\phi(a), \phi(x)-\phi(a)\rangle = |\phi(x)-\phi(a)|^2.$$

Multiplying both sides by k^2 and using eq. (12.11) we obtain:

$$k^2 r^2 = |k\phi(x) - k\phi(a)|^2 = |f(x) - b - k\phi(a)|^2 = |f(x) - (b + k\phi(a))|^2 = |f(x) - f(a)|^2,$$

so, if $x \in S^{n-1}_{a,r}$, then $|f(x) - f(a)| = kr$, i.e.

$$\boxed{f(S^{n-1}_{a,r}) = S^{n-1}_{f(a),kr}} \qquad \forall f \in \mathcal{S}(\mathbb{R}^n), \tag{12.13}$$

i.e. the image of a Euclidean sphere of center a and radius r is still a Euclidean sphere of center $f(a)$ and radius kr.

2. : we first prove the thesis for hyperplanes and then for spheres.

Hyperplanes: let us fix two hyperplanes $P(a,t)$ and $P(a',t')$, with $a, a' \in \mathbb{R}^n$ such that $|a| = |a'| = 1$ and $t, t' \geqslant 0$. Our aim is to prove that it always exists an isometry $f \in \mathcal{I}(\mathbb{R}^n)$ such that $f(P(a,t)) = P(a',t')$. Thanks to eq. (12.12) with $k = 1$ because f is an isometry, we can rewrite the last equation as $P(\phi(a), t + \langle b, \phi(a)\rangle) = P(a', t')$ so that our problem is equivalent to showing that there exist $\phi \in \mathrm{O}(n)$ and $b \in \mathbb{R}^n$ such that the system

$$\begin{cases} \phi(a) = a' \\ t + \langle b, \phi(a)\rangle = t' \end{cases}$$

has at least one solution for all vectors $a, a' \in \mathbb{R}$ such that $|a| = |a'| = 1$, i.e. $a, a' \in S^{n-1}$. Thanks to the transitivity of $\mathrm{O}(n)$ on S^{n-1}, it surely exists $\phi \in \mathrm{O}(n)$ such that $a' = \phi(a)$, if we introduce this in the second equation we get $t + \langle b, a'\rangle = t'$, or $\langle b, a'\rangle = t' - t$, which is satisfied by all vectors $b \in \mathbb{R}^n$ such that $b \in P(a'\mathrm{signum}(t' - t), |t' - t|)$.

Euclidean spheres: analogously to the previous case, once fixed any two spheres $S^{n-1}_{a,r}$ and $S^{n-1}_{a',r'}$, $a, a' \in \mathbb{R}^n$ and $r, r' > 0$, we must prove that it exists a similarity $f \in \mathcal{S}(\mathbb{R}^n)$ such that $f(S^{n-1}_{a,r}) = S^{n-1}_{a',r'}$. Thanks to eq. (12.13), we can rewrite the last equation as $S^{n-1}_{f(a),kr} = S^{n-1}_{a',r'}$, or $S^{n-1}_{b+k\phi(a),kr} = S^{n-1}_{a',r'}$, and so our problem is equivalent to the existence of $\phi \in \mathrm{O}(n)$, $k > 0$ and $b \in \mathbb{R}^n$ such that the system

$$\begin{cases} b + k\phi(a) = a' \\ kr = r' > 0 \end{cases}$$

has at least one solution for all $a, a' \in \mathbb{R}^n$ and $r, r' > 0$. We have immediately that $k = r'/r > 0$, which implies $b + \frac{r'}{r}\phi(a) = a'$. If we set $\phi = id_{\mathbb{R}^n} \in \mathrm{O}(n)$, then we get $b + \frac{r'}{r}a = a'$, which leads to $b = a' - \frac{r'}{r}a$. So, in conclusion, $k = r'/r$, $\phi = id_{\mathbb{R}^n}$ and $b = a' - \frac{r'}{r}a$ solve the system above, thus implying the transitivity of $\mathcal{S}(\mathbb{R}^n)$ on the set of Euclidean spheres in $\mathbb{R}^n$. $\square$

Note that if we consider similarities in $\hat{\mathbb{R}}^n$, since they fix ∞, this lemma states that similarities in $\hat{\mathbb{R}}^n$ map hyperplanes $\cup \{\infty\}$ into hyperplanes $\cup \{\infty\}$ and Euclidean spheres into Euclidean spheres. Hence, a weaker, but useful, reformulation of this assertion is contained in the following corollary.

Corollary 12.4.3 *The set of spheres in $\hat{\mathbb{R}}^n$ is stable w.r.t. isometries and similarities in $\hat{\mathbb{R}}^n$.*

As a consequence, reflections, which are particular types of isometries, are stable on the set of spheres in $\hat{\mathbb{R}}^n$.

Lemma 12.4.2 *Let $a \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$ satisfying $\alpha\beta < |a|^2$. Then, the set of points defined by*

$$\Sigma_{\alpha,\beta} := \{x \in \mathbb{R}^n \ : \ \alpha|x|^2 + 2\langle x, a\rangle + \beta = 0\} \tag{12.14}$$

represents either a hyperplane or a sphere in $\mathbb{R}^n$.

Proof. First of all we note that if $a = 0$ we cannot hope to find the equation of a $(n-1)$-dimensional hyperplane in $\mathbb{R}^n$ simply because the orthogonal complement of the null vector of $\mathbb{R}^n$ is $\mathbb{R}^n$ itself. However, it is clear that when $a = 0$, eq. (12.14) represents the equation of a sphere centered in 0 and with radius $r = \sqrt{-\beta/\alpha}$. So, for $a = 0$, eq. (12.14) represents all the spheres centered in 0 and with arbitrary (strictly positive) radius provided that $\alpha \neq 0$ and $\alpha\beta < 0$.

Let us now consider the case $a \in \mathbb{R}^n \backslash \{0\}$. If $\alpha = 0$, eq. (12.14) represents the hyperplane $P(\frac{a}{|a|}, -\frac{\beta}{2|a|})$.

If $\alpha \neq 0$, eq. (12.14) represents (under a constraint that we will determine below) the sphere $S^{n-1}_{-\frac{a}{\alpha}, \frac{\sqrt{|a|^2-\alpha\beta}}{|\alpha|}}$. In fact, by dividing both sides of eq. (12.14) by $-\alpha$ we get

$$-|x|^2 - \frac{2}{\alpha}\langle x, a\rangle - \frac{\beta}{\alpha} = 0 \iff -|x|^2 + 2\left\langle x, -\frac{a}{\alpha}\right\rangle - \frac{\beta}{\alpha} = 0 \iff |x|^2 - 2\left\langle x, -\frac{a}{\alpha}\right\rangle + \frac{\beta}{\alpha} = 0,$$

which coincides with the equation of a sphere of radius r and center c, i.e.

$$|x-c|^2 = r^2 \iff |x|^2 - 2\langle x, c\rangle + |c|^2 - r^2 = 0,$$

if and only if the center is $c = -\frac{a}{\alpha}$ and the radius r satisfies $\frac{\beta}{\alpha} = \frac{|a|^2}{\alpha^2} - r^2$, i.e. $r^2 = \frac{|a|^2}{\alpha^2} - \frac{\beta}{\alpha}$, or $r = \frac{\sqrt{|a|^2-\alpha\beta}}{|\alpha|}$. Thus, the constraint that allows eq. (12.14) to represent any sphere of arbitrary radius centered in $a \neq 0$ is $\alpha\beta < |a|^2$.

We can resume our analysis by saying that eq. (12.14) represents any sphere or hyperplane in $\mathbb{R}^n$ provided that $\alpha\beta < |a|^2$ for all $a \in \mathbb{R}^n$, as it was to be proven. □

The next lemma shows that also inversions are stable on the set of spheres in $\hat{\mathbb{R}}^n$.

Lemma 12.4.3 *The set of spheres in $\hat{\mathbb{R}}^n$ is stable w.r.t. inversions.*

Proof. Thanks to lemma 12.2.1, we can write any inversion $\sigma = \sigma_{a,r}$ as $\sigma = \phi \circ \sigma_{0,1} \circ \phi^{-1}$, with $\phi(x) = a + rx$ for all $x \in \mathbb{R}^n$. Thanks to corollary 11.1.1, ϕ and ϕ^{-1} are similarities, which are stable on the set of spheres in $\hat{\mathbb{R}}^n$ thanks to lemma 12.4.1. Hence we can reduce the proof to the case of $\sigma = \sigma_{0,1}$, i.e. from now on we will consider $\sigma(x) = \frac{x}{|x|^2}$, $x \neq 0$, and what we have to prove is that if we apply σ to either a sphere or a hyperplane we get back another sphere or hyperplane.

Since eq. (12.14) represents all possible hyperplane or sphere in $\mathbb{R}^n$ provided that $\alpha\beta < |a|^2$, to finish the proof of the theorem it is enough to show that σ preserves the structure of that equation. This turns out to be very easy: let x satisfy eq. (12.14), i.e. $\alpha|x|^2 + 2\langle x, a\rangle + \beta = 0$, which is equivalent to

$$\alpha|x|^2 + 2\langle x, a\rangle + \beta = 0 \iff \alpha + 2\left\langle \frac{x}{|x|^2}, a\right\rangle + \frac{\beta}{|x|^2} = 0 \iff \alpha + 2\langle \sigma(x), a\rangle + \frac{\beta}{|x|^2} = 0,$$

but $|\sigma(x)|^2 = \left|\frac{x}{|x|^2}\right|^2 = \frac{|x|^2}{|x|^4} = \frac{1}{|x|^2}$, so $\frac{\beta}{|x|^2} = \beta|\sigma(x)|^2$ and so we obtain that

$$\alpha|x|^2 + 2\langle x, a\rangle + \beta = 0 \iff \beta|\sigma(x)|^2 + 2\langle \sigma(x), a\rangle + \alpha = 0,$$

which shows that $\sigma(x)$ satisfies an equation of the same form as the one satisfied by x with the same constraint $\alpha\beta < |a|^2$. □

Theorem 12.4.3 *The set of spheres in $\hat{\mathbb{R}}^n$ is stable w.r.t. Möbius transformations on $\hat{\mathbb{R}}^n$.*

Proof. The proof will be just a sequence of considerations based on results that we have already proven that will allow us to greatly simplify the rest of the proof.

First of all, if $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$ fixes ∞, then, by theorem 12.4.2, ϕ is a similarity on $\hat{\mathbb{R}}^n$ and so it is stable on the set of spheres in $\hat{\mathbb{R}}^n$ by lemma 12.4.1.

If $\phi(\infty) \neq \infty$, then, again thanks to theorem 12.4.2, we can decompose $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$ as $\phi = \psi \circ \sigma$, where ψ is a Euclidean isometry and σ is an inversion w.r.t. a sphere. Again by lemma 12.4.1, ψ will be stable on the set of spheres in $\hat{\mathbb{R}}^n$, so what is left to prove is just that an inversion σ is stable on the set of spheres in $\hat{\mathbb{R}}^n$, which is guaranteed by lemma 12.4.3. $\square$

Thanks to this theorem, the natural action of the group of Möbius transformations on the set of spheres in $\hat{\mathbb{R}}^n$ defined by

$$\begin{array}{ccc} \mathcal{M}(\hat{\mathbb{R}}^n) \times \text{Spheres in } \hat{\mathbb{R}}^n & \longrightarrow & \text{Spheres in } \hat{\mathbb{R}}^n \\ (\phi, \Sigma) & \longmapsto & \phi(\Sigma) \end{array}$$

is well-defined.

Theorem 12.4.4 *The action of $\mathcal{M}(\hat{\mathbb{R}}^n)$ on the set of spheres of $\hat{\mathbb{R}}^n$ is transitive.*

Proof. Property 2. of Lemma 12.4.1 says that the group of similarities $\mathcal{S}(\hat{\mathbb{R}}^n) \subset \mathcal{M}(\hat{\mathbb{R}}^n)$ acts transitively on the set of hyperplanes united with $\{\infty\}$ of $\hat{\mathbb{R}}^n$ and on the set of Euclidean spheres in the following sense: for every fixed couple of spheres in $\hat{\mathbb{R}}^n$, Σ_1 and Σ_2 (*both* hyperplanes united with $\{\infty\}$ or *both* Euclidean spheres), there exists a similarity $\psi \in \mathcal{S}(\hat{\mathbb{R}}^n)$ such that $\psi(\Sigma_1) = \Sigma_2$.

However, the set of similarities of $\hat{\mathbb{R}}^n$ is not transitive on the whole set of generalized spheres in $\hat{\mathbb{R}}^n$. Indeed, it is not possible to map a hyperplane united with $\{\infty\}$ into a Euclidean sphere, or vice-versa, through a similarity. A simple explanation of this fact is that, clearly, ∞ belongs to the first category of sphere in $\hat{\mathbb{R}}^n$ but not to the second. Moreover, property 1. of theorem 12.4.2 says that similarities in $\hat{\mathbb{R}}^n$ leave the point ∞ fixed, thus they cannot map an object containing ∞ into an object not containing it, or vice-versa.

So, to conclude the proof, we must show that if we have two spheres Σ_1, Σ_2 in $\hat{\mathbb{R}}^n$, such that $\Sigma_1 = P(a,t) \cup \{\infty\}$ and $\Sigma_2 = S^{n-1}_{b,r}$, there exists a Möbius transformation ϕ, necessarily in $\mathcal{M}(\hat{\mathbb{R}}^n) \backslash \mathcal{S}(\hat{\mathbb{R}}^n)$ for what we have just observed, such that $\phi(\Sigma_1) = \Sigma_2$.

By a straightforward computation, it can be verified that $\sigma_{0,1}(P(e_1, \frac{1}{2}) \cup \{\infty\}) = S^{n-1}_{e_1,1}$, notice that $\sigma_{0,1}(\infty) = 0 \in S^{n-1}_{e_1,1}$. By property 2. of Lemma 12.4.1, there exist $\psi_1, \psi_2 \in \mathcal{S}(\hat{\mathbb{R}}^n)$ such that $\psi_1(\Sigma_1) = P(e_1, \frac{1}{2}) \cup \{\infty\}$ and $\psi_2(\Sigma_2) = S^{n-1}_{e_1,1}$. Hence $\phi \equiv \psi_2^{-1} \circ \sigma_{0,1} \circ \psi_1$ is a Möbius transformation, because composition of two similarities and an inversion, moreover, clearly, $\phi(\Sigma_1) = \Sigma_2$.

$\square$

Theorem 12.4.5 *Let $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$ and let Σ be a sphere of $\hat{\mathbb{R}}^n$ such that $\phi(x) = x \; \forall x \in \Sigma$. Then ϕ is either $id_{\hat{\mathbb{R}}^n}$ or the reflection or inversion w.r.t. Σ, depending on the fact that Σ is a hyperplane united with ∞ or a Euclidean sphere, respectively.*

Proof. Σ is either a hyperplane $\cup \{\infty\}$ or a $(n-1)$-dimensional sphere in $\mathbb{R}^n$. Thus an inversion w.r.t. Σ can be either a reflection w.r.t. a hyperplane (that fixes ∞) or an inversion w.r.t. a Euclidean sphere.

We start by assuming that $\Sigma = P(e_n, 0) \cup \{\infty\}$, but

$$P(e_n, 0) = \operatorname{span}(e_n)^\perp = \operatorname{span}(e_1, \ldots, e_{n-1}) = \mathbb{R}^{n-1},$$

hence $\Sigma = \hat{\mathbb{R}}^{n-1}$. By the hypothesis that ϕ fixes all the points of Σ we have, in particular:

- $\phi(\infty) = \infty \implies \phi$ is a Euclidean similarity by th. 12.4.2, i.e. $\phi(x) = a + kAx$, $k > 0$, $A \in \mathrm{O}(n)$, for all $x \in \mathbb{R}^n$
- $\phi(0) = 0 \implies \phi = kA$
- $\phi(e_1) = e_1 \implies |\phi(e_1) - \phi(0)| = |e_1 - 0| = |e_1| = 1$, but since $\phi = kA$ this is equivalent to $|kAe_1 - kA0| = k|Ae_1| \underset{A \in \mathrm{O}(n)}{=} k|e_1| = k$, which implies $k = 1$, so $A \in \mathrm{O}(n)$
- $\phi(e_j) = e_j$, $j = 2, \ldots, n-1$ implies that the matrix $A \in \mathrm{O}(n)$ associated to ϕ w.r.t. the canonical basis of $\mathbb{R}^n$ is either

$$A = \begin{pmatrix} I_{n-1} & 0 \\ 0 & 1 \end{pmatrix} = id_{\mathbb{R}^n} \quad \text{or} \quad A = \begin{pmatrix} I_{n-1} & 0 \\ 0 & -1 \end{pmatrix},$$

  because these are the only possible options compatible with the fact that $\det(A) = \pm 1$.

So, either ϕ is the identity on $\mathbb{R}^n$, extended to the identity on $\hat{\mathbb{R}}^n$ because $\phi(\infty) = \infty$, or ϕ is the reflection w.r.t Σ. Hence the thesis is proven when $\Sigma = \hat{\mathbb{R}}^{n-1}$.

We now assume that Σ is an arbitrary sphere of $\hat{\mathbb{R}}^n$ and that ϕ fixes Σ. By the transitivity of $\mathcal{M}(\hat{\mathbb{R}}^n)$ on the set of spheres of $\hat{\mathbb{R}}^n$, there exists a Möbius transformation $\psi \in \mathcal{M}(\hat{\mathbb{R}}^n)$ such that $\psi(\Sigma) = \hat{\mathbb{R}}^{n-1}$, i.e. $\psi(s) = x \in \hat{\mathbb{R}}^{n-1}$ for all $s \in \Sigma$. It follows that, for all $x \in \hat{\mathbb{R}}^{n-1}$,

$$(\psi \circ \phi \circ \psi^{-1})(x) = \psi(\phi(s)) = \psi(s) = x,$$

i.e. ψ fixes $\hat{\mathbb{R}}^{n-1}$ so, thanks to what proven above, $\psi \circ \phi \circ \psi^{-1} = id_{\hat{\mathbb{R}}^n}$ or $\psi \circ \phi \circ \psi^{-1} \equiv \rho$, the reflection w.r.t. $\hat{\mathbb{R}}^{n-1}$. By composing on the left both members by ψ^{-1} and on the right by ψ, we have that it is either $\phi = \psi^{-1} \circ \psi = id_{\hat{\mathbb{R}}^n}$ or $\phi = \psi^{-1} \circ \rho \circ \psi$.

We now want to understand what kind of transformation $\psi^{-1} \circ \rho \circ \psi$ is. To this scope, let us consider, instead of the generic $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$, a reflection or inversion σ w.r.t. Σ, which is not the identity. By repeating the argument above on σ, we obtain that $\psi \circ \sigma \circ \psi^{-1} = \rho$, the reflection or inversion w.r.t. Σ, i.e. $\sigma = \psi^{-1} \circ \rho \circ \psi = \phi$. □

This result will be fundamental to prove theorem 12.5.1.

We know that reflections and inversions fix the points of the hyperplane or sphere w.r.t. they act, respectively. The theorem just proven tells us that *this condition is sufficient to determine if a Möbius transformation is a pure reflection or inversion*, provided that we have excluded the possibility that it is the identity on the whole $\hat{\mathbb{R}}^n$, which is particularly easy because it is sufficient to consider any point not belonging to the hyperplane or the sphere.

The last property of Möbius transformations that we prove here refers to inverse points.

Def. 12.4.4 *Let Σ be a sphere of $\hat{\mathbb{R}}^n$ and σ the reflection or inversion w.r.t. Σ. Two points $x, y \in \hat{\mathbb{R}}^n$ are said to be* ***inverse points*** *w.r.t. Σ if $y = \sigma(x)$.*

Theorem 12.4.6 *Let $\phi \in \mathcal{M}(\hat{\mathbb{R}}^n)$ and let Σ be a sphere of $\hat{\mathbb{R}}^n$. If x and y are inverse points w.r.t. Σ, then $\phi(x)$ and $\phi(y)$ are also inverse points w.r.t. $\Sigma' = \phi(\Sigma)$.*

Proof. The thesis of the theorem is trivially true if ϕ is the identity. So, let us assume that ϕ is not the identity and that σ is the reflection or inversion w.r.t. Σ. Then, $\phi \circ \sigma \circ \phi^{-1}$ fixes each point of $\Sigma' = \phi(\Sigma)$ and so $\phi \circ \sigma \circ \phi^{-1} = \rho$ is the reflection or the inversion w.r.t. Σ'. Finally, if x and y are inverse points w.r.t. Σ, i.e. $y = \sigma(x)$, then

$$\rho(\phi(x)) = (\rho \circ \phi)(x) = (\phi \circ \sigma \circ \phi^{-1} \circ \phi)(x) = (\phi \circ \sigma)(x) = \phi(\sigma(x)) = \phi(y),$$

i.e. $y = \sigma(x)$ implies $\phi(y) = \rho(\phi(x))$. □

12.4.3 The conformality of Möbius transformations

A conformal transformation is a map that maintains angles. In this section we show that conformal and Möbius transformations are tightly interconnected to the point of being confounded in a Euclidean vector space of dimension higher or equal to 3.

An intuitive idea behind this fact can be obtained by considering two intersecting spheres of $\hat{\mathbb{R}}^n$, Σ_1 and Σ_2: if ϕ is a Möbius transformation, then $\phi(\Sigma_1)$ and $\phi(\Sigma_2)$ are two other intersecting spheres Σ_1' and Σ_2' of $\hat{\mathbb{R}}^n$. It is natural to ask oneself how Σ_1' and Σ_2' are positioned to one another when compared to Σ_1 and Σ_2, since Möbius transformations are continuous functions and contain Euclidean similarities, intuitively we imagine that they are positioned more or less in the same way.

Furthermore, if this is the case, then the angle between the normal vectors n_1 of Σ_1 and n_2 of Σ_2 at an intersecting point $x \in \Sigma_1 \cap \Sigma_2$ should not change.

The path that we will follow to make this argument rigorous starts with a definition.

Def. 12.4.5 *Let $U \subseteq \mathbb{R}^n$ open and $\phi : U \to \mathbb{R}^n$, $f \in \mathscr{C}^1(U)$, i.e. all the partial derivatives $\frac{\partial \phi_i}{\partial x_j}$ exists and they are continuous functions on U. ϕ is said to be* ***conformal*** *if there is a function $\kappa : U \to \mathbb{R}^+$, called the* ***scale factor*** *of ϕ, such that*

$$\frac{1}{\kappa(x)} J_\phi(x) \in \mathrm{O}(n) \qquad \forall x \in U,$$

$J_\phi(x)$ being the Jacobian matrix of ϕ calculated in x.

In other words, a conformal function is a continuously differentiable map whose Jacobian matrix can be turned into an orthogonal one simply by re-scaling its coefficients with a positive factor that is allowed to change in every point of the function domain.

Def. 12.4.6 *Given $x, y \in \mathbb{R}^n$, $x, y \neq 0$, we denote with $\theta(x, y)$ the angle between them, i.e. the only angle in $[0, \pi]$ that verifies this equation:*

$$\cos(\theta(x, y)) = \frac{\langle x, y \rangle}{|x||y|}. \tag{12.15}$$

$f : \mathbb{R}^n \to \mathbb{R}^n$ preserves the angle between non-zero vectors if $\theta(f(x), f(y)) = \theta(x, y)$ for all $x, y \in \mathbb{R}^n$.

It is clear that an orthogonal tranformation $f \in \mathrm{O}(n)$ preserves the angle between non-zero vectors because it preserves the scalar product between them and their norms. The next lemma says that, among linear transformations, the orthogonal ones are the only angle preserving maps modulo a scalar coefficient.

Lemma 12.4.4 *Let A a $n \times n$ real matrix. Then, there is a $k \in \mathbb{R}^+$ such that $k^{-1}A \in \mathrm{O}(n)$ if and only if A preserves the angle between non-zero vectors.*

Proof.

$\boxed{\Longrightarrow}$: we assume there is a k such that $k^{-1}A$ is an orthogonal matrix, then $k^{-1}A$ is non-singular and so, for all $x, y \in \mathbb{R}^n$, $x, y \neq 0$, also Ax and Ay are non-zero vectors and we can write:

$$\begin{aligned} \cos(\theta(Ax, Ay)) &= \frac{\langle Ax, Ay\rangle}{|Ax||Ay|} = \frac{\langle k^{-1}Ax, k^{-1}Ay\rangle}{|k^{-1}Ax||k^{-1}Ay|} \\ &\underset{k^{-1}A \in \mathrm{O}(n)}{=} \frac{\langle x, y\rangle}{|x||y|} = \cos(\theta(x, y)). \end{aligned}$$

$\boxed{\Longleftarrow}$: conversely, we suppose that A preserves the angle between non-zero vectors. Then, in particular,

$$\theta(Ae_i, Ae_j) = \theta(e_i, e_j) = \frac{\pi}{2} \qquad \forall i, j \in \{1, \ldots, n\},\ i \neq j.$$

Hence, $(Ae_1, \ldots, Ae_n)$ is an orthogonal basis of $\mathbb{R}^n$, so, if we normalize each vector and we set it as a column of a matrix B, i.e.

$$B = \begin{pmatrix} | & \cdots & | \\ \frac{Ae_1}{|Ae_1|} & \cdots & \frac{Ae_n}{|Ae_n|} \\ | & \cdots & | \end{pmatrix},$$

then B belongs to $\mathrm{O}(n)$ and so does B^{-1} because $\mathrm{O}(n)$ is a group.

By direct computation we get $Be_i = \frac{Ae_i}{|Ae_i|}$ for all $i = 1, \ldots, n$, so, if we multiply both members by $|Ae_i|$ and compose them with B^{-1} we get $B^{-1}Ae_i = |Ae_i|e_i \equiv c_i e_i$, with $c_i > 0$, for all $i = 1, \ldots, n$.

Finally, notice that $B^{-1}A$ preserves the angles between non-zero vectors because it is the composition of two angle-preserving operators, so, using definition (12.15) and the injectivity of the cosine function in $[0, \pi]$ we have that, for all $i, j = 1, \ldots, n$, $i \neq j$,

$$\begin{aligned} \theta(B^{-1}A(e_i + e_j), B^{-1}Ae_j) = \theta(e_i + e_j, e_j) \quad &\Longleftrightarrow \quad \frac{\langle c_i e_i + c_j e_j, c_j e_j\rangle}{|c_j e_j||c_i e_i + c_j e_j|} = \frac{\langle e_i + e_j, e_j\rangle}{|e_i + e_j||e_j|} \\ &\Longleftrightarrow \quad \frac{c_j^2}{c_j\sqrt{c_i^2 + c_j^2}} = \frac{1}{\sqrt{2}} \\ &\Longleftrightarrow \quad \frac{c_j}{\sqrt{c_i^2 + c_j^2}} = \frac{1}{\sqrt{2}} \\ &\Longleftrightarrow \quad \frac{\sqrt{2c_j^2}}{\sqrt{c_i^2 + c_j^2}} = 1 \iff 2c_j^2 = c_i^2 + c_j^2 \\ &\Longleftrightarrow \quad c_j = c_i, \end{aligned}$$

thanks to the strict positivity of the coefficients c_i. Thus all the coefficients can be identified with a constant $k > 0$, which implies $B^{-1}Ae_i = ke_i$, $\forall i = 1, \dots, n$, i.e., by direct computation, $B^{-1}A = kI_n$, or $\frac{1}{k}A = B \in \mathrm{O}(n)$. □

We recall that, given a differentiable curve $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$, the tangent vector to γ at $\gamma(0)$ is the vector of $\mathbb{R}^n$ defined by the formula:

$$\gamma'(0) = \lim_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t}.$$

Def. 12.4.7 *Let $\alpha, \beta : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ be two differentiable curves with $\alpha(0) = \beta(0)$ and $\alpha'(0), \beta'(0) \neq 0$. The angle between α and β is defined as the angle between the vectors of $\mathbb{R}^n$ given by $\alpha'(0)$ and $\beta'(0)$.*

We can now give a characterization of conformality that it is often used as an alternative definition of this property.

Theorem 12.4.7 *Let $U \subseteq \mathbb{R}^n$ be open, $\phi : U \to \mathbb{R}^n$, $\phi \in \mathscr{C}^1(U)$. Then, ϕ is conformal if and only if ϕ preserves the angle between curves.*

Proof.

$\boxed{\Longrightarrow}$: if ϕ is conformal, then there is a scale factor $\kappa : U \to \mathbb{R}^+$ such that $\kappa^{-1}(x)J_\phi(x) \in \mathrm{O}(n)$ for all $x \in U$. Let $\alpha, \beta : (-\varepsilon, \varepsilon) \to U$ be two $\mathscr{C}^1$ curves such that $\alpha(0) = \beta(0)$ and $\alpha'(0), \beta'(0) \neq 0$. Then, $\kappa(\alpha(0))J_\phi(\alpha(0)) = \kappa(\beta(0))J_\phi(\beta(0))$ is an orthogonal matrix and so, by Lemma 12.4.4, $J_\phi(\alpha(0)) = J_\phi(\beta(0))$ preserves angles between the non-zero (by hypothesis) vectors $\alpha'(0)$ and $\beta'(0)$. Hence

$$\begin{aligned} \theta\big((\phi \circ \alpha)'(0), (\phi \circ \beta)'(0)\big) &= \theta\big(J_\phi(\alpha(0))\alpha'(0), J_\phi(\beta(0))\beta'(0)\big) \\ &= \theta\big(\alpha'(0), \beta'(0)\big), \end{aligned}$$

which shows that the angle between α and β is the same as the one between $\phi \circ \alpha$ and $\phi \circ \beta$, i.e. ϕ preserves the angle between curves.

$\boxed{\Longleftarrow}$: conversely, by Lemma 12.4.4, if ϕ preserves angles, then $J_\phi(x)$ preserves angles between non-zero vectors for each fixed $x \in U$. Hence, there exists a $\kappa > 0$ such that $\kappa(x)^{-1}J_\phi(x)$ is orthogonal for all $x \in U$ and so ϕ is conformal on U. □

Def. 12.4.8 *Let $U \subseteq \mathbb{R}^n$ open and let $\phi : U \to \mathbb{R}^n$ be a differentiable function. ϕ preserves (resp. reverses) orientation at a point $x \in U$ if $\det J_\phi(x) > 0$ (resp. $\det J_\phi(x) < 0$).*

ϕ preserves (resp. reverses) orientation if ϕ preserves (resp. reverses) orientation at each point of its domain.

Theorem 12.4.8 *Every reflection and inversion in $\mathbb{R}^n$ is conformal and reverses orientation.*

Proof.

Reflections. Let ρ be a reflection w.r.t. a hyperplane. The easiest way to prove that ρ is conformal is by recalling that it is an isometry, hence there exist $b \in \mathbb{R}^n$ and $B \in \mathrm{O}(n)$ such that

$\rho(x) = b + B(x)$ for all $x \in \mathbb{R}^n$, thus $J_\rho(x) = B$ and so ρ verifies the definition of conformality with $\kappa(x) = 1$ $\forall x \in \mathbb{R}^n$.

However, for later use in this proof, let us also verify the conformality of ρ by computing directly the Jacobian of the original expression of the reflection, i.e. $\rho(x) = x + 2(t - \langle a, x\rangle a) = id_{\mathbb{R}^n}(x) + 2t - 2\langle a, x\rangle a$, $|a| = 1$, $t \geqslant 0$. Thanks to eq. (B.11) we have

$$J_\rho(x) = I - 2A,$$

where A is the matrix $A = (a_i a_j)_{1 \leqslant i,j \leqslant n}$. Notice that $J_\rho(x)$ does not depend on the parameter t, so we are allowed to set it to 0, but then ρ becomes an orthogonal (hence linear transformation), i.e. $\rho(x) = J_\rho(x)$, $\forall x \in \mathbb{R}^n$, hence, by property **4.** of theorem 12.2.1 this implies that

$$J_\rho(x) = I - 2A \in \mathrm{O}(n) \tag{12.16}$$

for all $A = (a_i a_j)_{1 \leqslant i,j \leqslant n}$ with $a \in \mathbb{R}^n$, $|a| = 1$.

Let us now prove that ρ reverses orientation. By the transitivity of the action of $\mathrm{SO}(n)$ on S^{n-1}, there is a $\psi \in \mathrm{SO}(n)$ such that $\psi(a) = e_1$ and so for any $x \in \mathbb{R}^n$,

$$\begin{aligned}
(\psi \circ \rho \circ \psi^t)(x) &= \psi(\rho(\psi^t(x))) = \psi\big(\psi^t(x) + 2(t - \big\langle a, \psi^t(x)\big\rangle)a\big) \\
&\underset{\psi^t = \psi^{-1}}{=} x + 2(t - \langle \psi(a), x\rangle)\psi(a) \\
&= x + 2(t - \langle e_1, x\rangle)e_1,
\end{aligned}$$

but $\langle e_1, x\rangle = (x_1, 0, \ldots, 0)^t$, so, by direct computation we get

$$x + 2(t - \langle e_1, x\rangle)e_1 = (-x_1 + 2t, x_2, \ldots, x_n)^t = \eta x + 2te_1,$$

with $\eta = \mathrm{diag}(-1, 0, \ldots, 0)$, hence $J_{\psi\circ\rho\circ\psi^t}(x) = \eta$ and so $\det(J_{\psi\circ\rho\circ\psi^t}(x)) = -1$ for all $x \in \mathbb{R}^n$. Furthermore, $J_{\psi\circ\rho\circ\psi^t}(x) = J_{\psi\circ\rho\circ\psi^{-1}}(x)$ and the functions ψ and ρ are linear and affine, respectively, so their Jacobian matrices do not depend of the evaluation point, which can be arbitrarily taken to be x. Thanks to these considerations and to the chain rule for Jacobian matrices we have

$$J_{\psi\circ\rho\circ\psi^t}(x) = J_\psi(x) J_\rho(x) J_{\psi^{-1}}(x) = J_\psi(x) J_\rho(x) J_\psi(x)^{-1},$$

and so, by Binet's theorem:

$$-1 = \det(J_{\psi\circ\rho\circ\psi^t}(x)) = \cancel{\det(J_\psi(x))} \det(J_\rho(x)) \cancel{\det(J_\psi(x))^{-1}} = \det(J_\rho(x))$$

for all $x \in \mathbb{R}^n$, hence ρ reverses orientation.

<u>Inversions</u>. Let us start by considering an inversion w.r.t. a sphere centered in 0, i.e. $\sigma_{0,r}(x) = \frac{r^2}{|x|^2}x$, defined for $x \neq 0$, we will consider the generic case later. By theorem B.0.4, the computation of the Jacobian matrix of $\sigma_{0,r}$ gives:

$$J_{\sigma_{0,r}}(x) = \frac{r^2}{|x|^2}\left(I - 2\frac{x_i x_j}{|x|^2}\right) = \frac{r^2}{|x|^2}(I - 2A_x) \equiv \kappa(x) B_x,$$

where $A_x = \left(\frac{x_i}{|x|}\frac{x_j}{|x|}\right)_{1 \leqslant i,j \leqslant n}$, $B_x = I - 2A_x$ and $\kappa(x) = \frac{r^2}{|x|^2} \in \mathbb{R}^+$ for all $x \neq 0$. Notice that the entries of the matrix A_x are the components of the normalized vector $\frac{x}{|x|}$, so, thanks to eq. (12.16), $B_x = I - 2A_x$ is orthogonal, for all $x \neq 0$. This proves that $\sigma_{0,r}$ is conformal.

Let us now prove that $\sigma_{0,r}$ reverses orientation: the properties of the determinant imply

$$\det(J_{\sigma_{0,r}}(x)) = \frac{r^{2n}}{|x|^{2n}} \det(I - 2A_x) = \frac{r^{2n}}{|x|^{2n}} \det(J_\rho(x)) \underset{\det(J_\rho(x))=-1}{=} -\frac{r^{2n}}{|x|^{2n}} < 0.$$

Note that $\det(I - 2A_x) = \det(I - 2A) = \det(J_\rho(x))$, because for every fixed $x \neq 0$ the matrix $J_{\sigma_{0,r}}(x) = I - 2A_x = J_\rho(x)$, with $\rho = \rho_{\frac{x}{|x|},0}$.

Finally, let us consider the generic inversion $\sigma_{a,r}(x) = a + \frac{r^2}{|x-a|^2}(x-a)$. If $\tau_a(x) = x + a$ is the translation operator by a, then it is clear that $\sigma_{a,r} = \tau_a \circ \sigma_{0,r} \circ \tau_a^{-1}$. So, since $J_{\tau_a}(x) = I$ for all x, the chain rule for the Jacobian gives:

$$J_{\sigma_{a,r}}(x) = J_{\sigma_{0,r}}(x - a),$$

which allows us to conclude that also $\sigma_{a,r}$ is conformal and reverses orientation for all $a \in \mathbb{R}^n$ thanks to the previous analysis of $\sigma_{0,r}$. The procedure is totally analogous, paying attention to impose the condition $x \neq a$ instead of $x \neq 0$. □

Since Möbius transformations are finite compositions of reflections and inversions, they are conformal too.

Corollary 12.4.4 *Every Möbius transformation is conformal.*

In 2 dimensions, all holomorphic[3] and anti-holomorphic[4] functions with a non-vanishing Jacobian are conformal mappings.

However, as soon as we pass to the third dimensions, conformal mappings are completely determined by Möbius transformations. This result has been first proven by Liouville [**?**] in 1850 in the case of $\mathscr{C}^3$ mappings in $\mathbb{R}^3$ and then it has been quickly extended to higher dimensions. Nonetheless, it remained an open problem for over a century how to relax the hypothesis of this theorem by considering only $\mathscr{C}^1$ functions, until Hartman solved it [**?**], [**?**].

Theorem 12.4.9 (Liouville-Hartman theorem of conformal mappings) *Let $U \subseteq \mathbb{R}^n$ open, $n \geqslant 3$ and $f : U \to \mathbb{R}^n$ a $\mathscr{C}^1$ map. Then, f is conformal is and only if f is the restriction of a Möbius transformation on U.*

12.5 Möbius transformations in the upper half space $\mathcal{U}^n$ and the open unit ball $\mathcal{B}^n$

Up to now we have analyzed the set of Möbius transformations $\mathcal{M}(\hat{\mathbb{R}}^n)$ *on the whole space* $\hat{\mathbb{R}}^n$. In this section we will focus our attention on subgroups of $\mathcal{M}(\hat{\mathbb{R}}^n)$ given by *Möbius transformations that preserve proper subsets of* $\hat{\mathbb{R}}^n$ and on the relationship between them. The information that we will gather will prove to be of crucial importance in the analysis of the hyperbolic models that we will discuss in chapter 13.

The first proper subset of $\hat{\mathbb{R}}^n$ that we will consider is

$$\hat{\mathbb{R}}^{n-1} \cong P(e_n, 0) \cup \{\infty\} = \{x \in \mathbb{R}^n : \langle x, e_n \rangle = 0\} \cup \{\infty\},$$

[3]A function $f : \mathbb{C} \to \mathbb{C}$, $f(z) = u(z) + iv(z)$, is said to be holomorphic if $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ and $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$.

[4]anti-holomorphic if $\frac{\partial u}{\partial x} = -\frac{\partial v}{\partial y}$ and $\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}$.

then we will also deal with the two disjoint, connected subsets of dimension n in which $\hat{\mathbb{R}}^n$ is separated by $\hat{\mathbb{R}}^{n-1}$, i.e. the upper and the lower half space. Finally, we will also discuss the open connected subsets of $\hat{\mathbb{R}}^n$ given by the unit ball and the complementary of its closure. The explicit definitions of these subsets are given below:

- the upper-half space: $\mathcal{U}^n = \{x \in \mathbb{R}^n \ : \ x_n > 0\} = \{x \in \mathbb{R}^n : \langle x, e_n \rangle > 0\}$;
- the lower-half space: $\mathcal{L}^n = \{x \in \mathbb{R}^n \ : \ x_n < 0\} = \{x \in \mathbb{R}^n \ : \ \langle x, e_n \rangle < 0\}$;
- the open unit ball: $\mathcal{B}^n = \{x \in \mathbb{R}^n \ : \ |x| < 1\}$;
- the complementary in $\hat{\mathbb{R}}^n$ of the closed unit ball: $(\overline{\mathcal{B}^n})^c = \{x \in \mathbb{R}^n \ : \ |x| > 1\} \cup \{\infty\}$.

Clearly,

$$\hat{\mathbb{R}}^n = \mathcal{U}^n \sqcup P(e_n, 0) \sqcup \{\infty\} \sqcup \mathcal{L}^n = \mathcal{B}^n \sqcup S^{n-1} \sqcup (\overline{\mathcal{B}^n})^c.$$

The mathematical analysis starts with the stereographic projection relative to the sphere embedded in $\hat{\mathbb{R}}^n$, which is the isomorphism between $\hat{\mathbb{R}}^{n-1} \cong P(e_n, 0) \cup \{\infty\}$ and S^{n-1} given by

$$\begin{array}{rccl} \hat{\pi}: & \hat{\mathbb{R}}^{n-1} & \xrightarrow{\sim} & S^{n-1} \\ & x & \longmapsto & \hat{\pi}(x) := \begin{cases} \left(\frac{2x_1}{1+|x|^2}, \ldots, \frac{2x_{n-1}}{1+|x|^2}, \frac{|x|^2-1}{1+|x|^2}\right) & \text{if } x \neq \infty \\ e_n \quad \text{if } x = \infty \end{cases} \end{array}.$$

For the aim of this section, it is fundamental to use theorem 12.3.2 which guarantees that we can interpret $\hat{\pi}$ as the restriction of an inversion w.r.t. a sphere in $\hat{\mathbb{R}}^n$, precisely $\hat{\pi} = \sigma_{e_n, \sqrt{2}}\big|_{\hat{\mathbb{R}}^{n-1}}$, or

$$\begin{array}{rccl} \hat{\pi}: & \hat{\mathbb{R}}^{n-1} \subset \hat{\mathbb{R}}^n & \xrightarrow{\sim} & S^{n-1} \subset \hat{\mathbb{R}}^n \\ & x & \longmapsto & \begin{cases} e_n + \frac{2}{|x-e_n|^2}(x - e_n) & \text{if } x \neq \infty \\ e_n \quad \text{if } x = \infty \end{cases} \end{array}.$$

We have all the information we need in order to understand the action of both $\sigma_{e_n, \sqrt{2}}$ and $\hat{\pi}$, which is depicted in Figure 12.5 in the two-dimensional case: by the properties of an inversion w.r.t. a sphere, $\sigma_{e_n, \sqrt{2}}$ will leave the sphere $S^{n-1}_{e_n, \sqrt{2}}$ fixed and it will inverse the points in its interior to points in its exterior. In particular, its center, given by e_n, will be sent to ∞ and the points on S^{n-1} will be mapped on $\hat{\mathbb{R}}^{n-1}$.

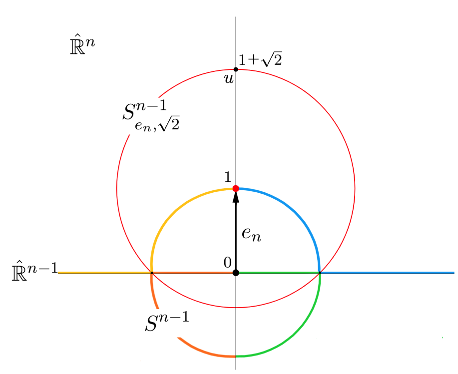


Figure 12.5: Two-dimensional representation of the geometric objects involved in the stereographic projection in $\hat{\mathbb{R}}^n$.

Let us now study the action of $\sigma_{e_n,\sqrt{2}}$ on $\mathcal{U}^n, \mathcal{L}^n, \mathcal{B}^n, (\overline{\mathcal{B}^n})^c$. The most important piece of information that we need to understand this action consists in recalling that $\sigma_{e_n,\sqrt{2}}$ is a homeomorphism, so it maps connected subsets of $\hat{\mathbb{R}}^n$ into connected subsets of $\hat{\mathbb{R}}^n$.

In order to single out the image of each connected subset via $\sigma_{e_n,\sqrt{2}}$ it is sufficient to think about the fact that $\hat{\mathbb{R}}^{n-1}$ splits $\hat{\mathbb{R}}^n$ into $\mathcal{U}^n$ and $\mathcal{L}^n$, and also that $\hat{\mathbb{R}}^n$ it is mapped to the spherical surface S^{n-1}, thus:

- either $\sigma_{e_n,\sqrt{2}}$ maps $\mathcal{U}^n$ to $\mathcal{B}^n$ and $\mathcal{L}^n$ to $(\overline{\mathcal{B}^n})^c$
- or, $\sigma_{e_n,\sqrt{2}}$ maps $\mathcal{U}^n$ to $(\overline{\mathcal{B}^n})^c$ and $\mathcal{L}^n$ to $\mathcal{B}^n$.

In order to choose between these two mutually exclusive options, it is enough to consider the image via $\sigma_{e_n,\sqrt{2}}$ of a wisely chosen point, i.e. $u = (1+\sqrt{2})e_n$. In fact, $u \in S^{n-1}_{e_n,\sqrt{2}} \cap \mathcal{U}^n \cap (\overline{\mathcal{B}^n})^c$ and u will remain fixed after the application of $\sigma_{e_n,\sqrt{2}}$ thanks to property 1. in 12.2.2, thus u, as all the other points belonging to the upper half space, will be mapped to $(\overline{\mathcal{B}^n})^c$, i.e.

$$\sigma_{e_n,\sqrt{2}}(\mathcal{U}^n) = \overline{\mathcal{B}^n}^c \quad \text{and} \quad \sigma_{e_n,\sqrt{2}}(\mathcal{L}^n) = \mathcal{B}^n.$$

We can say more, thanks to property 2. in 12.2.2, $\sigma^{-1}_{e_n,\sqrt{2}} = \sigma_{e_n,\sqrt{2}}$, so also the opposite is true, i.e.

$$\sigma_{e_n,\sqrt{2}}(\overline{\mathcal{B}^n}^c) = \mathcal{U}^n \quad \text{and} \quad \sigma_{e_n,\sqrt{2}}(\mathcal{B}^n) = \mathcal{L}^n.$$

The result that we have obtained can be reached in an alternative way. To do that, we need the following preliminary results, direct consequences of a straightforward computation:

$$|\sigma_{e_n,\sqrt{2}}(x)|^2 = \begin{cases} 1 + \frac{4x_n}{|x-e_n|^2} & \text{if } x \neq \infty \\ 1 & \text{if } x = \infty \end{cases}, \tag{12.17}$$

and

$$\langle \sigma_{e_n,\sqrt{2}}(x), e_n \rangle = \begin{cases} \frac{|x|^2-1}{|x-e_n|^2} & \text{if } x \neq \infty \\ 1 & \text{if } x = \infty \end{cases}. \tag{12.18}$$

Let $x \in \mathcal{L}^n$, i.e. $x_n = \langle x, e_n \rangle < 0$, then, using eq. (12.17) we obtain that $|\sigma_{e_n,\sqrt{2}}(x)|^2 < 1$, hence $\sigma_{e_n,\sqrt{2}}(x) \in \mathcal{B}^n$, i.e. $\mathcal{L}^n$ is mapped into $\mathcal{B}^n$. Furthermore, for all $x \in \mathcal{B}^n$, i.e. such that $|x| < 1$, by using eq. (12.18) we get $\langle \sigma_{e_n,\sqrt{2}}(x), e_n \rangle < 0$, i.e. $\sigma_{e_n,\sqrt{2}}(x) \in \mathcal{L}^n$, so also $\mathcal{B}^n$ is mapped into $\mathcal{L}^n$. With analogous arguments it is possible to verify that $\mathcal{U}^n$ is mapped into $\overline{\mathcal{B}^n}^c$ and vice-versa.

Historically, the upper half space and the interior of the unit ball have been, arbitrarily, privileged w.r.t. their counterparts. This explains why, in general, we prefer to identify $\mathcal{U}^n$ with $\mathcal{B}^n$ instead of $(\overline{\mathcal{B}^n})^c$. This can be achieved very easily by swapping $\mathcal{U}^n$ with $\mathcal{L}^n$ thanks to the reflection w.r.t. $\hat{\mathbb{R}}^{n-1}$, i.e. $\rho_{e_n,0}$.

Clearly, $\sigma_{e_n,\sqrt{2}} \circ \rho_{e_n,0} \in \mathcal{M}(\hat{\mathbb{R}}^n)$ and so it is an isomorphism between $\hat{\mathbb{R}}^n$ and itself. Thus, the transformation $\sigma_{e_n,\sqrt{2}} \circ \rho_{e_n,0}|_{\mathcal{U}^n} : \mathcal{U}^n \xrightarrow{\sim} \mathcal{B}^n$ is an isomorphism between the upper half space and the interior of the unit ball.

Def. 12.5.1 *The Möbius transformation $\eta \equiv \sigma_{e_n,\sqrt{2}} \circ \rho_{e_n,0} \in \mathcal{M}(\hat{\mathbb{R}}^n)$, whose restriction to $\mathcal{U}^n$ allows us to identify $\mathcal{U}^n$ and $\mathcal{B}^n$ is called* ***standard transformation****.*

Def. 12.5.2 *We denote by $\mathcal{M}(\mathcal{U}^n)$ and $\mathcal{M}(\mathcal{B}^n)$, the set of Möbius transformations stable on the upper-half space and the open unit ball, respectively, i.e.*

$$\mathcal{M}(\mathcal{U}^n) = \{\phi \in \mathcal{M}(\hat{\mathbb{R}}^n) \ : \ \phi(\mathcal{U}^n) = \mathcal{U}^n\} \tag{12.19}$$

$$\mathcal{M}(\mathcal{B}^n) = \{\phi \in \mathcal{M}(\hat{\mathbb{R}}^n) \ : \ \phi(\mathcal{B}^n) = \mathcal{B}^n\}. \tag{12.20}$$

It is possible to verify that both of them are subgroups of $\mathcal{M}(\hat{\mathbb{R}}^n)$.

Since $\mathcal{U}^n$ and $\mathcal{B}^n$ are identified through the Möbius transformation η, it is possible to define in a natural way, via η, an isomorphism that permits to identify their Möbius subgroups as shown in the following commutative diagram:

$$\begin{tikzcd}
\mathcal{U}^n \arrow[r, "\eta"] \arrow[d, "\phi"'] & \mathcal{B}^n \arrow[d, dashed, "\eta\circ\phi\circ\eta^{-1}"] \\
\mathcal{U}^n \arrow[r, "\eta"'] & \mathcal{B}^n.
\end{tikzcd}$$

The function

$$\begin{array}{rccl} \iota: & \mathcal{M}(\mathcal{U}^n) & \xrightarrow{\sim} & \mathcal{M}(\mathcal{B}^n) \\ & \phi & \longmapsto & \iota(\phi) := \eta \circ \phi \circ \eta^{-1} \end{array}$$

is clearly an isomorphism of groups.

Let us now focus on the link between $\mathcal{M}(\hat{\mathbb{R}}^n)$ and $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$. In particular, the problem of extending an element of $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$ to the whole $\hat{\mathbb{R}}^n$ is related to the following concept.

Def. 12.5.3 (Poincaré extension) *Let $t \geqslant 0$, $r > 0$, $a \in \mathbb{R}^{n-1}$, $|a| = 1$, and $\tilde{a} = (a, 0) \in \mathbb{R}^n$. The Poincaré extension $\tilde{\phi} \in \mathcal{M}(\hat{\mathbb{R}}^n)$ of $\phi \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$ is defined as follows:*

- *if $\phi = \rho_{a,t}$, then $\tilde{\phi} := \rho_{\tilde{a},t}$;*
- *if $\phi = \sigma_{a,r}$, then $\tilde{\phi} = \sigma_{\tilde{a},r}$;*
- *if $\phi = \phi_1 \circ \cdots \circ \phi_m$, then $\tilde{\phi} := \tilde{\phi}_1 \circ \ldots \tilde{\phi}_m$, where ϕ_i is a reflection or an inversion of $\hat{\mathbb{R}}^{n-1}$, $\forall i \in \{1, \ldots, m\}$.*

To show the usefulness of the Poincaré extension we need the following intermediate result.

Lemma 12.5.1 *Let $\phi \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$, then its Poincaré extension $\tilde{\phi}$ is stable on the hyperplane $\hat{\mathbb{R}}^{n-1} = P(e_n, 0) \cup \{\infty\}$, i.e. $\tilde{\phi}(\hat{\mathbb{R}}^{n-1}) = \hat{\mathbb{R}}^{n-1}$.*

Proof. By definition of Poincaré extension, to prove the statement it is sufficient to prove it for the simple cases of $\phi = \rho_{a,t}$ and $\phi = \sigma_{a,r}$.

- If $\phi = \rho_{a,t}$, then $\tilde{\phi} = \rho_{\tilde{a},t}$, with $\tilde{a} = (a, 0)$. Let us consider the hyperplane $\hat{\mathbb{R}}^{n-1} = P(e_n, 0) \cup \{\infty\}$. Clearly $\rho_{\tilde{a},t}(\infty) = \infty$. Let us consider $x \in P(e_n, 0)$, i.e. $\langle x, e_n\rangle = 0$. By definition $\rho_{\tilde{a},t}(x) = x + 2(t - \langle x, \tilde{a}\rangle)\tilde{a}$, since $\langle \tilde{a}, e_n\rangle = 0$, then

$$\langle \rho_{\tilde{a},t}(x), e_n\rangle = \langle x + 2(t - \langle x, \tilde{a}\rangle)\tilde{a}, e_n\rangle = \langle x, e_n\rangle + 2(t - \langle x, \tilde{a}\rangle)\langle \tilde{a}, e_n\rangle = 0,$$

  hence $\rho_{\tilde{a},t}(x) \in P(e_n, 0)$, so $\hat{\mathbb{R}}^{n-1}$ is globally fixed by $\tilde{\phi}$.

- If $\phi = \sigma_{a,r}$, then $\tilde{\phi} = \sigma_{\tilde{a},r}$, with $\tilde{a} = (a, 0)$ because $\langle \tilde{a}, e_n \rangle = 0$. Let us consider the hyperplane $\hat{\mathbb{R}}^{n-1} = P(e_n, 0) \cup \{\infty\}$. Thence $\sigma_{\tilde{a},t}(\infty) = \tilde{a} \in P(e_n, 0)$. Let us consider $x \in P(e_n, 0)$, i.e. $\langle x, e_n \rangle = 0$. By definition $\sigma_{\tilde{a},r}(x) = \tilde{a} + \frac{r^2}{|x-\tilde{a}|^2}(x - \tilde{a})$, since $\langle \tilde{a}, e_n \rangle = 0$, then

  $$\langle \sigma_{\tilde{a},t}(x), e_n \rangle = \langle \tilde{a} + \frac{r^2}{|x-\tilde{a}|^2}(x - \tilde{a}), e_n \rangle = \langle \tilde{a}, e_n \rangle + \frac{r^2}{|x-\tilde{a}|^2}(\langle x, e_n \rangle - \langle \tilde{a}, e_n \rangle) = 0.$$

  hence $\sigma_{\tilde{a},r}(x) \in P(e_n, 0)$, so $\hat{\mathbb{R}}^{n-1}$ is globally fixed by $\tilde{\phi}$. □

The following theorem gives a further link between the two subgroups of $\mathcal{M}(\hat{\mathbb{R}}^n)$, $\mathcal{M}(\mathcal{U}^n)$ or its isomorphic group $\mathcal{M}(\mathcal{B}^n)$, and $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$, through the Poincaré extension.

Theorem 12.5.1 *Let $\tilde{\phi} \in \mathcal{M}(\hat{\mathbb{R}}^n)$. Then, $\tilde{\phi} \in \mathcal{M}(\mathcal{U}^n)$ if and only if $\tilde{\phi}$ is the Poincaré extension of $\phi \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$.*

Proof.

$\boxed{\Longleftarrow}$: as we did in the previous lemma, it is sufficient to prove the statement for $\tilde{\phi}$ as the Poincaré extension of a reflection $\phi = \rho_{a,t}$, i.e. $\tilde{\phi} = \rho_{\tilde{a},t}$, or an inversion $\phi = \sigma_{a,r}$, i.e. $\tilde{\phi} = \sigma_{\tilde{a},r}$, with $\tilde{a} = (a, 0)$. Let $x \in \mathcal{U}^n$, i.e. $\langle x, e_n \rangle > 0$. Note that $\langle \tilde{a}, e_n \rangle = 0$.

- if $\tilde{\phi} = \rho_{\tilde{a},t}$, then $\rho_{\tilde{a},t}(x) = x + 2(t - \langle x, \tilde{a} \rangle)\tilde{a}$, then

  $$\langle \rho_{\tilde{a},t}(x), e_n \rangle = \langle x, e_n \rangle + 2(t - \langle x, \tilde{a} \rangle)\langle \tilde{a}, e_n \rangle = \langle x, e_n \rangle > 0,$$

  thus $\rho_{\tilde{a},t}(x) \in \mathcal{U}^n$;

- if $\tilde{\phi} = \sigma_{\tilde{a},r}$, then $\sigma_{\tilde{a},r}(x) = \tilde{a} + \frac{r^2}{|x-\tilde{a}|^2}(x - \tilde{a})$, then

  $$\langle \sigma_{\tilde{a},r}(x), e_n \rangle = \langle \tilde{a}, e_n \rangle + \frac{r^2}{|x-\tilde{a}|^2}(\langle x, e_n \rangle - \langle \tilde{a}, e_n \rangle) = \frac{r^2}{|x-\tilde{a}|^2}\langle x, e_n \rangle > 0,$$

  thus $\sigma_{\tilde{a},r}(x) \in \mathcal{U}^n$.

Note that we could repeat an analogous procedure with $\mathcal{L}^n$ instead of $\mathcal{U}^n$, with the opposite inequality, obtaining that, if $\tilde{\phi}$ is the Poincaré extension of $\phi \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$, then it preserves also the lower-half space.

$\boxed{\Longrightarrow}$: suppose $\psi \in \mathcal{M}(\mathcal{U}^n) \subset \mathcal{M}(\hat{\mathbb{R}}^n)$, we must prove that it exists $\phi \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$ such that $\tilde{\phi} = \psi$. As a natural candidate we consider $\phi \equiv \psi|_{\hat{\mathbb{R}}^{n-1}}$.

First of all, let us check that our candidate is suitable, i.e. that its domain and image are $\hat{\mathbb{R}}^{n-1}$. This is an immediate consequence of the fact that ψ, as a Möbius transformation, is an homeomorphism, so it leaves $\partial\mathcal{U}^n \cong \hat{\mathbb{R}}^{n-1}$ fixed.

Moreover $\phi \equiv \psi|_{\hat{\mathbb{R}}^{n-1}}$ is an homeomorphism on $\hat{\mathbb{R}}^{n-1}$. Moreover, since ψ is a Möbius transformation on $\hat{\mathbb{R}}^n$, ψ preserves the cross ratios in $\hat{\mathbb{R}}^n$, so its restriction ϕ must preserve the cross ratios in $\hat{\mathbb{R}}^{n-1}$, hence, by theorem 12.4.1 $\phi \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$.

The only thing that remains to be done is to prove that $\tilde{\phi}$, the Poincaré extension of ϕ, is ψ, or, analogously, that $\tilde{\phi} \circ \psi^{-1} = id_{\hat{\mathbb{R}}^n}$, in fact, this implies that ψ is the right inverse of $\tilde{\phi}$, which is invertible as an element of $\mathcal{M}(\hat{\mathbb{R}}^n)$, so it coincides with the inverse $\tilde{\phi}^{-1}$. In order to obtain this result, we need two preliminary facts:

1. $\tilde{\phi} \circ \psi^{-1}$ is stable on $\mathcal{U}^n$ as composition of functions that are stable on $\mathcal{U}^n$, indeed $\psi^{-1} \in \mathcal{M}(\mathcal{U}^n)$ by hypothesis and $\tilde{\phi} \in \mathcal{M}(\mathcal{U}^n)$ thanks to the first implication of this theorem;

2. $\tilde{\phi} \circ \psi^{-1}$ fixes $\hat{\mathbb{R}}^{n-1}$ pointwise, i.e. $\tilde{\phi} \circ \psi^{-1}(x) = x$ for all $x \in \hat{\mathbb{R}}^{n-1}$, in fact, by definition $\tilde{\phi}\Big|_{\hat{\mathbb{R}}^{n-1}} = \phi = \psi|_{\hat{\mathbb{R}}^{n-1}}$, so $\tilde{\phi} \circ \psi^{-1}\Big|_{\hat{\mathbb{R}}^{n-1}} = \tilde{\phi}\Big|_{\hat{\mathbb{R}}^{n-1}} \circ \psi^{-1}|_{\hat{\mathbb{R}}^{n-1}} = \psi|_{\hat{\mathbb{R}}^{n-1}} \circ \psi^{-1}|_{\hat{\mathbb{R}}^{n-1}} = id_{\hat{\mathbb{R}}^{n-1}}$.

The result 2. guarantees that we can apply theorem 12.4.5 with $\Sigma = \hat{\mathbb{R}}^{n-1}$. This implies that, either $\tilde{\phi} \circ \psi^{-1} = id_{\hat{\mathbb{R}}^n}$ or $\tilde{\phi} \circ \psi^{-1} = \rho_{e_n,0}$. However, this second option is not possible because, by 1. it is stable on $\mathcal{U}^n$ while $\rho_{e_n,0}(\mathcal{U}^n) = \mathcal{L}^n$, hence $\tilde{\phi} = \psi$. □

An immediate consequence of this last theorem is the following corollary:

Corollary 12.5.1 *$\mathcal{M}(\mathcal{U}^n)$ and $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$ are isomorphic (as subgroups of $\mathcal{M}(\hat{\mathbb{R}}^n)$).*

The isomorphism p between $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$ and $\mathcal{M}(\mathcal{U}^n)$ is given by the Poincaré extension as follows:

$$\begin{array}{rccc} p: & \mathcal{M}(\hat{\mathbb{R}}^{n-1}) & \xrightarrow{\sim} & \mathcal{M}(\mathcal{U}^n) \\ & \phi & \longmapsto & p(\phi) := \tilde{\phi} \end{array}$$

Indeed by the definition of the Poincaré extension it is clear that for all ϕ it exists only Poincaré extension $\tilde{\phi}$ and $\tilde{\phi} \in \mathcal{M}(\mathcal{U}^n)$ because of the first implication we proved in the previous theorem. Moreover, thanks to the second implication of the theorem, for all $\psi \in \mathcal{M}(\mathcal{U}^n)$ it exists only one $\phi \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$ such that $\tilde{\phi} = \psi$. By direct computation, it can be proven that p is also a group homomorphism.

Now we are going to analyze the link between $\mathcal{M}(\mathcal{U}^n)$ and $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$ from another perspective which involves angles. This is not surprising since we have already underlined the conformality of Möbius transformations in subsection 12.4.3.

To proceed gradually we need to introduce the concept of orthogonality between (generalized) spheres in $\hat{\mathbb{R}}^n$.

Def. 12.5.4 *Two spheres Σ_1 and Σ_2 of $\hat{\mathbb{R}}^n$ are said to be orthogonal if $\Sigma_1 \cap \Sigma_2 \in \mathbb{R}^n$ and, for all $x \in \Sigma_1 \cap \Sigma_2$, the two normal vectors at x to each sphere are orthogonal.*

The condition $\Sigma_1 \cap \Sigma_2 \in \mathbb{R}^n$ is introduced to guarantee that Σ_1 and Σ_2 actually intersect in at least a point in $\mathbb{R}^n$. The normal vector to a hyperplane has already been defined. The normal vector to a sphere at a point is a vector that is normal to the tangent space to the sphere in the that point.

Since a generalized sphere in $\hat{\mathbb{R}}^n$ can be either a hyperplane $\cup\{\infty\}$ or a Euclidean sphere, there are three possible scenarios, depicted in Figure 12.6 in the 2-dimensional case:

1. if $\Sigma_1 = P(a,t) \cup \{\infty\}$ and $\Sigma_2 = P(b,s) \cup \{\infty\}$, then they are orthogonal if and only if a and b are orthogonal vectors;

2. if $\Sigma_1 = P(a,t) \cup \{\infty\}$ and $\Sigma_2 = S^{n-1}_{b,r}$, then they are orthogonal if and only if $b \in P(a,t)$;

3. if $\Sigma_1 = S^{n-1}_{a,r}$ and $\Sigma_2 = S^{n-1}_{b,s}$, then they are orthogonal if and only if $|a-b|^2 = r^2 + s^2$.

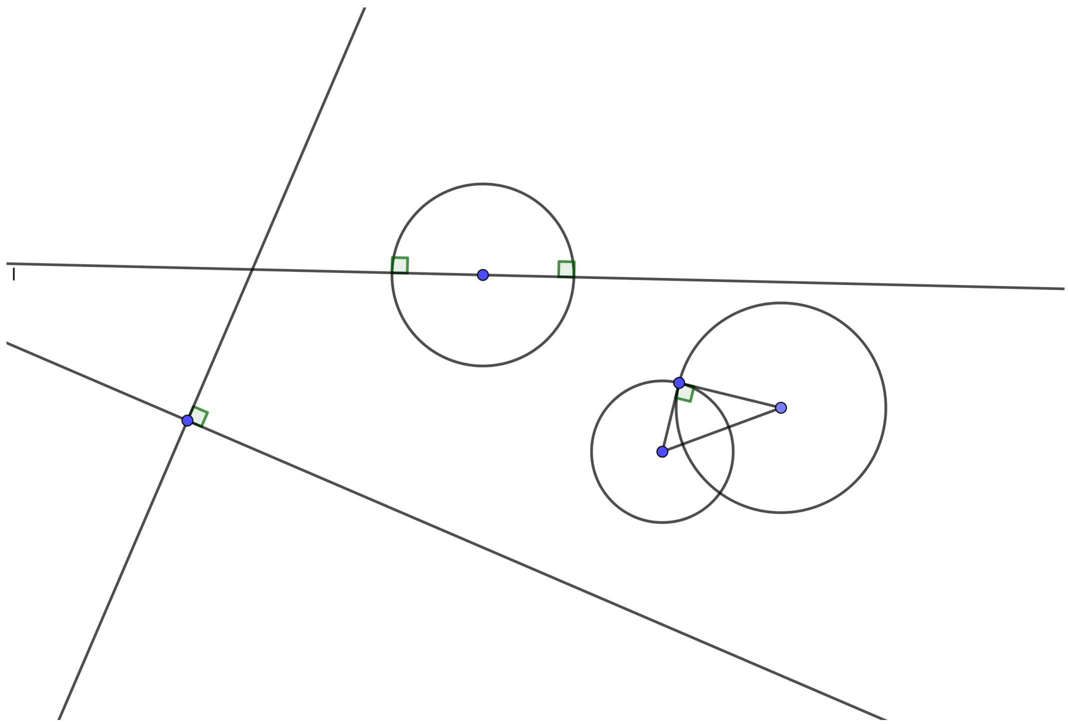


Figure 12.6: the three types of orthogonal spheres in $\hat{\mathbb{R}}^2$.

Note that, by symmetry, in cases 2. and 3. it is sufficient to check the orthogonality condition of the normal vectors in just one of the two points of intersection between the spheres.

Theorem 12.5.2 *Let ϕ be a reflection or inversion with respect to a sphere Σ in $\hat{\mathbb{R}}^n$, then $\phi \in \mathcal{M}(\mathcal{U}^n)$ if and only if Σ is orthogonal to $\hat{\mathbb{R}}^{n-1}$.*

Proof. Let us recall that $\hat{\mathbb{R}}^{n-1} \cong P(e_n, 0) \cup \{\infty\}$. We will consider the cases of reflection and inversion separately.

1. Let $\phi = \rho_{\tilde{a},t}$ be the reflection with respect to $\Sigma = P(\tilde{a}, t) \cup \{\infty\}$, then the following chain of equivalent assertions holds:

$$\begin{aligned} \rho_{\tilde{a},t} \in \mathcal{M}(\mathcal{U}^n) &\overset{\text{th.12.5.1}}{\iff} \exists \rho_{a,t} \in \mathcal{M}(\hat{\mathbb{R}}^{n-1}) \text{ such that } \tilde{\rho}_{a,t} = \rho_{\tilde{a},t} \text{ and } \tilde{a} = (a, 0) \\ &\iff \langle \tilde{a}, e_n \rangle = 0 \\ &\iff \tilde{a} \text{ and } e_n \text{ are orthogonal vectors} \\ &\iff \Sigma = P(\tilde{a}, t) \cup \{\infty\} \text{ and } \hat{\mathbb{R}}^{n-1} = P(e_n, 0) \cup \{\infty\} \text{ are orthogonal.} \end{aligned}$$

2. Let $\phi = \sigma_{\tilde{a},r}$ be the inversion with respect to $\Sigma = S^{n-1}_{\tilde{a},r}$, then the following chain of equivalent assertions holds:

$$\begin{aligned} \sigma_{\tilde{a},r} \in \mathcal{M}(\mathcal{U}^n) &\overset{\text{th.12.5.1}}{\iff} \exists \sigma_{a,r} \in \mathcal{M}(\hat{\mathbb{R}}^{n-1}) \text{ s.t. } \tilde{\sigma}_{a,r} = \sigma_{\tilde{a},r} \text{ and } \tilde{a} = (a, 0) \\ &\iff \langle \tilde{a}, e_n \rangle = 0 \\ &\iff \tilde{a} \in P(e_n, 0) \\ &\iff \Sigma = S^{n-1}_{\tilde{a},r} \text{ and } \hat{\mathbb{R}}^{n-1} = P(e_n, 0) \cup \{\infty\} \text{ are orthogonal.} \end{aligned}$$

□

Since every Möbius transformation can be written as the composition of reflections and inversions, a direct consequence of the previous theorem is the following.

Corollary 12.5.2 *Every Möbius transformation $\phi \in \mathcal{M}(\mathcal{U}^n)$ is the composition of reflections and inversions with respect to spheres in $\hat{\mathbb{R}}^n$ which are orthogonal to $\hat{\mathbb{R}}^{n-1}$.*

For the following theorem we need to define the subgroup of $\mathcal{I}(\hat{\mathbb{R}}^n)$ stable on the upper half space:

$$\mathcal{I}(\mathcal{U}^n) = \{\psi \in \mathcal{I}(\hat{\mathbb{R}}^n) \ : \ \psi(\mathcal{U}^n) = \mathcal{U}^n\} = \mathcal{I}(\hat{\mathbb{R}}^n) \cap \mathcal{M}(\mathcal{U}^n). \tag{12.21}$$

Theorem 12.5.3 *Let $\phi \in \mathcal{M}(\mathcal{U}^n)$ such that $\phi(\infty) \neq \infty$. Let Σ be the isometric sphere of ϕ and $\phi = \psi \circ \sigma$ its decomposition (see theorem 12.4.2), with σ the reflection or inversion w.r.t. Σ and $\psi \in \mathcal{I}(\hat{\mathbb{R}}^n)$. Then Σ is orthogonal to $\hat{\mathbb{R}}^{n-1}$ and $\psi \in \mathcal{I}(\mathcal{U}^n)$.*

Proof. From corollary 12.5.2, since $\phi \in \mathcal{M}(\mathcal{U}^n)$ and $\phi = \psi \circ \sigma$, Σ is orthogonal to $\hat{\mathbb{R}}^{n-1}$. What remains to be proven is that $\psi \in \mathcal{I}(\mathcal{U}^n)$. σ is the reflection or inversion with respect to Σ, which is orthogonal to $\hat{\mathbb{R}}^{n-1}$. From theorem 12.5.2 this implies that $\sigma \in \mathcal{M}(\mathcal{U}^n)$. Since $\sigma^{-1} = \sigma$, then $\psi = \sigma \circ \phi$. Now $\sigma \in \mathcal{M}(\mathcal{U}^n)$ and $\phi \in \mathcal{M}(\mathcal{U}^n)$, hence $\psi \in \mathcal{M}(\mathcal{U}^n)$ as composition of elements of $\mathcal{M}(\mathcal{U}^n)$. Moreover $\psi \in \mathcal{I}(\hat{\mathbb{R}}^n)$, so we can conclude that $\psi \in \mathcal{I}(\mathcal{U}^n)$. □

Now we are going to analyze the properties of the last Möbius subgroup: $\mathcal{M}(\mathcal{B}^n)$.

As we already know it is possible to identify $\mathcal{U}^n$ and $\mathcal{B}^n$ through the standard transformation η defined in 12.5.1 as $\eta = \sigma_{e_n,\sqrt{2}} \circ \rho_{e_n,0}$. Moreover, it permits to define the isomorphism of subgroups ι as follows:

$$\begin{array}{rccl} \iota : & \mathcal{M}(\mathcal{U}^n) & \xrightarrow{\sim} & \mathcal{M}(\mathcal{B}^n) \\ & \phi & \longmapsto & \iota(\phi) := \eta \circ \phi \circ \eta^{-1}. \end{array}$$

Up to now we have analyzed in detail the properties of $\mathcal{M}(\mathcal{U}^n)$. Because of the isomorphism between $\mathcal{M}(\mathcal{U}^n)$ and $\mathcal{M}(\mathcal{B}^n)$ it is reasonable to think that analogous properties should hold for $\mathcal{M}(\mathcal{B}^n)$. This is actually the case and it order to prove it, we will make large use of the isomorphism between the two subgroups.

As we have seen in corollary 12.5.1 the Poincaré extension induces the isomorphism p between $\mathcal{M}(\mathcal{U}^n)$ and $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$.

$$\begin{array}{rccl} p : & \mathcal{M}(\hat{\mathbb{R}}^{n-1}) & \xrightarrow{\sim} & \mathcal{M}(\mathcal{U}^n) \\ & \phi & \longmapsto & p(\phi) := \tilde{\phi}. \end{array}$$

Note that $\hat{\mathbb{R}}^{n-1} = \partial\mathcal{U}^n$, hence the Poincaré extension gives a correspondence between the Möbius subgroup of $\mathcal{U}^n$ and the Möbius subgroup of its border $\partial\mathcal{U}^n$. Analogously we would like to define a Poincaré extension which links the Möbius subgroups associated to $\mathcal{B}^n$ and its border $\partial\mathcal{B}^n = S^{n-1}$, respectively.

First of all we need to identify $\mathcal{M}(S^{n-1})$. For that, we need to search for an analogous version of p for $\mathcal{B}^n$, to do that we will clearly make use of p, which connects $\mathcal{M}(\mathcal{U}^n)$ with $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$. Thence it is important to define $\mathcal{M}(S^{n-1})$ as something related to $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$.

Before giving this definition let us recall that the extended stereographic projection $\hat{\pi} : \hat{\mathbb{R}}^{n-1} \longrightarrow S^{n-1}$ maps bijectively $\hat{\mathbb{R}}^{n-1}$ onto S^{n-1}, $\hat{\pi}(\hat{\mathbb{R}}^{n-1}) = S^{n-1}$ and $\hat{\pi}^{-1}(S^{n-1}) = \mathbb{R}^{n-1}$.

Now we can define the Möbius group of S^{n-1} as follows:

$$\mathcal{M}(S^{n-1}) = \{\phi : S^n \to S^n \text{ such that } \hat{\pi}^{-1} \circ \phi \circ \hat{\pi} \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})\}, \tag{12.22}$$

the following commutative diagram visualizes the action of such Möbius transformations:

$$\begin{tikzcd} S^{n-1} \arrow[d, "\phi"'] & \hat{\mathbb{R}}^{n-1} \arrow[l, "\hat{\pi}"'] \arrow[d, dashed, "\hat{\pi}^{-1}\circ\phi\circ\hat{\pi}"] \\ S^{n-1} \arrow[r, "\hat{\pi}^{-1}"'] & \hat{\mathbb{R}}^{n-1}. \end{tikzcd}$$

Clearly $\hat{\pi}$ allows us to define the group isomorphism μ between $\mathcal{M}(\hat{\mathbb{R}}^{n-1})$ and $\mathcal{M}(S^{n-1})$ as follows:

$$\begin{array}{rccc} \mu : & \mathcal{M}(S^{n-1}) & \xrightarrow{\sim} & \mathcal{M}(\hat{\mathbb{R}}^{n-1}) \\ & \phi & \longmapsto & \mu(\phi) := \hat{\pi}^{-1} \circ \phi \circ \hat{\pi}. \end{array}$$

The definition of the Poincaré extension p' for the elements of $\mathcal{M}(S^{n-1})$ to elements of $\mathcal{M}(\mathcal{B}^n)$ is given below.

Def. 12.5.5 *Let $\phi \in \mathcal{M}(S^{n-1})$, let $\psi = \hat{\pi}^{-1} \circ \phi \circ \hat{\pi} = \mu^{-1}(\phi) \in \mathcal{M}(\hat{\mathbb{R}}^{n-1})$ and let $\tilde{\psi} \in \mathcal{M}(\mathcal{U}^n)$ be the Poincaré extension of ψ, $\tilde{\psi} = p(\psi)$. We define the Poincaré extension of ϕ as $\tilde{\phi} \equiv p'(\phi) = \eta \circ \tilde{\psi} \circ \eta^{-1} = \iota \circ p \circ \mu^{-1}(\phi) \in \mathcal{M}(\mathcal{B}^n)$.*

The following commutative diagram visualizes the action of p':

$$\begin{tikzcd} \mathcal{M}(\hat{\mathbb{R}}^{n-1}) \arrow[r, "\mu"] \arrow[d, "p"'] & \mathcal{M}(S^{n-1}) \arrow[d, dashed, "p'"] \\ \mathcal{M}(\mathcal{U}^n) \arrow[r, "\iota"'] & \mathcal{M}(\mathcal{B}^n). \end{tikzcd}$$

Two immediate consequences of this definition are the analogous versions of theorem 12.5.1 and corollary 12.5.1, that can be proven analogously.

Theorem 12.5.4 *$\phi \in \mathcal{M}(\mathcal{B}^n)$ if and only if ϕ is the Poincaré extension of an element of $\mathcal{M}(S^{n-1})$.*

Corollary 12.5.3 *The Poincaré extension p' is an isomorphism between the Möbius groups $\mathcal{M}(S^{n-1})$ and $\mathcal{M}(\mathcal{B}^n)$.*

We will now analyze the analogous version of theorem 12.5.2.

Theorem 12.5.5 *Let ϕ be a reflection or inversion w.r.t. a sphere Σ in $\hat{\mathbb{R}}^n$, then $\phi \in \mathcal{M}(\mathcal{B}^n)$ if and only if Σ is orthogonal to S^{n-1}.*

Proof. Since $\phi \in \mathcal{M}(\mathcal{B}^n)$, let us consider $\psi = \iota^{-1}(\phi) = \eta^{-1} \circ \phi \circ \eta \in \mathcal{M}(\mathcal{U}^n)$. Let us call $\Sigma' = \eta^{-1}(\Sigma)$, after a straightforward computation it is immediate to verify that ψ fixes Σ' pointwise. Moreover $\psi \neq id_{\hat{\mathbb{R}}^n}$, indeed if $\psi = id_{\hat{\mathbb{R}}^n} = \eta^{-1} \circ \phi \circ \eta$, then $\phi = id_{\hat{\mathbb{R}}^n}$, which is false. Theorem 12.4.5 allows us to conclude that ψ is the reflection or inversion w.r.t. Σ'.

Moreover, because of theorem 12.5.2, $\psi \in \mathcal{M}(\mathcal{U}^n)$ if and only if Σ' is orthogonal to $\hat{\mathbb{R}}^{n-1}$. Note that if we apply η to both Σ' and $\hat{\mathbb{R}}^{n-1}$ we obtain $\eta(\Sigma') = \Sigma$ and $\eta(\hat{\mathbb{R}}^{n-1}) = \sigma_{e_n,\sqrt{2}} \circ \rho_{e_n,0}(\hat{\mathbb{R}}^{n-1}) = \sigma_{e_n,\sqrt{2}}(\hat{\mathbb{R}}^{n-1}) = S^{n-1}$.

By corollary 12.4.4, $\eta \in \mathcal{M}(\hat{\mathbb{R}}^n)$ is conformal, thence it preserves angles, and so, in particular, it preserves orthogonality. Finally we can conclude that Σ' is orthogonal to $\hat{\mathbb{R}}^{n-1}$ if and only if Σ is orthogonal to S^{n-1}. $\square$

The following corollary is the analogous version for $\mathcal{B}^n$ of 12.5.2.

Corollary 12.5.4 *Every Möbius transformation in $\mathcal{M}(\mathcal{B}^n)$ is the composition of reflections and inversions w.r.t. spheres of $\hat{\mathbb{R}}^n$ which are orthogonal to S^{n-1}.*

We will now analyze a similar result to 12.5.3.

Theorem 12.5.6 *Let $\phi \in \mathcal{M}(\mathcal{B}^n)$, then:*

1. *if $\phi(\infty) = \infty$, then $\phi \in \mathrm{O}(n) = \mathcal{I}(\mathcal{B}^n)$;*
2. *if $\phi(\infty) \neq \infty$, let Σ be its isometric sphere and let $\phi = \psi \circ \sigma$ be its decomposition*[5]*, with $\psi \in \mathcal{I}(\hat{\mathbb{R}}^n)$ and σ the inversion w.r.t. Σ, then Σ is orthogonal to S^{n-1} and $\psi \in \mathrm{O}(n) = \mathcal{I}(\mathcal{B}^n)$.*

Proof.

1. Let us consider the case $\phi(\infty) = \infty$. By point 1. of theorem 12.4.2, $\phi \in \mathcal{S}(\hat{\mathbb{R}}^n)$ hence it can be written as $\phi(x) = b + kAx$, with $k > 0$, $A \in \mathrm{O}(n)$ and $b \in \mathbb{R}^n$. Notice that, since $\phi \in \mathcal{M}(\mathcal{B}^n)$, the vector b should belong to $\mathcal{B}^n$, hence $|b| < 1$. Indeed if $|b| \geqslant 1$, then $\phi(0) = b \notin \mathcal{B}^n$, but $0 \in \mathcal{B}^n$ and this is contradictory with the hypothesis $\phi \in \mathcal{M}(\mathcal{B}^n)$.

   Let us suppose $b \neq 0$. Clearly $\eta^{-1} \circ \phi \circ \eta \in \mathcal{M}(\mathcal{U}^n)$. Theorem 12.5.1 and lemma 12.5.1 allow us to say that $\eta^{-1} \circ \phi \circ \eta$ is stable on $\hat{\mathbb{R}}^{n-1}$, i.e. $\eta^{-1}(\phi(\eta(\hat{\mathbb{R}}^{n-1}))) = \hat{\mathbb{R}}^{n-1}$, hence $\phi(\eta(\hat{\mathbb{R}}^{n-1})) = \eta(\hat{\mathbb{R}}^{n-1})$, but $\eta(\hat{\mathbb{R}}^{n-1}) = S^{n-1}$, this means that $\phi(S^{n-1}) = S^{n-1}$, i.e. ϕ is stable on S^{n-1}. Since we have supposed $b \neq 0$, we can define $\tilde{b} = A^t \frac{b}{|b|}$. It is easy to verify that $|\tilde{b}| = 1$, thus $\tilde{b} \in S^{n-1}$.

   Now since ϕ is stable on S^{n-1} we must have that $|\phi(\tilde{b})| = 1$. Explicitly:

   $$|\phi(\tilde{b})| = \left| b + k\frac{b}{|b|} \right| = ||b| + k| = |b| + k = 1, \tag{12.23}$$

   which implies that $k = 1 - |b|$, which is positive because $b \in \mathcal{B}^n$.

   Clearly also $-\tilde{b} \in S^{n-1}$, i.e. $|-\tilde{b}| = 1$ and $|\phi(-\tilde{b})| = 1$. Developing the computation and using the fact that $k = 1 - |b|$ we obtain:

   $$|\phi(-\tilde{b})| = \left| b - k\frac{b}{|b|} \right| = ||b| - k| = |2|b| - 1| = 1. \tag{12.24}$$

   Hence $|b| = 0$ or $|b| = 1$ which is contradictory because we assumed that $b \neq 0$ and $|b| < 0$.

   This means that $b = 0$, thus $\phi(x) = kAx$. Since ϕ is stable on S^{n-1}, let us consider $x \in S^{n-1}$, $|x| = 1$ and $|\phi(x)| = 1$, but $1 = |\phi(x)| = k|Ax| = k|x| = k$, hence $k = 1$ and $\phi = A \in \mathrm{O}(n) = \mathcal{I}(S^{n-1})$.

[5] given by 2. in theorem 12.4.2.

2. Let us consider the case $\phi(\infty) \neq \infty$. Let $a = \phi^{-1}(\infty) \in \mathbb{R}^n$, using the decomposition given by point 2. in theorem 12.4.2 $\phi = \psi \circ \sigma$ we have that $\phi(a) = \psi(\sigma(a)) = \infty$, so $\sigma(a) = \psi^{-1}(\infty) = \infty$, hence $\sigma(a) = \infty$. This implies that a is the center of the isometric sphere $\Sigma = S^{n-1}_{a,r}$ and $\sigma = \sigma_{a,r}$. Moreover by corollary 12.5.4, since $\phi \in \mathcal{M}(\mathcal{B}^n)$ the spheres $\Sigma = S^{n-1}_{a,r}$ and S^{n-1} are orthogonal, hence r is such that $|a|^2 = r^2 + 1$. Now, by theorem 12.5.5 we know that $\sigma \in \mathcal{M}(\mathcal{B}^n)$, moreover $\phi \in \mathcal{M}(\mathcal{B}^n)$, hence $\psi \in \mathcal{M}(\mathcal{B}^n)$, but also $\psi \in \mathcal{I}(\hat{\mathbb{R}}^n)$, so $\psi \in \mathcal{M}(\mathcal{B}^n) \cap \mathcal{I}(\hat{\mathbb{R}}^n) = \mathcal{I}(\mathcal{B}^n) = \mathrm{O}(n)$.

□

A direct consequence of the previous theorem is the following corollary.

Corollary 12.5.5 *Let $\phi \in \mathcal{M}(\mathcal{B}^n)$, then $\phi(0) = 0$ if and only if $\phi \in \mathrm{O}(n)$.*

Proof. If $\phi(\infty) = \infty$ then $\phi \in \mathrm{O}(n)$ because of point 1. in the previous theorem.

Let us consider the case of $\phi(\infty) \neq \infty$. Because of the previous theorem we have the decomposition $\phi = \psi \circ \sigma$, with $\psi \in \mathrm{O}(n)$ and σ the inversion w.r.t. the sphere $S^{n-1}_{a,r}$, with $r^2 = |a|^2 - 1$. The condition $\phi(0) = 0$ corresponds to $\phi(0) = \psi(\sigma(0)) = 0$, but, since $\psi \in \mathrm{O}(n)$, $\psi(0) = 0$, thence the previous condition is equivalent to $\sigma(0) = 0$. Now, because of property 1. in theorem 12.2.2, this means that $0 \in S^{n-1}_{a,r}$, hence $|0 - a| = |a| = r$, but $r^2 = |a|^2 - 1$, so $|a|^2 = |a|^2 - 1$ which gives a contradiction. Hence, $\phi \in \mathrm{O}(n)$ if and only if $\phi(0) = 0$. □

Chapter 13

The hyperbolic models

Around 300 B.C, Euclid wrote his celebrated 'Elements' [?], a thirteen-volume work where he presented the fundamentals of Greek geometry and number theory. In the first pages, he exposes his five postulates of planar geometry:

1. *'Let it have been postulated to draw a straight-line from any point to any point'*
2. *'And to produce a finite straight-line continuously in a straight-line'*
3. *'And to draw a circle with any center and radius'*
4. *'And that all right-angles are equal to one another'*
5. *'And that if a straight-line falling across two (other) straight-lines makes internal angles on the same side (of itself whose sum is) less than two right-angles, then the two (other) straight-lines, being produced to infinity, meet on that side (of the original straight-line) that the (sum of the internal angles) is less than two right-angles (and do not meet on the other side)'.*

This last postulate is best known as the parallel postulate and it is equivalent to Playfair's axiom when combined with the first four axioms:

'In a plane, given a line and a point not on it, at most one line parallel to the given line can be drawn through the point'

For over two thousand years, mathematicians have tried to simplify Euclid's axioms of geometry, by proving the fifth axiom from the first four (known as the fifth postulate problem), but without success. However, in the 19th century, things took a surprising turn when mathematicians discovered that in fact the fifth axiom was *independent* from the first four while trying to prove the fifth postulate problem by contradiction by denying the fifth axiom. To the general astonishment of mathematicians at the time, *geometries that refute the fifth axiom (while keeping the first four), turned out to be highly consistent.*

The geometries that reject some of Euclid's postulates are fittingly designated as non-Euclidean geometries. Hyperbolic geometry is a non-Euclidean geometry where we keep the first four postulates and we refute the fifth postulate by replacing it with the following:

'In a plane, given a line and a point not on it, there are infinitely many lines parallel to the given line that can be drawn through the point.'

Funnily enough, while Gauss is thought to be the one of the first mathematicians to have worked on hyperbolic geometry, he never published anything about it out of fear of the '*uproar of the Boeotians*' (1829, letter from Gauss to W. Bessel), to the extend that Gauss' visionary work on non-Euclidean geometry was only found among his papers after his death in 1855.

While the first publications on hyperbolic geometry were independently given by Nikolai Lobachevsky and János Bolyai in 1829 and 1832 respectively, it was only during the second half of the 19th century that hyperbolic geometry was fully developed by mathematicians such as Poincaré and Hilbert, with the culmination point being at the start of the 20th century with Einstein's groundbreaking use of hyperbolic geometry in his formulation of special relativity, thus showing that hyperbolic geometry was not just meant to be left in the dark cupboards of the mathematics department. More recently, hyperbolic geometry has made a come back with its use in artificial intelligence and information processing, such as in [**?**]or [**?**] which make a nice use of Poincaré's and Klein's disk embedding, respectively).

The set of hyperbolic models that we will discuss in this chapter is not exhaustive, we will just concentrate on four of them, listed below.

$$\begin{tikzcd}
 & \text{Poincaré disk} \arrow[r] & \text{Upper-half space} & \text{(Conformal model)} \\
\text{Hyperboloid} \arrow[ur] \arrow[dr] & & & \\
 & \text{Klein model} & & \text{(Projective model)}
\end{tikzcd}$$

The most natural model is the hyperboloid one, from which we can deduce the other ones. The Poincaré and upper-half space models are called conformal because they preserve the Euclidean notion of angle, while lines can be distorted. On the other hand, the Beltrami-Cayley-Klein model (or Klein model for simplicity) has opposite features: the concept of Euclidean straight line is the same, while Euclidean angles are distorted.

Figure 13.1 gives a visual depiction of 2-dimensional hyperbolic models showing how the Poincaré and Klein disks can be obtained from the hyperboloid via two different projection: for the Klein, the center of projection is 0 and the disk is placed at height 1 along the vertical axis, while the Poincaré disk is placed at height 0 and the center of projection is -1, both values are on the vertical axis.

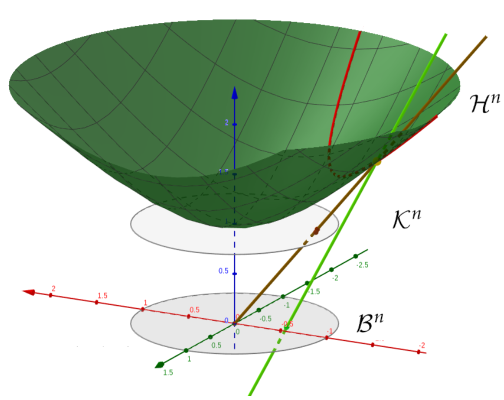


Figure 13.1: The two different projections from the hyperpoloid model onto Poincaré and Klein disks.

13.1 The hyperboloid model and the hyperbolic metric

In this section we are going to define and analyze in detail the first model of hyperbolic geometry: the hyperboloid. A quick recap about the concept of distance and angle in the Euclidean setting will help us underlying similarities and differences between spherical and hyperbolic geometry.

13.1.1 Memories of spherical geometry

The classical way of introducing the concept of angle and spherical distance is based on the Cauchy-Schwarz inequality (lemma 11.1.1). In fact, as a direct consequence, we have that for all $x, y \in \mathbb{R}^n$, there is a number $\alpha(x, y) \in [-1, 1]$ such that

$$\langle x, y\rangle = \alpha(x, y)|x||y|, \tag{13.1}$$

which, for non-zero vectors, satisfies the following properties: $\alpha(x, y) = 0$ if and only if x and y are orthogonal and $\alpha(x, y) = \pm 1$ if and only is x and y are linearly dependent. Being $\cos|_{[0,\pi]}$ a bijective function between $[0, \pi]$ and $[-1, 1]$, with $\cos(0) = 1$, $\cos(\pi/2) = 0$ and $\cos(\pi) = -1$, we have the identification:

$$\alpha(x, y) = \cos(\theta(x, y)), \tag{13.2}$$

where $\theta(x, y) \in [0, \pi]$ is *defined* to be the **angle** between x and y. $\theta(x, y)$ is related to the so-called spherical distance between two vectors, that we recall next.

Def. 13.1.1 *The spherical distance $d_S(x, y)$ between two vectors $x, y \in \mathbb{R}^n$ is the angle between the projections of x and y on the unit sphere S^{n-1}.*

Of course the projection p_{S^1} on the unit sphere of a vector $v \in \mathbb{R}^n$ is simply $p_{S^1}(v) = \frac{v}{|v|}$, to get $|p_{S^1}(v)| = 1$. Using eqs. (13.1) and (13.1) when $|x| = |y| = 1$, by definition of spherical distance $d_S(x, y)$ we have:

$$\cos(d_S(x, y)) = \langle x, y\rangle \iff d_S(x, y) = \arccos(\langle x, y\rangle) \tag{13.3}$$

and, since $\cos^2(d_S(x, y)) + \sin^2(d_S(x, y)) = 1$, $\sin(d_S(x, y)) = \sqrt{1 - \langle x, y\rangle^2}$, where only the positive determination of the square root makes sense here because $d_S(x, y)$ has been defined as the angle between x, y, which belongs to $[0, \pi]$, so $\sin(d_S(x, y)) \geqslant 0$.

The straight line on the sphere through $x, y \in S^{n-1}$ is

$$\ell_{x,y} = \operatorname{span}(x, y) \cap S^{n-1},$$

and the shortest (geodesic) arc between $x, y \in S^{n-1}$ has the expression

$$\gamma(t) = \cos(t)x + \frac{\sin(t)}{\sqrt{1 - \langle x, y\rangle^2}}\Big(y - \langle x, y\rangle\, x\Big), \qquad t \in [0, d_S(x, y)],$$

notice that $\gamma(0) = x$ and

$$\begin{aligned}\gamma(d_S(x,y)) &= \cos(d_S(x,y))x + \frac{\sin(d_S(x,y))}{\sqrt{1-\langle x,y\rangle^2}}\Big(y - \langle x,y\rangle x\Big)\\ &= \langle x,y\rangle x + \frac{\sqrt{1-\langle x,y\rangle^2}}{\sqrt{1-\langle x,y\rangle^2}}\Big(y - \langle x,y\rangle x\Big)\\ &= y,\end{aligned}$$

or

$$\gamma(t) = \cos(t)x + \sin(t)y, \qquad t \in [0, d_S(x,y)],$$

if x and y are orthogonal.

We also remark that sine and cosine are the only functions verifying

$$1 = \cos(t)^2 + \sin(t)^2 = \left|\begin{pmatrix}\cos(t)\\ \sin(t)\end{pmatrix}\right|^2$$

and

$$S^1 = \left\{\begin{pmatrix}\cos(t)\\ \sin(t)\end{pmatrix} \in \mathbb{R}^2 \;:\; t \in \mathbb{R}\right\}.$$

The hyperboloid model that we will analyze now will show analogous features, the major difference being represented by the fact that the circular functions sine and cosine must be replaced by the their hyperbolic counterparts:

$$\cosh(t) = \frac{e^x + e^{-x}}{2}, \;\; \sinh(t) = \frac{e^x - e^{-x}}{2},$$

which, for all $t \in \mathbb{R}$, satisfy

$$\left|\begin{pmatrix}\cosh(t)\\ \sinh(t)\end{pmatrix}\right|^2 = \cosh^2(t) - \sinh^2(t) = 1$$

and

$$\left\|\begin{pmatrix}\cosh(t)\\ \sinh(t)\end{pmatrix}\right\|^2 = -\cosh^2(t) + \sinh^2(t) = -1.$$

13.1.2 The hyperboloid model and its metric

We have just seen how we can build a distance on the sphere from the scalar product and the cosine function. Here we follow exactly the same path by replacing the unit sphere with the unit hyperboloid, i.e. the one defined by $q(x) = -1$ and the cosine by the hyperbolic cosine.

Def. 13.1.2 *The hyperboloid model of hyperbolic geometry is defined as the upper connected part of the level set defined by $q(x) = -1$ in $\mathbb{R}^{1,n}$, explicitly,*

$$\mathcal{H}^n = \{x \in \mathbb{R}^{1,n} : \|x\|^2 = -1, \; x_1 > 0\}.$$

The hyperboloid model can thus be described also as **the set of all unit positive time-like vectors** in $\mathbb{R}^{1,n}$. The analysis of the hyperboloid model starts with a variation of the Cauchy-Schwarz inequality specific for positive time-like vectors.

Theorem 13.1.1 *Let $x, y \in \mathbb{R}^{1,n}$ be two positive time-like vectors. Then:*

$$x \circ y \leqslant \|x\| \, \|y\| \, ,$$

with equality if and only if x and y are linearly dependent.

Proof. Set $t := |\, \|x\| \,| > 0$ ($\|x\|^2 = -t^2$) and because $x \in \mathrm{span}(x)$ is a one-dimensional time-like vector subspace of $\mathbb{R}^{1,n}$, by theorem 11.3.4 there exists $\phi \in \mathrm{PO}(1, n)$ such that $\phi(\mathrm{span}(x)) = \mathrm{span}(e_1)$ and consequently we have $\phi(x) = te_1$. Set $z \equiv \begin{pmatrix} z_1 \\ \bar{z} \end{pmatrix} := \phi(y)$. Then,

$$\begin{aligned} \|x\|^2 \, \|y\|^2 &= \|\phi(x)\|^2 \, \|\phi(y)\|^2 = -t^2(-z_1^2 + |\bar{z}|) = t^2 z_1^2 - t^2 |\bar{z}| \\ &\leqslant t^2 z_1^2 = (te_1 \circ z)^2 = (\phi(x) \circ \phi(y))^2 = (x \circ y)^2, \end{aligned}$$

thus $\|x\|^2 \, \|y\|^2 \leqslant (x \circ y)^2$. Notice that the equality $\|x\|^2 \, \|y\|^2 = (x \circ y)^2$ holds if and only if $\bar{z} = 0$, which implies $\phi(y) \in \mathrm{span}(e_1)$ and, since the action of $\mathrm{PO}(1, n)$ on time-like vector subspaces is stable, $y \in \mathrm{span}(x)$, i.e. x and y are linearly dependent.

Finally, theorem 11.2.1 guarantees that $x \circ y < 0$, hence $(\|x\| \, \|y\|)^2 = \|x\|^2 \, \|y\|^2 \leqslant (x \circ y)^2$ is an inequality between two negative real numbers, which implies

$$x \circ y \leqslant \|x\| \, \|y\|$$

since the function $\xi \mapsto \xi^2$ is decreasing, and thus order-reversing, in $(-\infty, 0]$. $\square$

Taking into account that for all $x, y \in \mathcal{H}^n$ we have $\|x\|^2 = \|y\|^2 = -1$, the proof of the previous theorem implies directly the following result.

Corollary 13.1.1 *Let $x, y \in \mathcal{H}^n$. Then:*

$$x \circ y \leqslant -1,$$

with equality if and only if $x = y$.

Recall now that $\cosh(\alpha) \geqslant 1$ for all $\alpha \in \mathbb{R}$ and that $\cosh(-\alpha) = \cosh(\alpha)$, thus we can consider just positive entries $\alpha \geqslant 0$ and formulate the following corollary.

Corollary 13.1.2 *Let $x, y \in \mathbb{R}^{1,n}$ be two positive time-like vectors, then there exists a unique $\alpha(x, y) \geqslant 0$ such that*

$$x \circ y = \cosh(\alpha(x, y)) \, \|x\| \, \|y\| \, . \tag{13.4}$$

In particular, if $x, y \in \mathcal{H}^n$, then $\|x\| \|y\| = -1$ and so it exists only one $\alpha(x, y) \geqslant 0$ such that $\cosh(\alpha(x, y)) = -x \circ y$.

Following the lead given to us by spherical geometry, we introduce the hyperbolic distance on $\mathcal{H}^n$ as follows.

Def. 13.1.3 (Hyperbolic distance on $\mathcal{H}^n$) *The hyperbolic distance between two elements x, y of $\mathcal{H}^n$ is $d_{\mathcal{H}}(x, y) = \alpha(x, y)$, where $\alpha(x, y) \geqslant 0$ is the only non-negative real number that satisfies the equation:*

$$\cosh(d_{\mathcal{H}}(x, y)) = -x \circ y, \tag{13.5}$$

or, equivalently,

$$\boxed{d_{\mathcal{H}}(x, y) = \operatorname{arcosh}(-(x \circ y))}\,. \tag{13.6}$$

The non-negative real number $\alpha(x, y)$ is called the ***Lorentzian time-like angle*** *between $x, y \in \mathcal{H}^n$.*

A transformation $T : \mathcal{H}^n \to \mathcal{H}^n$ is a ***hyperbolic isometry*** *on $\mathcal{H}^n$ if it verifies the following condition:*

$$d_{\mathcal{H}}(T(x), T(y)) = d_{\mathcal{H}}(x, y), \qquad \forall x, y \in \mathcal{H}^n. \tag{13.7}$$

The set of hyperbolic isometries on $\mathcal{H}^n$ is denoted with $\mathcal{I}(\mathcal{H}^n)$.

By (13.3), we have that the spherical distance is $d_S(x, y) = \arccos(\langle x, y\rangle)$, thus, apart from the minus sign in front of the Lorentz pseudo-scalar product, the only change that is required to pass from the spherical to the hyperbolic distance on $\mathcal{H}^n$ is to replace the inverse circular function arccos with the inverse hyperbolic function arcosh.

Clearly, $d_{\mathcal{H}}$ is positive, symmetric and $d_{\mathcal{H}}(x, y) = 0$ if and only if $x = y$ by corollary 13.1.1. All that is left to prove to verify that $d_{\mathcal{H}}$ is actually a distance is the triangular inequality, which is far from being trivial.

The proof of the triangular inequality of $d_{\mathcal{H}}$ needs a result that is important by its own: the possibility to identify the isometries of $\mathcal{H}^n$ with positive Lorentz transformations. The proof of this result requires the following lemma, which is proven with a technical reasoning of vast applicability that we will encounter again in this chapter.

Lemma 13.1.1 *A generic transformation $S : \mathcal{H}^n \to \mathcal{H}^n$ that preserves the Lorentz pseudo-scalar product can be extended to a positive Lorentz transformation $\phi_S \in \mathrm{PO}(1, n)$ if and only if there exists a transformation $T : \mathcal{H}^n \to \mathcal{H}^n$ that preserves the Lorentz pseudo-scalar product and that has an arbitrary fixed point $h \in \mathcal{H}^n$, i.e. $T(h) = h$, which can be extended to a positive Lorentz transformation $\phi_T \in \mathrm{PO}(1, n)$.*

Proof. If a generic transformation $S : \mathcal{H}^n \to \mathcal{H}^n$ that preserves the Lorentz pseudo-scalar product can be extended to a positive Lorentz transformation ϕ_S, then this property is also shared by a map T of this kind that also has the additional property of having a fixed point $h \in \mathcal{H}^n$. So, the non-trivial part of the proof consists in showing that the opposite is true.

To this end, write $S(h) = x \in \mathcal{H}^n$ and recall that $\mathrm{PO}(1, n)$ is transitive, in particular, on the set of 1-dimensional time-like vector subspaces of $\mathbb{R}^{1,n}$, so it surely exists $\tilde{R} \in \mathrm{PO}(1, n)$ such that $\tilde{R}(x) = h$. Since both x and h belong to $\mathcal{H}^n$, we can consider $R := \tilde{R}\Big|_{\mathcal{H}^n}$ and compute $(R \circ S)(h) = R(x) = h$, which shows that h is a fixed point for $T := R \circ S$, which surely preserves the Lorentz pseudo-scalar product since it is the composition of two functions that share this property.

Notice now that, since $\mathrm{PO}(1, n)$ is a group, it exists a transformation $\tilde{R}^{-1} \in \mathrm{PO}(1, n)$ such that the restriction $R^{-1} := \tilde{R}^{-1}\Big|_{\mathcal{H}^n}$ satisfies the equation $S = R^{-1} \circ T$.

Finally, if T can be extended to a positive Lorentz transformation, i.e. if there exists $\phi_T \in \mathrm{PO}(1,n)$ such that $T = \phi_T|_{\mathcal{H}^n}$, then $S = R^{-1} \circ T = \tilde{R}^{-1}\Big|_{\mathcal{H}^n} \circ \phi_T|_{\mathcal{H}^n} = (\tilde{R}^{-1} \circ \phi_T)\Big|_{\mathcal{H}^n}$, hence we recognize the extension of S to $\mathrm{PO}(1,n)$ to be $\phi_S := \tilde{R}^{-1} \circ \phi_T$. □

Theorem 13.1.2 *Every hyperbolic isometry on $\mathcal{H}^n$ can be extended to be a positive Lorentz transformation and every positive Lorentz transformation is a hyperbolic isometry on $\mathcal{H}^n$. Thus, we have the identification:*

$$\mathrm{PO}(1,n) \cong \mathcal{I}(\mathcal{H}^n).$$

Proof. If $\phi \in \mathrm{PO}(1,n-1)$, then $\phi : \mathcal{H}^n \to \mathcal{H}^n$ and, by definition (13.5), we have:

$$\cosh(d_{\mathcal{H}}(x,y)) = -x \circ y = -(\phi(x) \circ \phi(y)) = \cosh(d_{\mathcal{H}}(\phi(x),\phi(y))), \qquad \forall x,y \in \mathcal{H}^n,$$

but cosh is injective on $\mathbb{R}^+$, so $d_{\mathcal{H}}(\phi(x),\phi(y)) = d_{\mathcal{H}}(x,y)$ for all $x,y \in \mathcal{H}^n$.

Conversely, let $T : \mathcal{H}^n \to \mathcal{H}^n$ be a hyperbolic isometry, $T \equiv (T_1,\dots,T_{n+1})$, $T_j : \mathcal{H}^n \to \mathbb{R}$ being the j-th component function of T, i.e for $j \in \{1,\dots,n+1\}$,

$$\begin{array}{cccc} T: & \mathcal{H}^n & \longrightarrow & \mathcal{H}^n \\ & x & \longmapsto & \begin{pmatrix} T_1(x) \\ T_2(x) \\ \vdots \\ T_{n+1}(x) \end{pmatrix} \equiv \begin{pmatrix} T_1(x) \\ \overline{T}(x) \end{pmatrix} \end{array} .$$

We must prove that there exists $\phi \in \mathrm{PO}(1,n)$ such that $T = \phi|_{\mathcal{H}^n}$. Since ϕ preserves the Lorentz pseudo-scalar product on $\mathbb{R}^{1,n}$, for this problem to be well-posed, we must first check if T preserves the Lorentz pseudo-scalar product on $\mathcal{H}^n$. In order to do that we use eq. (13.5) and the fact that T preserves the hyperbolic distance to write, for all $x,y \in \mathcal{H}^n$,

$$\begin{aligned} d_{\mathcal{H}}(T(x),T(y)) = d_{\mathcal{H}}(x,y) \quad &\iff \quad \cosh(d_{\mathcal{H}}(T(x),T(y))) = \cosh(d_{\mathcal{H}}(x,y)) \\ &\iff \quad T(x) \circ T(y) = x \circ y. \end{aligned}$$

Having proven that a hyperbolic isometry T preserves the Lorentz pseudo-scalar product on $\mathcal{H}^n$ has another important consequence, i.e. the possibility to invoke lemma 13.1.1: if we solve our problem w.r.t.just one hyperbolic isometry $T : \mathcal{H}^n \to \mathcal{H}^n$ with a fixed point, then we automatically solve it for all the other hyperbolic isometries of $\mathcal{H}^n$.

A particularly clever choice of such a fixed point is represented by e_1, that clearly belongs to $\mathcal{H}^n$. The reason underlying this choice can be understood by recalling that the matrix

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}, \quad A \in \mathrm{O}(n), \tag{13.8}$$

is a positive Lorentzian matrix thanks to corollary 11.3.3. Since a Lorentzian matrix is associated to a Lorentz transformation w.r.t. the canonical basis $(e_1,\dots,e_n)$, the fact that the first column of Λ coincides is $(1,0,\dots,0)^t$ means that e_1 is a fixed point for the transformation.

As a consequence, the only thing that remains to do in order to prove the theorem is to use the properties of T to build a suitable orthogonal matrix such that expression in (13.8) extends T from $\mathcal{H}^n$ to the whole $\mathbb{R}^{1,n}$.

We start by observing that $T(e_1) = e_1$ implies, for all $u \in \mathcal{H}^n$:

$$T(u) \circ T(e_1) = T(u) \circ e_1 = -T_1(u) + 0 + \dots 0 = -T_1(u),$$

on the other side, since the Lorentz pseudo-scalar product is preserved by T, we have:

$$T(u) \circ T(e_1) = u \circ e_1 = -u_1 + 0 + \dots 0 = -u_1,$$

so $T_1(u) = u_1$ for all $u \in \mathcal{H}^n$.

Consider now $x, y \in \mathcal{H}^n$ and recall that $x \circ y = -x_1 y_1 + \langle \bar{x}, \bar{y} \rangle$, $\bar{x} = (x_2, \dots, x_{n+1})^t$, $\bar{y} = (y_2, \dots, y_{n+1})^t$, so

$$x \circ y = T(x) \circ T(y) \iff -\cancel{x_1 y_1} + \langle \bar{x}, \bar{y} \rangle = -\cancel{T(x_1) T(y_1)} + \left\langle \overline{T}(x), \overline{T}(y) \right\rangle,$$

where the first terms in each member of the second equality above cancel out because, as we have just proven, $T_1(x) = x_1$ and $T_1(y) = y_1$. So,

$$x \circ y = T(x) \circ T(y) \iff \langle \bar{x}, \bar{y} \rangle = \left\langle \overline{T}(x), \overline{T}(y) \right\rangle,$$

notice that this property is not yet enough to say that $\overline{T} = (T_2, \dots, T_{n+1})^t$ is an orthogonal transformation on $\mathbb{R}^n$, that we could associated to the $\mathrm{O}(n)$ matrix that we are searching for, because up to now we have shown that $\overline{T}$ preserves the Euclidean scalar product only when we apply it to the vectors $\bar{x}$ and $\bar{y}$, which were obtained by extracting the last n components of $x, y \in \mathcal{H}^n$. The extension to $\mathbb{R}^n$ can be achieved by considering the following bijection:

$$\begin{array}{rccc} p: & \mathcal{H}^n & \xrightarrow{\sim} & \mathbb{R}^n \\ & u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n+1} \end{pmatrix} & \longmapsto & p(u) := \begin{pmatrix} u_2 \\ \vdots \\ u_{n+1} \end{pmatrix}, \end{array}$$

which allows us to build the function $\tilde{T} := \overline{T} \circ p^{-1}$, explicitly

$$\begin{array}{rccl} \tilde{T}: & \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ & u & \longmapsto & (T_2(p^{-1}(u)), \dots, T_{n+1}(p^{-1}(u)))^t. \end{array}$$

$\tilde{T}$ is an orthogonal transformation on $\mathbb{R}^n$, lemma 11.1.2 guarantees that $\tilde{T}$ is linear and, by denoting with $A_{\tilde{T}}$ the associated matrix w.r.t. the canonical basis of $\mathbb{R}^n$, we have that

$$\begin{pmatrix} 1 & 0 \\ 0 & A_{\tilde{T}} \end{pmatrix}$$

is the Lorentzian matrix of a transformation $\phi \in \mathrm{PO}(1, n)$ such that $\phi|_{\mathcal{H}^n} = T$. □

We now start the proof of the triangular inequality for $d_{\mathcal{H}}$: a fundamental tool for this proof is given by the so-called Lorentzian cross product, which is the hyperbolic variant of the classical cross (or vector) product $x \times y$ between two vectors x, y in $\mathbb{R}^3$.

Recall that the $x \times y$ is a vector *orthogonal to the plane that contains x and y*, i.e. $\langle x, x \times y \rangle = \langle y, x \times y \rangle = 0$ and defined as follows:

$$x \times y := \det \begin{pmatrix} e_1 & e_2 & e_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} = (x_2 y_3 - x_3 y_2) e_1 - (x_1 y_3 - x_3 y_1) e_2 + (x_1 y_2 - x_2 y_1) e_3 = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}.$$

Def. 13.1.4 *Let $x, y \in \mathbb{R}^{1,2}$. The* ***Lorentzian cross-product*** *is defined as*

$$x \otimes y := \eta(x \times y) = \begin{pmatrix} -x_2y_3 + x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}.$$

Remark 13.1.1 *The Lorentzian cross product of x and y in $\mathbb{R}^{1,2}$ is Lorentz-orthogonal to both x and y:*

$$x \circ (x \otimes y) = x \circ \eta(x \times y) = \langle x, \eta(\eta(x \times y))\rangle = \langle x, x \times y\rangle = 0,$$

and analogously for y. Then, if x or y belong to $\mathcal{H}^2$, their Lorentzian cross product is space-like.

The proof of the following result can be obtained by direct computation.

Theorem 13.1.3 *For all $x, y, w, z \in \mathbb{R}^{1,2}$ we have:*

1. $x \otimes y = -y \otimes x$, *'antisymmetry';*
2. $(x \otimes y) \circ z = \det \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix}$, *'Lorentz mixed product formula' ;*
3. $x \otimes (y \otimes z) = (x \circ y)z - (x \circ z)y$;
4. $(x \otimes y) \circ (z \otimes w) = \det \begin{pmatrix} x \circ w & x \circ z \\ y \circ w & y \circ z \end{pmatrix}$, *'Lorentz version of Lagrange identity' .*
5. $x \otimes y = -\eta(x) \times \eta(y) = \eta(y) \times \eta(x)$.

Corollary 13.1.3 *For all $x, y \in \mathbb{R}^{1,2}$ we have:*

$$\|x \otimes y\|^2 = (x \circ y)^2 - \|x\|^2 \|y\|^2 .$$

Proof. By using property 4. of theorem 13.1.3 we get:

$$\|x \otimes y\|^2 = (x \otimes y) \circ (x \otimes y) = \det \begin{pmatrix} x \circ y & x \circ x \\ y \circ y & y \circ x \end{pmatrix} = (x \circ y)^2 - \|x\|^2 \|y\|^2 .$$

□

The three statements that follow are direct consequences of the previous corollary.

Corollary 13.1.4 *If $x, y \in \mathbb{R}^{1,2}$ are space-like, then*

1. $|x \circ y| < \|x\| \|y\| \iff x \otimes y$ *is time-like;*
2. $|x \circ y| = \|x\| \|y\| \iff x \otimes y$ *is light-like;*

3. $|x \circ y| > \|x\| \, \|y\| \iff x \otimes y$ *is space-like.*

Corollary 13.1.5 *Let $x, y \in \mathbb{R}^{1,2}$ be two linearly independent, positively-oriented, time-like vectors. Then, $x \otimes y$ is space-like and*

$$\|x \otimes y\| = -\|x\| \, \|y\| \sinh(\alpha(x,y)),$$

where $\alpha(x,y)$ is the Lorentzian time-like angle between x and y.

In particular, if $x, y \in \mathcal{H}^n$, then $\|x \otimes y\| = \sinh(\alpha(x,y))$.

Proof. Thanks to corollary 13.1.3 we have:

$$\begin{aligned}\|x \otimes y\|^2 &= (x \circ y)^2 - \|x\|^2 \|y\|^2 \underset{(13.4)}{=} \|x\|^2 \|y\|^2 \cosh^2(\alpha(x,y)) - \|x\|^2 \|y\|^2 \\ &= \|x\|^2 \|y\|^2 (\cosh^2(\alpha(x,y)) - 1) = \|x\|^2 \|y\|^2 \sinh^2(\alpha(x,y)).\end{aligned}$$

From remark 13.1.1, $x \otimes y$ is Lorentz-orthogonal to x which is time-like so $x \otimes y$ must space-like by corollary 11.2.2. The space-likeness of $x \otimes y$ implies $\|x \otimes y\| > 0$, so:

$$\|x \otimes y\| = -\|x\| \, \|y\| \sinh(\alpha(x,y)).$$

□

We are now ready to prove the triangular inequality of the hyperbolic distance on $\mathcal{H}^n$. As we have said, we will have to use the properties of the Lorentz cross-product, however, since it is defined only on $\mathbb{R}^{1,2}$, it seems not appropriate to use this operation to prove a property of $d_{\mathcal{H}}$ on $\mathcal{H}^n$ for n different than 3.

In fact, as we will see in the proof below, the clever idea that will allow us to circumvent this problem consists in the very simple observation that *only three vectors are involved in the triangular inequality* so, proving the triangular inequality of $d_{\mathcal{H}}$ in the 3-dimensional vector subspace generated by those three vectors or with $\mathbb{R}^{1,2}$ will be enough to infer the same property of $d_{\mathcal{H}}$ on $\mathcal{H}^n$ thanks to the transitivity of $\mathrm{PO}(1,n)$ on time-like vector subspaces and to the isometric nature of positive Lorentz transformations.

Theorem 13.1.4 *The hyperbolic distance $d_{\mathcal{H}}$ is a metric on $\mathcal{H}^n$.*

Proof. As previously said, only the triangular inequality for $d_{\mathcal{H}}$ remains to be proven. Let $\tilde{x}, \tilde{y}, \tilde{z} \in \mathcal{H}^n$ distinct and $\tilde{V} = \mathrm{span}(\tilde{x}, \tilde{y}, \tilde{z})$. Thanks to theorem 11.3.4, it exists $\phi \in \mathrm{PO}(1,n)$ such that $\phi(\tilde{V}) = \mathrm{span}(e_1, e_2, e_3) \cong \mathbb{R}^{1,2}$. We set $x = \phi(\tilde{x})$, $y = \phi(\tilde{y})$ and $z = \phi(\tilde{z})$.

As proven in theorem 13.1.2, positive Lorentz transformations preserve the hyperbolic distance, thus proving the triangular inequality for $x, y, z \in \mathbb{R}^{1,2}$ is equivalent to prove it for the vectors $\tilde{x}, \tilde{y}, \tilde{z}$. To this aim, let us use corollary 13.1.5 to write

$$\|x \otimes y\| = \sinh(d_{\mathcal{H}}(x,y)) \qquad \text{and} \qquad \|y \otimes z\| = \sinh(d_{\mathcal{H}}(y,z)), \tag{13.9}$$

then, by property 3. of theorem 13.1.3, we have

$$(x \otimes y) \otimes (y \otimes z) = -\underbrace{((x \otimes y) \circ y)}_{=0} z - ((x \otimes y) \circ z) y = -((x \otimes y) \circ z) y, \tag{13.10}$$

$-((x \otimes y) \circ z) \in \mathbb{R}$, thus $(x \otimes y) \otimes (y \otimes z)$ and y are linearly dependent, so $(x \otimes y) \otimes (y \otimes z)$ is either time-like or it is the zero vector. Corollary 13.1.4 implies the following inequality

$$|(x \otimes y) \circ (y \otimes z)| \leqslant \|x \otimes y\| \, \|y \otimes z\| \,. \tag{13.11}$$

Finally, we recall the formula $\cosh(a+b) = \cosh a \cosh b + \sinh a \sinh b$ for all $a, b \in \mathbb{R}$. We have gathered all the information that we need to prove the triangular inequality for x, y, z:

$$\begin{aligned}
\cosh(d_{\mathcal{H}}(x,y) + d_{\mathcal{H}}(y,z)) &= \cosh(d_{\mathcal{H}}(x,y)) \cosh(d_{\mathcal{H}}(y,z)) + \sinh(d_{\mathcal{H}}(x,y)) \sinh(d_{\mathcal{H}}(y,z)) \\
&\underset{(13.5) \text{ and } (13.9)}{=} (x \circ y)(y \circ z) + \|x \otimes y\| \, \|y \otimes z\| \\
&\underset{(13.11)}{\geqslant} (x \circ y)(y \circ z) + |(x \otimes y) \circ (y \otimes z)| \\
&\geqslant (x \circ y)(y \circ z) + (x \otimes y) \circ (y \otimes z) \\
&\underset{(4. \text{ of th. } 13.1.3)}{\geqslant} \cancel{(x \circ y)(y \circ z)} + (x \circ z)(y \circ y) - \cancel{(x \circ y)(y \circ z)} \\
&= (x \circ z)\|y\|^2 = -(x \circ z) \\
&\underset{(13.5)}{=} \cosh(d_{\mathcal{H}}(x,z)),
\end{aligned}$$

i.e. $\cosh(d_{\mathcal{H}}(x,y) + d_{\mathcal{H}}(y,z)) \geqslant \cosh(d_{\mathcal{H}}(x,z))$, but cosh is a strictly increasing function on $\mathbb{R}^+$, so it preserves the order and we can write $d_{\mathcal{H}}(x,y) + d_{\mathcal{H}}(y,z) \geqslant d_{\mathcal{H}}(x,z)$, which is the triangular inequality that we wanted to prove. □

Def. 13.1.5 *The metric space $(\mathcal{H}^n, d_{\mathcal{H}})$ is called the **hyperbolic n-space**.*

In the geometry of the sphere, the geodesic lines are given by the intersection of the sphere S^n with a 2-dimensional vector subspace of $\mathbb{R}^{n+1}$ (and thus results in circles). Once again the hyperboloid model has very similar features as those of spherical geometry.

Def. 13.1.6 *A **hyperbolic line** in $\mathcal{H}^n$ is the intersection of $\mathcal{H}^n$ with a 2-dimensional time-like vector subspace of $\mathbb{R}^{1,n}$.*

Since a 2-dimensional time-like vector subspace of $\mathbb{R}^{1,n}$ must pass through the origin, its intersection with $\mathcal{H}^n$ will always be a hyperbola, so, in turn, **a hyperbolic line in $\mathcal{H}^n$ is just a hyperbola**.

Lemma 13.1.2 *Two distinct elements x, y of the hyperboloid $\mathcal{H}^n$ are linearly independent and so they span a 2-dimensional time-like vector subspace of $\mathbb{R}^{1,n}$.*

Proof. By absurd, let $x, y \in \mathcal{H}^n$, $x \neq y$, be linearly dependent, then $y = \lambda x$, $\lambda \in \mathbb{R} \backslash \{1\}$, then $\|x\|^2 = -1 = \|y\|^2 = \|\lambda x\|^2 = \lambda^2 \|x\|^2$, $\lambda = -1$. However, λ cannot be -1 because otherwise y would not belong to $\mathcal{H}^n$ anymore since its first coordinate would be negative. □

Remark 13.1.2 Given two distinct $x, y \in \mathcal{H}^n$, we have

$$\ell_{x,y} := \mathcal{H}^n \cap \operatorname{span}(x, y)$$

that is the unique hyperbolic line of $\mathcal{H}^n$ that contains both x and y.

We will show that these hyperbolic lines are the 'straight lines' of the hyperbolic metric, i.e. the curves that minimize the hyperbolic distance between two points.

Def. 13.1.7 *Three points x, y and z of $\mathcal{H}^n$ are said to be hyperbolically collinear if there is a hyperbolic line ℓ passing through x, y and z.*

Lemma 13.1.3 *If $x, y, z \in \mathcal{H}^n$ are such that*

$$d_{\mathcal{H}}(x, z) = d_{\mathcal{H}}(x, y) + d_{\mathcal{H}}(y, z),$$

then x, y and z are hyperbolically collinear.

Proof. As shown in the proof of the triangular inequality for $d_{\mathcal{H}}$, it is possible to consider the Lorentzian cross product of vectors belonging to $\mathcal{H}^n$ by associating them to vectors belonging to $\mathbb{R}^{1,2}$, in what follows this assumption will be implicitly assumed.

Let $x, y, z \in \mathcal{H}^n$ verify the equality $d_{\mathcal{H}}(x, z) = d_{\mathcal{H}}(x, y) + d_{\mathcal{H}}(y, z)$ and apply cosh to both members, then, using the already quoted property $\cosh(a + b) = \cosh a \cosh b + \sinh a \sinh b$ for all $a, b \in \mathbb{R}$, we get:

$$\begin{aligned}
\cosh(d_{\mathcal{H}}(x, z)) &= \cosh(d_{\mathcal{H}}(x, y) + d_{\mathcal{H}}(y, z)) \\
&= \cosh(d_{\mathcal{H}}(x, y)) \cosh(d_{\mathcal{H}}(y, z)) + \sinh(d_{\mathcal{H}}(x, y)) \sinh(d_{\mathcal{H}}(y, z)) \\
&= (-x \circ y)(-y \circ z) + \|x \otimes y\| \, \|y \otimes z\| \\
&= (x \circ y)(y \circ z) + \|x \otimes y\| \, \|y \otimes z\| ,
\end{aligned}$$

but $\cosh(d_{\mathcal{H}}(x, z)) = -x \circ z$, so

$$-x \circ z - (x \circ y)(y \circ z) = \|x \otimes y\| \, \|y \otimes z\| .$$

We can interpret the left-hand side of the previous equality as the following determinant:

$$\det \begin{pmatrix} x \circ z & x \circ y \\ y \circ z & y \circ y \end{pmatrix} = (x \circ z) \, \|y\|^2 - (x \circ y)(y \circ z) = -x \circ z - (x \circ y)(y \circ z),$$

but, thanks to property 4. of theorem 13.1.3, we have

$$\det \begin{pmatrix} x \circ z & x \circ y \\ y \circ z & y \circ y \end{pmatrix} = (x \otimes y) \circ (y \otimes z),$$

which implies

$$(x \otimes y) \circ (y \otimes z) = \|x \otimes y\| \, \|y \otimes z\| .$$

Thanks to remark 13.1.1, $x \otimes y$ and $y \otimes z$ are space-like vectors, thus their norm is positive, so $(x \otimes y) \circ (y \otimes z) = |(x \otimes y) \circ (y \otimes z)|$ and we can write:

$$|(x \otimes y) \circ (y \otimes z)| = \|x \otimes y\| \, \|y \otimes z\| .$$

Property 2. of corollary 13.1.4 implies that $(x \otimes y) \otimes (y \otimes z)$ is light-like, moreover, thanks to eq. (13.10),

$$(x \otimes y) \otimes (y \otimes z) = -((x \otimes y) \circ z) y,$$

but $-((x \otimes y) \circ z) \in \mathbb{R}$ and y is time-like, hence $(x \otimes y) \otimes (y \otimes z)$ is a light-like vector collinear with a time-like vector, which is possible if and only if $(x \otimes y) \otimes (y \otimes z) = 0$, i.e. $((x \otimes y) \circ z)y = 0$, but $y \in \mathcal{H}^n$, so the only possibility that remains is that $(x \otimes y) \circ z = 0$.

Finally, by property 4. of theorem 13.1.3 we have:

$$\det \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} = ((x \otimes y) \circ z) = 0$$

and so x, y, z are linearly dependent, so each vector belong to $\mathcal{H}^n$ and to the span of the other two vectors, thus, by definition, x, y, z are hyperbolically collinear. □

In order to prove that hyperbolic lines minimize the hyperbolic distance, we start with the definition and analysis of hyperbolic geodesic arcs.

Def. 13.1.8 (Geodesic arc) *A geodesic arc in a generic metric space (X, d) is a distance preserving function $\gamma : [a, b] \subseteq \mathbb{R} \to X$, with $a < b$.*

Explicitly, this means that $\forall t, s \in [a, b]$, $s \leqslant t$, we have: $d(\gamma(s), \gamma(t)) = d(s, t)$, but $d(s, t) = t - s$, so the request for a geodesic arc can be explicitly restated as follows:

$$d(\gamma(s), \gamma(t)) = t - s, \qquad \forall t, s \in [a, b],\ s \leqslant t.$$

Def. 13.1.9 (Hyperbolic geodesic arc) *A geodesic arc in the metric space $(\mathcal{H}^n, d_{\mathcal{H}})$ is called a hyperbolic geodesic arc.*

Theorem 13.1.5 *Let $\gamma : [a, b] \to \mathcal{H}^n$ be a curve. The following statements are equivalent.*

1. *The curve γ is a hyperbolic geodesic arc.*
2. *There exist Lorentz-orthonormal vectors $x, y \in \mathbb{R}^{1,n}$ such that*
$$\gamma(t) = \cosh(t-a)x + \sinh(t-a)y. \tag{13.12}$$
3. *The curve satisfies the differential equation $\gamma'' - \gamma = 0$.*

Proof.

$\boxed{1 \implies 2}$: we assume γ to be a geodesic arc on $(\mathcal{H}^n, d_{\mathcal{H}})$. Then for all $t \in [a, b]$, we have

$$\begin{aligned} d_{\mathcal{H}}(\gamma(a), \gamma(b)) &= b - a = t - a + b - t \\ &= d_{\mathcal{H}}(\gamma(a), \gamma(t)) + d_{\mathcal{H}}(\gamma(t), \gamma(b)), \end{aligned} \tag{13.13}$$

which, by lemma 13.1.3, shows that $\gamma(t), \gamma(a)$ and $\gamma(b)$ are hyperbolically collinear for all $t \in [a, b]$, i.e $\gamma(t) \in \operatorname{span}(\gamma(a), \gamma(b))$, and so

$$\gamma([a, b]) \subset \ell_{\gamma(a), \gamma(b)}.$$

Since the image of γ belongs to $\mathcal{H}^n$, $\operatorname{span}((\gamma(a), \gamma(b))$ is a time-like vector subspace of $\mathbb{R}^{1,n}$, so, thanks to the transitivity of $\mathrm{PO}(1, n)$ on the set of time-like vector subspaces of $\mathbb{R}^{1,n}$, there exists $\phi \in \mathrm{PO}(1, n)$ such that $\phi(\operatorname{span}(\gamma(a), \gamma(b))) = \operatorname{span}(e_1, e_2) \cong \mathbb{R}^{1,1}$ and $\phi(\gamma(a)) = e_1$.

For all $t \in [a,b]$, let $z_t := \phi(\gamma(t))$. To obtain eq. (13.12), the decomposition of z_t on the $\mathbb{R}^{1,1}$ orthonormal basis $(e_1, \epsilon e_2)$, where $\epsilon = \pm 1$, will prove very helpful. In fact, if we write

$$z_t = \langle z_t, e_1 \rangle e_1 + \langle z_t, e_2 \rangle e_2 \tag{13.14}$$

and we apply ϕ^{-1} to both members we obtain

$$\phi^{-1}(z_t) = \langle z_t, e_1 \rangle \phi^{-1}(e_1) + \langle z_t, e_2 \rangle \phi^{-1}(e_2),$$

having used the linearity of Lorentz transformations. By definition, the last equation can be re-written as follows:

$$\gamma(t) = \langle z_t, e_1 \rangle \gamma(a) + \langle z_t, e_2 \rangle \phi^{-1}(e_2),$$

notice now that $x := \gamma(a) \in \mathcal{H}^n$ is a time-like vector and $y := \phi^{-1}(e_2)$ is a space-like vector because e_2 is space-like Lorentz transformations do not modify the likeness of vectors. Thus, x and y are Lorentz-orthogonal, plus, since ϕ^{-1} preserves the Lorentz norm, and $\|x\| = \left\|\phi^{-1}(e_1)\right\| = \|e_1\| = 1$ and $\|y\| = \left\|\phi^{-1}(e_2)\right\| = \|e_2\| = 1$, hence x and y are Lorentz-orthonormal vectors.

It follows that the only thing that remains to do is to prove that $\langle z_t, e_1 \rangle = \cosh(t-a)$ and $\langle z_t, e_2 \rangle = \sinh(t-a)$. Let us start with the first coefficient:

$$\begin{aligned}
\langle z_t, e_1 \rangle &= -z_t \circ e_1 = -\phi(\gamma(t)) \circ \phi(a) \\
&\underset{(\phi \in \mathrm{PO}(1,n))}{=} -\gamma(t) \circ \gamma(a) = \cosh(d_{\mathcal{H}}(\gamma(t), \gamma(a))) \\
&= \cosh(t-a).
\end{aligned}$$

Now, regarding the second coefficient, we remark that z_t belongs to $\mathcal{H}^n$, so, using the decomposition in eq. (13.14), we must have

$$-\langle z_t, e_1 \rangle^2 + \langle z_t, e_2 \rangle^2 = -1 \iff -\cosh^2(t-a) + \langle z_t, e_2 \rangle^2 = -1,$$

which implies that $\langle z_t, e_2 \rangle = \pm \sinh(t-a)$. By choosing the positive determination, we get precisely formula (13.12).

$\boxed{2 \implies 1}$: we assume $\gamma : [a,b] \to \mathcal{H}^n$ is such that there is $x, y \in \mathbb{R}^{1,n}$, Lorentz-orthonormal, that verify $\gamma(t) = \cosh(t-a)x + \sinh(t-a)y$ $\forall t \in [a,b]$. This means that $\gamma(t) \in \operatorname{span}(x,y)$, and so (x,y) is a Lorentz-orthonormal basis for this vector subspace. We recall that, by definition of Lorentz-orthonormal basis, x is time-like and y is space-like.

Now, given $s, t \in [a,b]$, $s \leqslant t$, we have

$$\begin{aligned}
\cosh(d_{\mathcal{H}}(\gamma(s), \gamma(t))) &= -\gamma(s) \circ \gamma(t) \\
&= -(\cosh(s-a)x + \sinh(s-a)y) \circ (\cosh(t-a)x + \sinh(t-a)y) \\
&\text{(by Lorentz-orthogonality of } x \text{ and } y) \\
&= -(\cosh(t-a)\cosh(s-a) \underbrace{\|x\|^2}_{=-1} + \sinh(t-a)\sinh(s-a) \underbrace{\|y\|^2}_{=1}) \\
&= \cosh(t-a)\cosh(s-a) - \sinh(t-a)\sinh(s-a) \\
&= \cosh((t-a)-(s-a)) \\
&= \cosh(t-s),
\end{aligned}$$

thus, since $\cosh(\xi)$ is injective for $\xi \geqslant 0$, $d_{\mathcal{H}}(\gamma(s), \gamma(t)) = t - s$ and so γ is a geodesic arc.

$\boxed{2 \implies 3}$: if $\gamma(t) = \cosh(t-a)x + \sinh(t-a)y$ $\forall t \in [a,b]$, then

$$\left.\begin{aligned} \cosh''(t-a) &= \cosh(t-a) \\ \sinh''(t-a) &= \sinh(t-a) \end{aligned}\right\} \implies \gamma''(t) - \gamma(t) = 0, \quad \forall t \in [a,b].$$

$\boxed{3 \implies 2}$: suppose $\gamma''(t) - \gamma(t) = 0$ $\forall t \in [a,b]$. From ODE calculus, we know that the general solution of the previous differential equation is:

$$\gamma(t) = \cosh(t-a)\gamma(a) + \sinh(t-a)\gamma'(a). \tag{13.15}$$

Thus, proving 2. comes down to proving that $\gamma(a)$ and $\gamma'(a)$ are Lorentz-orthonormal.

To this aim, we notice that, since $\gamma(t) \in \mathcal{H}^n$ for all $t \in [a,b]$, $\gamma(t) \circ \gamma(t) = -1$ for all $t \in [a,b]$, so $\gamma \circ \gamma : [a,b] \to \mathbb{R}$ is the constant function $t \mapsto -1$, thus $(\gamma \circ \gamma)'(t) = 0$. On the other side, by applying the Leibniz rule on the Lorentz pseudo-scalar product we get

$$(\gamma \circ \gamma)'(t) = \gamma'(t) \circ \gamma(t) + \gamma(t) \circ \gamma'(t) = 2\,\gamma(t) \circ \gamma'(t) \qquad \forall t \in [a,b],$$

where, in the last step, we have used the symmetry of $\circ$. By mixing these results we find $\gamma(t) \circ \gamma'(t) = 0$ for all $t \in [a,b]$, hence, in particular, $\gamma(a)$ and $\gamma'(a)$ are Lorentz-orthogonal. Moreover, using (13.15), for all $t \in [a,b]$ we have,

$$\begin{aligned} \|\gamma(t)\|^2 &= -1 = \gamma(t) \circ \gamma(t) \\ &= (\cosh(t-a)\gamma(a) + \sinh(t-a)\gamma'(a)) \circ (\cosh(t-a)\gamma(a) + \sinh(t-a)\gamma'(a)) \\ &= \cosh^2(t-a)\underbrace{\|\gamma(a)\|^2}_{=-1} + \sinh^2(t-a)\left\|\gamma'(a)\right\|^2 = -\cosh^2(t-a) + \sinh^2(t-a)\left\|\gamma'(a)\right\|^2, \end{aligned}$$

where the terms proportional to $\gamma(a) \circ \gamma'(a)$ have not been written because of the Lorentz-orthogonality between $\gamma(a)$ and $\gamma'(a)$. We conclude that

$$-\cosh^2(t-a) + \sinh^2(t-a)\left\|\gamma'(a)\right\|^2 = -1 \qquad \forall t \in [a,b],$$

which implies $\|\gamma'(a)\|^2 = 1$, i.e. $\gamma(a)$ and $\gamma'(a)$ are Lorentz-orthonormal. □

Remark 13.1.3 In the theory of dynamical systems, the differential equation satisfied by a hyperbolic geodesic arc, i.e. $\gamma'' - \gamma = 0$, is that of the harmonic repulsor, whose phase portrait is known to be given by hyperbolae. Instead, the differential equation satisfied by a spherical geodesic arc, i.e. $\gamma'' + \gamma = 0$, is that of the harmonic oscillator, whose phase portrait is represented by circles.

When we extend the arc parameterization interval $[a,b]$ to the whole $\mathbb{R}$, we say that a geodesic arc γ is a **geodesic line**.

Corollary 13.1.6 *A function $\gamma : \mathbb{R} \to \mathcal{H}^n$ is a hyperbolic geodesic line if and only if there are $x, y \in \mathbb{R}^{1,n}$ Lorentz-orthonormal such that*

$$\begin{aligned} \gamma(t) &= \cosh(t)x + \sinh(t)y \\ &= \cosh(t)\gamma(0) + \sinh(t)\gamma'(0). \end{aligned}$$

Corollary 13.1.7 *The hyperbolic geodesic lines of $\mathcal{H}^n$ are its hyperbolic lines.*

Proof. Let $x, y \in \mathcal{H}^n$, $x \neq y$, and $\ell_{x,y} = \mathcal{H}^n \cap \text{span}(x, y)$ be a hyperbolic line passing trough x and y, which defines, geometrically, a hyperbola connecting the points on the hyperboloid $\mathcal{H}^n$ identified by the vectors x and y. Thanks to the transitivity of $\text{PO}(1, n)$ on $\mathcal{V}_m^T$, the set of all time-like m-dimensional vector subspaces of $\mathbb{R}^{1,n}$, $m \leqslant n$, there is a $\phi \in \text{PO}(1, n)$ such that

$$\phi(\text{span}(x, y)) = \text{span}(e_1, e_2) \simeq \mathbb{R}^{1,1}.$$

Then, if we apply ϕ to $\ell_{x,y}$ we transform the hyperbola connecting x and y on $\mathcal{H}^n$ to a rectangular hyperbola on $\mathbb{R}^{1,1}$ relative to the canonical basis (e_1, e_2). We use $\mathcal{H}^1$ to denote this object, which is well-known to be parameterized by the hyperbolic functions as follows:

$$\phi(\ell_{x,y}) = \mathcal{H}^1 = \left\{\gamma(t) = \cosh(t)e_1 + \sinh(t)e_2,\ t \in \mathbb{R}\right\}$$

and so, thanks to the linearity of ϕ, we get

$$\ell_{x,y} = \left\{ \cosh(t)\phi^{-1}(e_1) + \sinh(t)\phi^{-1}(e_2),\ t \in \mathbb{R}\right\}.$$

Since ϕ preserves the likeness, the orientation and the norm of vectors, we have that $\ell_{x,y}$ is written as in formula (13.12), thus it is a hyperbolic geodesic line. □

Def. 13.1.10 *A metric space X is* ***geodesically complete*** *if each geodesic arc $\gamma : [a, b] \to X$ extends to a unique geodesic line $\lambda : \mathbb{R} \to X$.*

The previous results show us that each hyperbolic geodesic arc extends uniquely to a hyperbolic geodesic line, i.e. it can be seen as a piece of an infinite hyperbola, thus $\mathcal{H}^n$ **is geodesically complete**.

The final result that we discuss is the equivalence between the hyperbolic topology on $\mathcal{H}^n$ generated by $d_{\mathcal{H}}$ and the Euclidean topology on $\mathcal{H}^n$ inherited by $\mathbb{R}^{n+1}$ with the Euclidean distance d_E. In the proof of this result we will use the Taylor-MacLaurin series expansion for cosh:

$$\cosh(x) = 1 + \frac{x^2}{2} + \frac{x^4}{24} + \cdots = \sum_{k=0}^{m} \frac{x^{2k}}{(2k)!} + \text{O}(x^{2m+1}) \tag{13.16}$$

Theorem 13.1.6 *The metric topology on $\mathcal{H}^n$ given by $d_{\mathcal{H}}$ is equivalent to d_E.*

Proof. For all $x \in \mathcal{H}^n$ and $r > 0$, let us define the open neighborhoods of radius r around x w.r.t. the Euclidean and the hyperbolic distance, respectively, as follows:

$$B_E(x, r) := \{y \in \mathcal{H}^n \ :\ d_E(x, y) < r\}, \quad B_H(x, r) := \{y \in \mathcal{H}^n \ :\ d_{\mathcal{H}}(x, y) < r\}.$$

If we prove that $B_E(x, r) \subseteq B_H(x, r)$ and that $B_H(x, r) \subseteq B_E(x, r)$ for all $x \in \mathcal{H}^n$ and $r > 0$, then the theorem will be proven.

$\boxed{B_E(x, r) \subseteq B_H(x, r)}$: consider $x, y \in \mathcal{H}^n$ distinct, then, since $|x-y|^2 = (x-y)_1^2 + \cdots + (x-y)_n^2$ and $\|x - y\|^2 = -(x - y)_1^2 + \cdots + (x - y)_n^2$, we have $d_E(x, y)^2 = |x - y|^2 > \|x - y\|^2$, moreover,

$\|x-y\|^2 = (x-y) \circ (x-y) = \|x\|^2 - 2x \circ y + \|y\|^2 = -2x \circ y - 2 = 2\cosh(d_{\mathcal{H}}(x,y)) - 2 = 2(\cosh(d_{\mathcal{H}}(x,y)) - 1)$, i.e.

$$\|x-y\|^2 = 2(\cosh(d_{\mathcal{H}}(x,y)) - 1), \tag{13.17}$$

so

$$d_E(x,y)^2 > 2(\cosh(d_{\mathcal{H}}(x,y)) - 1) \underset{(13.16)}{>} 2\left(1 + \frac{d_{\mathcal{H}}(x,y)^2}{2} - 1\right) = d_{\mathcal{H}}(x,y)^2, \tag{13.18}$$

so, by positivity, $d_E(x,y) > d_{\mathcal{H}}(x,y)$. Let now $y \in B_E(x,r)$, then $d_{\mathcal{H}}(x,y) < d_E(x,y) < r$, so $y \in B_H(x,r)$ too, thus $B_E(x,r) \subseteq B_H(x,r)$ for all $x \in \mathcal{H}^n$ and all $r > 0$.

$\boxed{B_H(x,r) \subseteq B_E(x,r)}$: we start by noticing that, thanks to corollary 13.1.6, once we fix an arbitrary $x \in \mathcal{H}^n$, all the hyperbolic lines passing through x are parameterized by a unit space-like vector y Lorentz-orthogonal to x, i.e.

$$L_x := \{\ell_{x,z} = \mathrm{span}(x,z) \cap \mathcal{H}^n,\ z \in \mathcal{H}^n \backslash \{x\}\} \cong \{y \in \mathrm{span}(x)^L,\ \|y\|^2 = 1\} =: S_x^L.$$

Again corollary 13.1.6 tells us that the hyperbolic geodesic line associated to $y \in S_x^L$ is

$$\gamma_y(t) = \cosh(t)x + \sinh(t)y \qquad t \in \mathbb{R}.$$

γ_y is clearly continuous in the Euclidean topology on the whole $\mathbb{R}$, in particular, the continuity in $t = 0$ can be explicitly written as follows:

$$\forall \varepsilon > 0\ \exists \delta_y(\varepsilon) > 0\ :\ |t| < \delta_y(\varepsilon) \implies d_E(\gamma_y(0), \gamma_y(t)) < \varepsilon, \tag{13.19}$$

having interpreted the images of γ_y as points in $(\mathbb{R}^{n+1}, d_E)$. The key observation that let us introduce the hyperbolic distance in our reasoning is that, by definition of hyperbolic geodesic,

$$d_{\mathcal{H}}(x, \gamma_y(t)) = d_{\mathcal{H}}(\gamma_y(0), \gamma_y(t)) = |t - 0| = |t| \qquad t \in \mathbb{R},$$

so that expression (13.19) can be replaced by

$$\forall \varepsilon > 0\ \exists \delta_y(\varepsilon) > 0\ :\ d_{\mathcal{H}}(x, \gamma_y(t)) < \delta_y(\varepsilon) \implies d_E(\gamma_y(0), \gamma_y(t)) < \varepsilon$$

or, equivalently,

$$d_{\mathcal{H}}(x,z) < \delta_y(\varepsilon) \implies d_E(x,z) < \varepsilon \qquad z \in \gamma_y(\mathbb{R}). \tag{13.20}$$

Moreover, by the transitivity of $\mathrm{PO}(1,n)$ on time-like vector subspaces, there exists $\phi \in \mathrm{PO}(1,n)$ such that $\phi(x) = e_1$ and so we have

$$\phi(\mathrm{span}(x)^L) = \phi(\mathrm{span}(e_1)^L) = \mathrm{span}(e_2, \dots, e_{n+1}) \cong \mathbb{R}^n.$$

Now recall that S_x^L is the set of space-like vectors y belonging to $\mathrm{span}(x)^L$ such that $\|y\|^2 = 1$, but the Lorentz norm of a space-like vector is positive, so also $\|y\| = 1$. Since ϕ preserves the Lorentz norm, $\phi(S_x^L)$ is the set of vectors belonging to $\mathrm{span}(e_2, \dots, e_{n+1}) \cong \mathbb{R}^n$ with unit Lorentz norm, however, the Lorentz and the Euclidean norms coincide on $\mathrm{span}(e_2, \dots, e_{n+1})$, so $\phi(S_x^L) = \{y \in \mathbb{R}^n\ :\ |y| = 1\} \equiv S^{n-1}$, which is compact. By the continuity of ϕ^{-1}, we get that also $S_x^L = \phi^{-1}(S^{n-1})$ is compact. The compactness of S_x^L allows us to set

$$\delta(\varepsilon) := \inf_{y \in S_x^L} \{\delta_y(\varepsilon)\} > 0$$

which allows us to get rid of the dependence on y in the implication (13.20) and to write

$$d_{\mathcal{H}}(x,z) < \delta(\varepsilon) \implies d_E(x,z) < \varepsilon \qquad \forall \varepsilon > 0,$$

i.e. for all radius $\varepsilon > 0,\ z \in \gamma_y(\mathbb{R})$, it exists a radius $\delta(\varepsilon) > 0$ such that $B_H(x,\delta(\varepsilon)) \subseteq B_E(x,\varepsilon)$ which concludes the proof. □

13.1.3 The hyperbolic arc length

Let $\gamma : [a,b] \to \mathcal{H}^n$ be a curve. In this section we shall discuss how the metric given by the hyperbolic distance on $\mathcal{H}^n$ can be extended to compute the arc length. For that, we need to recall that a partition $P = \{t_0, \ldots, t_m\}$ of $[a,b]$ is an ordered finite set such that

$$a = t_0 < t_1 < \cdots < t_m = b.$$

We set a partial ordering on partitions

$$Q \leqslant P \iff P \subseteq Q$$

and set

$$|P| = \inf_{k \in \{1,\ldots,m\}} |t_k - t_{k-1}|,$$

$|P|$ is the finest partition interval. Notice that $|P| \to 0$ can be interpreted as 'P converges to $[a,b]$' and that if $Q \leqslant P$, then $|Q| \leqslant |P|$.

Def. 13.1.11 *Let $\gamma : [a,b] \to \mathcal{H}^n$ be a curve and let $P = \{t_0, \ldots, t_m\}$ be a partition of $[a,b]$. We define the* ***hyperbolic P-inscribed length*** *of γ as:*

$$L_H(\gamma, P) = \sum_{i=1}^{m} d_{\mathcal{H}}(\gamma(t_i), \gamma(t_{i-1})).$$

Moreover, the curve γ is ***rectifiable*** *if there is a real number $L(\gamma)$ such that for all $\varepsilon > 0$, there is a partition P_ε of $[a,b]$ such that for any partition Q verifying $Q \leqslant P_\varepsilon$, then*

$$|L(\gamma) - L_H(\gamma, Q)| < \varepsilon.$$

Lemma 13.1.4 *If γ is rectifiable, then for any partition P of $[a,b]$,*

$$L_H(\gamma, P) \leqslant L(\gamma).$$

Proof. Let P be a partition of $[a,b]$. First we note that if Q is a partition of $[a,b]$ such that $Q \leqslant P$, then

$$L(\gamma, P) \leqslant L(\gamma, Q) \tag{13.21}$$

by the triangular inequality of $d_{\mathcal{H}}$. Since γ is rectifiable, for any ε there is a P_ε such that for all $Q \leqslant P_\varepsilon$,

$$|L(\gamma) - L_H(\gamma, Q)| < \varepsilon.$$

Moreover, $Q_\epsilon := P_\varepsilon \cup P$ is such that $Q_\varepsilon \leqslant P_\varepsilon$ and $Q_\varepsilon \leqslant P$. Finally, for all $\varepsilon > 0$ we have

$$\begin{aligned} L_H(\gamma, P) - L(\gamma) &\underset{(13.21)}{\leqslant} L_H(\gamma, Q_\varepsilon) - L(\gamma) \\ &\leqslant \lim_{\varepsilon \to 0} L_H(\gamma, Q_\varepsilon) - L(\gamma) = 0, \end{aligned}$$

hence

$$L_H(\gamma, P) \leqslant L(\gamma).$$

□

Def. 13.1.12 *Let $x, y \in \mathcal{H}^n$. We define the Lorentzian distance as*

$$d_L(x, y) = \|x - y\|.$$

Lemma 13.1.5 *The Lorentzian distance d_L verifies the following properties.*

1. *$d_L(x, y) \geqslant 0$ with equality if and only if $x = y$*
2. *$d_L(x, y) = d_L(y, x)$*

Proof. Let $x, y \in \mathcal{H}^n$. Then,

$$\begin{aligned} \|x - y\|^2 &= \|x\|^2 - 2(x \circ y) + \|y\|^2 \\ &\underset{(13.1.1)}{\geqslant} -2 - 2\underbrace{\|x\|\,\|y\|}_{=-1} = 0, \end{aligned}$$

with equality if and only if they are linearly dependent, which implies $x = y$ since x, y belong to $\mathcal{H}^n$. □

Remark 13.1.4 *The Lorentzian distance is not a metric since it does not verify the triangular inequality. In fact, if we take $x, y, z \in \mathcal{H}^n$ hyperbolically collinear and such that y is between x and z, then it can be proven that*

$$d_L(x, z) > d_L(x, y) + d_L(y, z).$$

While the Lorentzian distance is not a metric, it will be useful because it can approximate the hyperbolic metric locally. To show this, we use formula (13.17) to write

$$\|x - y\|^2 = 2(\cosh(d_{\mathcal{H}}(x, y)) - 1) \underset{y \to x}{\sim} 2\left(1 - \frac{d_{\mathcal{H}}(x, y)^2}{2} - 1\right)^2 = d_{\mathcal{H}}(x, y)^2$$

and so, by positivity,

$$d_L(x, y) \underset{y \to x}{\sim} d_{\mathcal{H}}(x, y).$$

Def. 13.1.13 *Let $\gamma : [a, b] \to \mathcal{H}^n$ be a curve and let $P = \{t_0, \ldots, t_m\}$ be a partition of $[a, b]$. We define the **Lorentzian P-inscribed length** of γ as:*

$$L_L(\gamma, P) = \sum_{i=1}^{m} \|\gamma(t_i) - \gamma(t_{i-1})\|.$$

*Moreover, the curve γ is **Lorentz-rectifiable** if there is a real number $\mathcal{L}(\gamma)$ such that for all $\varepsilon > 0$, there is a partition P_ε of $[a, b]$ such that for any partition Q verifying $Q \leqslant P_\varepsilon$, then*

$$|\mathcal{L}(\gamma) - L_L(\gamma, Q)| < \varepsilon.$$

Since we do not have the triangular inequality for the Lorentzian distance, lemma 13.1.4 does not hold in the case of Lorentz-rectifiable curves.

Lemma 13.1.6 *If $\mathcal{L}(\gamma)$ exists, then it is unique.*

Proof. Assume γ is Lorentz-rectifiable with $\mathcal{L}_1(\gamma)$ and $\mathcal{L}_2(\gamma)$. If $\mathcal{L}_1(\gamma) \neq \mathcal{L}_1(\gamma)$, then there exists a $\varepsilon > 0$ such that $|\mathcal{L}_1(\gamma) - \mathcal{L}_2(\gamma)| > \varepsilon$. Let P, Q be partitions of $[a,b]$ such that

- $|\mathcal{L}_1(\gamma) - L_L(\gamma, P')| < \frac{\varepsilon}{2}$
- $|\mathcal{L}_2(\gamma) - L_L(\gamma, Q')| < \frac{\varepsilon}{2}$,

for all partitions P', Q' of $[a,b]$ verifying $P' \leqslant P$ and $Q' \leqslant Q$. The partition $R := P \cup Q$ is such that $R \leqslant P$ and $R \leqslant Q$, and we come to the following contradiction:

$$|\mathcal{L}_1(\gamma) - \mathcal{L}_2(\gamma)| \leqslant |\mathcal{L}_1(\gamma) - \mathcal{L}(\gamma, R)| + |\mathcal{L}_2(\gamma) - \mathcal{L}(\gamma, R)| < \varepsilon.$$

□

Def. 13.1.14 *Let $\gamma : [a,b] \to \mathcal{H}^n$ be a curve. We define the hyperbolic arc length as*

$$|\gamma|_H := \begin{cases} L(\gamma) & \text{if } \gamma \text{ is rectifiable} \\ \infty & \text{otherwise} \end{cases},$$

similarly, we define the Lorentzian length of γ as

$$\|\gamma\| = \begin{cases} \mathcal{L}(\gamma) & \text{if } \gamma \text{ is Lorentz-rectifiable} \\ \infty & \text{otherwise} \end{cases}.$$

Theorem 13.1.7 *Let $\gamma : [a,b] \to \mathcal{H}^n$ be a curve. Then γ is rectifiable in $\mathcal{H}^n$ if and only if γ Lorentz-rectifiable. Furthermore, the hyperbolic length is the same as the Lorentz length of γ, i.e.*

$$|\gamma|_H = \|\gamma\|.$$

Proof. We need to collect some preliminary results. Let $\eta > 0$, using the Taylor-MacLaurin series of cosh (13.16) we have

$$\begin{aligned}
\eta^2 \leqslant 2(\cosh(\eta) - 1) &= 2\left(1 + \frac{\eta^2}{2} + \frac{\eta^4}{24} + \sum_{k=3}^{m} \frac{\eta^{2k}}{(2k)!} + \mathrm{O}(x^{2m+1}) - 1\right) \\
&= 2\left(\frac{\eta^2}{2} + \frac{\eta^4}{24}\left(\sum_{k=0}^{m} 4!\frac{\eta^{2k}}{(2k+4)!} + \mathrm{O}(x^{2m+1})\right)\right) \\
&\leqslant 2\left(\frac{\eta^2}{2} + \frac{\eta^4}{24}\left(\sum_{k=0}^{m} \frac{\eta^{2k}}{(2k)!} + \mathrm{O}(x^{2m+1})\right)\right) \\
&= \eta^2 + \frac{\eta^4}{12}\cosh(\eta).
\end{aligned}$$

Consequently, if $\cosh(\eta) \leqslant 12$,

$$2(\cosh(\eta) - 1) \leqslant \eta^2(1 + \eta^2). \tag{13.22}$$

If we replace η in eq. (13.22) with $d_{\mathcal{H}}(x,y)$ and we suppose that $d_{\mathcal{H}}(x,y) \leqslant \operatorname{arcosh}(12)$, then, since $d_L(x,y)^2 = 2(\cosh(d_{\mathcal{H}}(x,y)) - 1)$ by eq. (13.17), we get

$$d_L(x,y)^2 \leqslant d_{\mathcal{H}}(x,y)^2(1 + d_{\mathcal{H}}(x,y)^2) \iff d_L(x,y) \leqslant d_{\mathcal{H}}(x,y)\sqrt{1 + d_{\mathcal{H}}(x,y)^2}.$$

On the other side, eq. (13.18) implies that $d_{\mathcal{H}}(x,y)^2 \leqslant 2(\cosh(d_{\mathcal{H}}(x,y)) - 1) = d_L(x,y)^2$, hence $d_{\mathcal{H}}(x,y) \leqslant d_L(x,y)$ for all $x, y \in \mathcal{H}^n$ not necessarily distinct.

So, for all $x, y \in \mathcal{H}^n$ such that $d_{\mathcal{H}}(x,y) \leqslant \operatorname{arcosh}(12)$, it holds that

$$d_{\mathcal{H}}(x,y) \leqslant d_L(x,y) \leqslant d_{\mathcal{H}}(x,y)\sqrt{1 + (d_{\mathcal{H}}(x,y))^2}. \tag{13.23}$$

We can now start with the proof of the equivalence.

$\boxed{\Longrightarrow}$: we start by assuming that γ is rectifiable. Let $\varepsilon > 0$ and P a partition of $[a,b]$ such that for all $Q \leqslant P$ we have by lemma 13.1.4

$$|\gamma|_H - L_H(\gamma, P) < \epsilon.$$

Let $\delta > 0$ and set

$$\mu(\gamma,\delta) = \sup_{a \leqslant s < t \leqslant b} \big\{ d_{\mathcal{H}}(\gamma(s), \gamma(t)) : |t - s| < \delta \big\}.$$

Note that since $[a,b]$ is compact and γ is continuous, γ is uniformly continuous and so $\mu(\gamma,\delta) \underset{\delta \to 0}{\to} 0$. Let $\delta > 0$ such that $\cosh(\mu(\gamma,\delta)) \leqslant 12$ and

$$|\gamma|_H \sqrt{1 + \mu(\gamma,\delta)^2} < |\gamma|_H + \varepsilon,$$

and P' a partition of $[a,b]$ such that $P' \leqslant P$ and $|P| < \delta$. Then for all partitions $Q = \{t_0, \ldots, t_m\}$ of $[a,b]$ such that $Q \leqslant P'$ we have:

$$|\gamma|_H - \epsilon \leqslant L_H(\gamma, Q) \leqslant L_L(\gamma, Q)$$

on one side, and

$$\begin{aligned}
L_L(\gamma, Q) &= \sum_{i=1}^{m} \|\gamma(t_i) - \gamma(t_{i-1})\| \\
&\leqslant \sum_{i=1}^{m} d_{\mathcal{H}}(\gamma(t_i), \gamma(t_{i-1}))\sqrt{1 + d_{\mathcal{H}}^2(\gamma(t_i), \gamma(t_{i-1}))} \\
&\leqslant L_H(\gamma, Q)\sqrt{1 + \mu(\gamma,\delta)^2} \\
&\leqslant |\gamma|_H + \varepsilon.
\end{aligned}$$

Hence, by combining both inequalities,

$$\big| |\gamma|_H - L_L(\gamma, Q) \big| < \varepsilon, \qquad \forall Q \leqslant P'$$

and by the unicity of the Lorentzian arc length (lemma 13.1.6), γ is Lorentz-rectifiable and $|\gamma|_H = \|\gamma\|$.

$\boxed{\Longleftarrow}$: we now suppose γ is Lorentz-rectifiable. Let $\varepsilon > 0$ and P be a partition of $[a,b]$ such that if $Q \leqslant P$, $\big| \|\gamma\| - L_L(\gamma, Q) \big| < \varepsilon$. Then,

$$L_H(\gamma, Q) - \|\gamma\| \leqslant L_L(\gamma, Q) - \|\gamma\| \leqslant \varepsilon$$

for all $Q \leqslant P$, and so γ is rectifiable. $\square$

Before proving the theorem regarding the metric of the arc length, we first make the following remark: for any $\mathcal{C}^1$-curve $\gamma : [a, b] \to \mathcal{H}^n$, we have that $\gamma(t)$ is Lorentz-orthogonal to $\gamma'(t)$ for all $t \in [a, b]$, i.e.

$$\gamma(t) \circ \gamma'(t) = 0,$$

in fact, by differentiating the Lorentz pseudo-scalar product and using the Leibniz property together with the symmetry of $\circ$ we get:

$$\begin{aligned} (\gamma(t) \circ \gamma(t))'(t) &= 2(\gamma(t) \circ \gamma'(t)) \\ &= (t \mapsto \|\gamma(t)\|^2 \equiv -1)' = 0. \end{aligned}$$

Theorem 13.1.8 *Let $\gamma : [a, b] \to \mathcal{H}^n$ a $\mathcal{C}^1$-curve. Then γ is rectifiable and the hyperbolic length of γ is given by*

$$\|\gamma\|_H = \int_a^b \|\gamma'\| \, dt.$$

Proof. By hypothesis $x, y \in \mathcal{B}^n$ are supposed to be linearly dependent, i.e. they are collinear with Let $F : [a, b]^{n+1} \to \mathbb{R}$ defined by

$$F(t_1, \ldots, t_{n+1}) = \big(- \gamma_1'(t_1)^2 + \cdots + \gamma_{n+1}'(t_{n+1})^2 \big)^{\frac{1}{2}}.$$

Since γ is $\mathcal{C}^1$ and $\gamma'(t)$ is space-like for all t because it is Lorentz-orthogonal to $\gamma(t) \in \mathcal{H}^n$, F is continuous on $[a, b]^{n+1}$ which is compact, thus F is uniformly continuous. The set

$$\big\{ |F(t) - F(s)|, \ t, s \in [a, b]^{n+1} \big\}$$

is bounded since F is continuous on a compact set.

For any fixed $\delta > 0$ we define

$$\mu(F, \delta) = \sup_{t, s \in [a,b]^{n+1}} \big\{ |F(t) - F(s)|, \ |t_i - s_i| \leqslant \delta, \ i \in \{1, \ldots, n+1\} \big\}.$$

As in the previous proof, F is uniformly continuous so $\mu(F, \delta) \underset{\delta \to 0}{\to} 0$ and if we set $P = \{t_0, \ldots, t_m\}$ such that $|P| \leqslant \delta$, we have by the mean-value theorem $\exists s_{ij} \in [t_{j-1}, t_j]$

$$|\gamma_i(t_j) - \gamma_i(t_{j-1})| = \gamma_i'(sij)(t_j - t_{j-1})$$

and if we set $s_j = (s_{1,j}, \ldots, s_{n+1,j})$, then

$$\begin{aligned} \|\gamma_i(t_j) - \gamma_i(t_{j-1})\| &= \big(- [\gamma_1(t_j) - \gamma_1(t_{j-1})]^2 + [\gamma_2(t_j) - \gamma_2(t_{j-1})]^2 + \cdots + [\gamma_{n+1}(t_j) - \gamma_{n+1}(t_{j-1})]^2 \big)^{\frac{1}{2}} \\ &= \big(- \gamma_1'(s_{1,j})^2 + \gamma_1'(s_{2,j})^2 + \cdots + \gamma_{n+1}'(s_{n+1,j})^2 \big)^{\frac{1}{2}} (t_j - t_{j-1}) \\ &= F(s_j)(t_j - t_{j-1}). \end{aligned}$$

Additionally, we set

$$S(\gamma, P) = \sum_{j=1}^{m} \|\gamma'(t_j)\| \, (t_j - t_{j-1})$$

and we remind

$$L_L(\gamma, P) = \sum_{j=1}^{m} \|\gamma(t_j) - \gamma(t_{j-1})\| .$$

As such, we have

$$\begin{aligned} \big|S(\gamma, P) - L_L(\gamma, P)\big| &= \big|\sum_{j=1}^{m} \|\gamma'(t_j)\| (t_j - t_{j-1}) - F(s_j)(t_j - tj - 1)\big| \\ &\leqslant \mu(F, \delta)(b-a) \qquad\qquad (*_2) \end{aligned}$$

and furthermore,

$$\begin{aligned} \big|\int_a^b \|\gamma'(t)\| \, dt - S(\gamma, P)\big| &= \big|\sum_{i=1}^{m} \big(\int_{t_{j-1}}^{t_j} \|\gamma'(t)\| - \|\gamma'(t_j)\| (t_j - t_{j-1}) dt\big)\big| \\ &\leqslant \sum_{i=1}^{m} \int_a^b \underbrace{\big| \|\gamma'(t)\| - \|\gamma'(t_j)\| (t_j - t_{j-1})\big|}_{\leqslant \mu(F,\delta)} dt \\ &\leqslant \mu(F, \delta)(b-a) \qquad\qquad (*_2) \end{aligned}$$

Finally, by combining $(*_1)$and $(*_2)$ we obtain

$$\begin{aligned} \big|\int_a^b \|\gamma'(t)\| \, dt - L_L(\gamma, P)\big| &\leqslant \big|\int_a^b \|\gamma'(t)\| \, dt - S(\gamma, P)\big| + \big|S(\gamma, P) - L_L(\gamma, P)\big| \\ &\leqslant 2\mu(F, \delta) \end{aligned}$$

and since $\mu(F, \delta) \underset{\delta \to 0}{\to} 0$ and $|P| \underset{\delta \to 0}{\to} 0$

$$\|\gamma\| = \int_a^b \|\gamma'(t)\| \, dt = \lim_{|P| \to 0} L_L(\gamma, P)$$

$\square$

Def. 13.1.15 *Let $\gamma : [a, b] \to \mathcal{H}^n$ curve. If $d_x = (dx_1, , \ldots, dx_{n+1})$, then*

$$\|dx\| = (-dx_1^2 + dx_2^2 + \cdots + dx_{n+1}^2)^{\frac{1}{2}}$$

and

$$\int_\gamma \|dx\| := \|\gamma\| .$$

Additionally if γ is a $\mathcal{C}^1-$curve,

$$\|\gamma\| = \int_\gamma \|dx\| = \int_a^b \|\gamma'(t)\| \, dt.$$

*The differential $\|dx\|$ is called the **element of hyperbolic arc length** of $\mathcal{H}^n$.*

13.1.4 The hyperboloid as a Riemannian manifold

In this short subsection we will prove that the hyperboloid $\mathcal{H}^n$ can be considered a Riemannian manifold. To this aim, we recall that the Lorentz pseudo-scalar product is a bilinear, symmetric non-degenerate form. Hence, if we set

$$\begin{array}{rccc} f: & \mathbb{R}^{n+1} & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & (x \circ x), \end{array}$$

then differential of f is

$$df_x(y) = 2(x \circ y).$$

In fact,

$$f(x+y) = (x+y) \circ (x+y) = f(x) + \underbrace{2(x \circ y)}_{\text{linear}} + \underbrace{(y \circ y)}_{\text{quadratic}},$$

since the differential represents the unique linear approximation of f, the result follows.

Theorem 13.1.9 *The hyperboloid $\mathcal{F}^n = f^{-1}(\{-1\})$ is a Riemannian n-manifold.*

Proof. The differential $df_x(y) = 2(x \circ y)$ is surjective and -1 is not a critical value of f, so, by the level set theorem 1.2.1, the hyperboloid is a differential manifold of dimension n. Moreover, thanks to eq. (2.34), for every $x \in \mathcal{F}^n$, the tangent space $T_x\mathcal{F}^n$ is given by

$$T_x\mathcal{F}^n = \ker df_x = \operatorname{span}(x)^L.$$

Now, since x is time-like, $\operatorname{span}(x)^L$ is a n-dimensional space-like vector subspace of $\mathbb{R}^{n+1}$ and so, for all $y \in T_x\mathcal{F}^n$ we have $y \circ y > 0$. Hence, the Lorentz pseudo-scalar product is positive-definite on the tangent spaces $T_x\mathcal{F}^n$ and so $\mathcal{F}^n$ is a Riemannian manifold. □

Finally, $\mathcal{H}^n$ can be defined as the biggest subset of F^n that contains e_1 and that is simply connected and so $\mathcal{H}^n$ is a complete, simply connected Riemannian manifold of dimension n with metric tensor $g_x(u, v) = u \circ v$.

13.2 The conformal hyperbolic models $\mathcal{B}^n$ and $\mathcal{U}^n$

There are essentially two so-called conformal models for hyperbolic geometry: the open unit ball $\mathcal{B}^n$ and the upper-half space $\mathcal{U}^n$. We have already mentioned them and their strict interdependence in section 12.5, of chapter 12, hence it is clear that Möbius transformations will play a key role in the analysis of these models. In this section we will introduce the **metric structure** that makes them hyperbolic spaces. Moreover we will stress their link with the hyperboloid model $\mathcal{H}^n$.

As we have already seen in subsection 12.4.3 the term conformal means angle-preserving. In this section we will see that $\mathcal{B}^n$ and $\mathcal{U}^n$, as hyperbolic spaces, are called **conformal models because their isometry group is given by conformal transformations**[1].

13.2.1 Link between $\mathcal{B}^n$ and $\mathcal{H}^n$

The unit ball $\mathcal{B}^n$ is classically defined as an open subset of $\mathbb{R}^n$ as follows:

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

Now we are going to show its link with the hyperboloid $\mathcal{H}^n$. Since $\mathcal{H}^n$ is contained in $\mathbb{R}^{n+1}$, it will be convenient to represent $\mathcal{B}^n$ embedded as well in $\mathbb{R}^{n+1}$ as follows:

$$\mathcal{B}^n = \{x \in \mathbb{R}^{n+1} : |\bar{x}| < 1 \text{ and } x_{n+1} = 0\} \subset \mathbb{R}^{n+1},$$

where $\bar{x} = (x_1, \ldots, x_n)^t$ for all $x \in \mathbb{R}^{n+1}$. Our aim is to define an invertible function ζ between $\mathcal{B}^n$ and $\mathcal{H}^n$ to eventually be able to transfer the metric structure of $\mathcal{H}^n$ onto $\mathcal{B}^n$. Let us recall that one of the possible definitions (see 11.2.1) of the Lorentzian scalar product in $\mathbb{R}^{n+1}$ is:

$$x \circ y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n - x_{n+1} y_{n+1}.$$

$\|x\|$ will denote the Lorentz pseudo-norm w.r.t. this Lorentzian pseudo-scalar product and, as usual, $|x|$ will stand for the Euclidean norm of $x \in \mathbb{R}^{n+1}$. Note that for all $x \in \mathcal{B}^n$ it holds that $|x| = \|x\|$. Moreover, the hyperboloid $\mathcal{H}^n$, as a hypersurface of $\mathbb{R}^{n+1}$, is described as follows:

$$\mathcal{H}^n = \{x \in \mathbb{R}^{n+1} : \|x\|^2 = -1 \text{ and } x_{n+1} > 0\}. \tag{13.24}$$

The construction of the map $\zeta : \mathcal{B}^n \to \mathcal{H}^n$ is very similar to the stereographic projection seen in section 12.3 with theorem 12.3.1. We are going to define ζ as a projection from the point $-e_{n+1}$ of $\mathbb{R}^{n+1}$ onto $\mathcal{H}^n$, in the same way as the stereographic projection is a projection from the north pole of S^n onto $\hat{\mathbb{R}}^n$ (both embedded in $\mathbb{R}^{n+1}$).

In figure 13.2 we depict the construction of ζ in the case $n = 1$, where $\mathcal{H}^1$ is a hyperbola and $\mathcal{B}^1$ is the open segment $(-1, 1)$ both embedded in the real plane $\mathbb{R}^2$.

Given an arbitrary $x \in \mathcal{B}^n$, if we consider the straight line $r(s) = -e_{n+1} + s(x + e_{n+1})$, then $r(0) = -e_{n+1}$ and $r(1) = x$. $\zeta(x)$ is the point of this straight line corresponding to the

[1] By theorem 12.4.9 we know that, for $n \geqslant 3$ the set of all (regular) conformal transformations that are stable on $\mathcal{B}^n$ (or $\mathcal{U}^n$, respectively) coincides with $\mathcal{M}(\mathcal{B}^n)$ ($\mathcal{M}(\mathcal{U}^n)$, respectively), while for $n = 2$ the set of all conformal transformations (preserving $\mathcal{B}^n$ or $\mathcal{U}^n$) is the set of all the holomorphic and anti-holomorphic transformations of $\mathbb{R}^2$ (stable on $\mathcal{B}^n$ or $\mathcal{U}^n$).

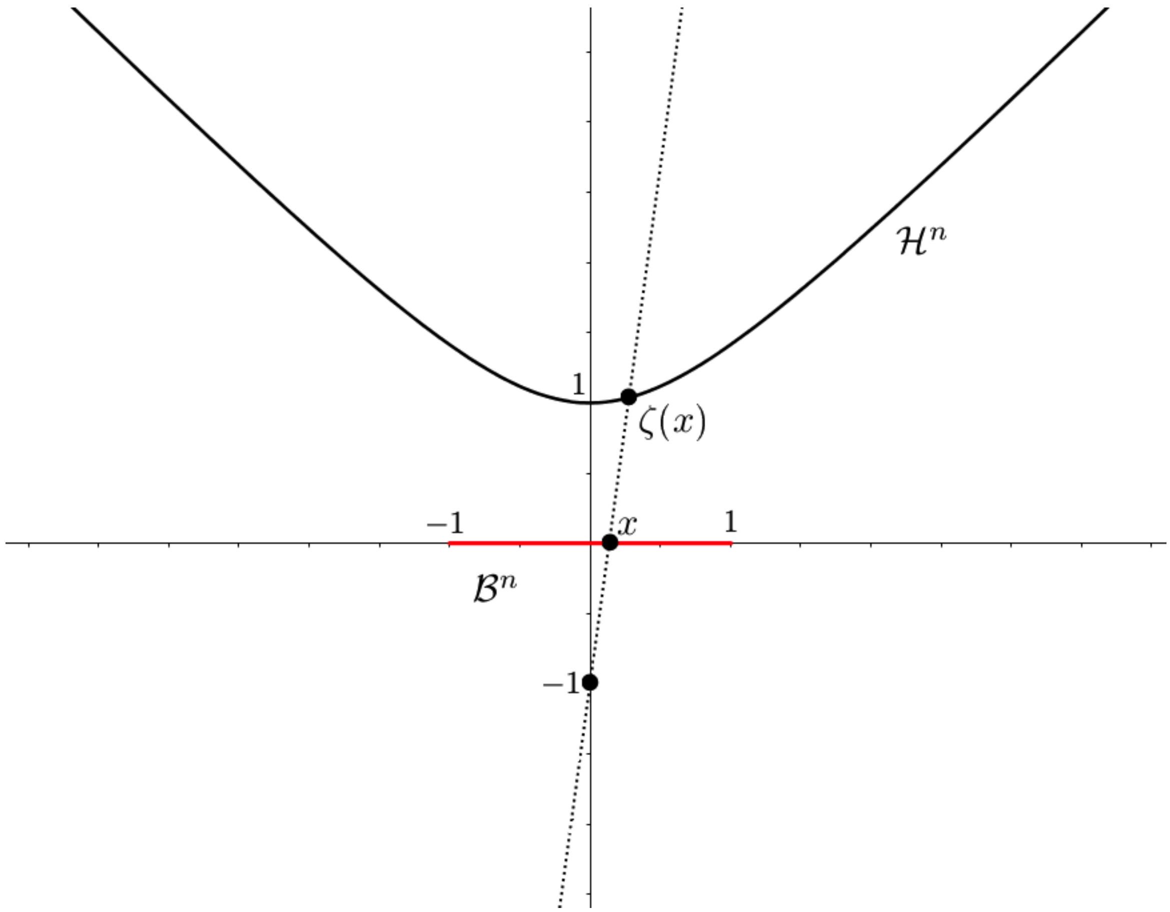


Figure 13.2: Graphic representation of the action of ζ in the case $n = 1$.

unique value $\bar{s}$ such that $r(\bar{s}) \in \mathcal{H}^n$. In order to obtain this value, let us notice that

$$r(s) = -e_{n+1} + s(x + e_{n+1}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix} + s \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{pmatrix} = \begin{pmatrix} sx_1 \\ \vdots \\ sx_n \\ s-1 \end{pmatrix},$$

so $r(s) \in \mathcal{H}^n$ if and only is $||r(s)|| = -1$ and $s - 1 > 0$ (see def. 13.24), but:

$$\begin{aligned} ||r(s)|| = -1 \quad &\iff \quad s^2x_1^2 + \cdots + s^2x_n^2 - (s-1)^2 = -1 \\ &\iff \quad s^2|x|^2 - s^2 + 2s - 1 = -1 \\ &\iff \quad s(1 - |x|^2) = 2 \end{aligned}$$

i.e. $\bar{s} = \frac{2}{1-|x|^2}$ which satisfies $\bar{s} - 1 > 0$ because $|x| < 1$. So, the explicit expression for ζ is:

$$\begin{array}{rcccl} \zeta : \ \mathcal{B}^n & \xrightarrow{\sim} & \mathcal{H}^n & & \\ x & \longmapsto & \zeta(x) & = & -e_{n+1} + \frac{2}{1-|x|^2}(x + e_{n+1}) \\ & & & = & \left(\frac{2x_1}{1-|x|^2}, \ldots, \frac{2x_n}{1-|x|^2}, \frac{1+|x|^2}{1-|x|^2}\right). \end{array}$$

The action of ζ for $n = 2$ is shown in figure 13.3.

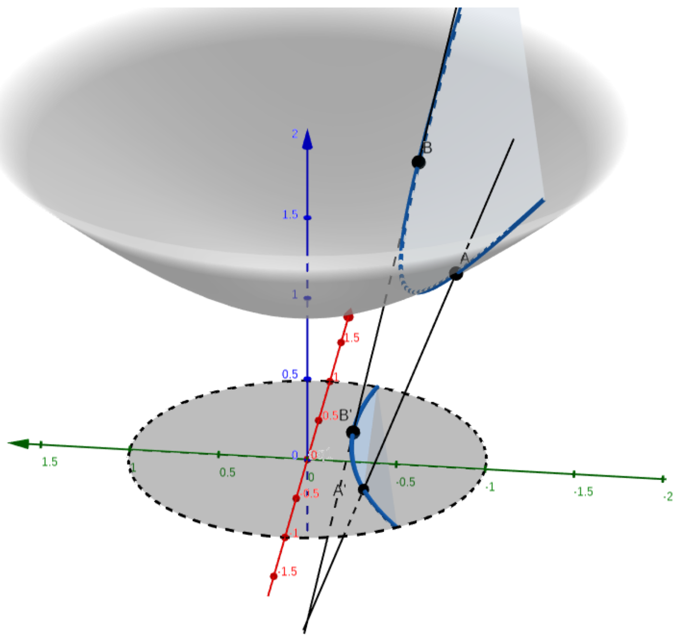


Figure 13.3: Illustration of the action of ζ between $\mathcal{H}^n$ and $\mathcal{B}^n$ when $n = 2$.

Proposition 13.2.1 *The function $\zeta : \mathcal{B}^n \to \mathcal{H}^n$ that we have just defined is invertible with inverse:*

$$\begin{array}{rcccl} \zeta^{-1}: & \mathcal{H}^n & \xrightarrow{\sim} & \mathcal{B}^n & \\ & y & \longmapsto & \zeta^{-1}(y) & = -e_{n+1} + \frac{1}{1+y_{n+1}}(y + e_{n+1}) \\ & & & & = \left(\frac{y_1}{1+y_{n+1}}, \dots, \frac{y_n}{1+y_{n+1}}, 0\right) \\ & & & & \cong \left(\frac{y_1}{1+y_{n+1}}, \dots, \frac{y_n}{1+y_{n+1}}\right). \end{array}$$

Proof. First of all we are going to construct ζ^{-1} in the same way as we did for ζ. We will eventually prove that it is injective and the right inverse of ζ (i.e. ζ is surjective) concluding that ζ is invertible with inverse ζ^{-1}.
For all $y \in \mathcal{H}^n$ we define $\zeta^{-1}(y)$ as the intersection with $\mathcal{B}^n$ of the straight line joining y and $-e_{n+1}$, as depicted in figure 13.4.

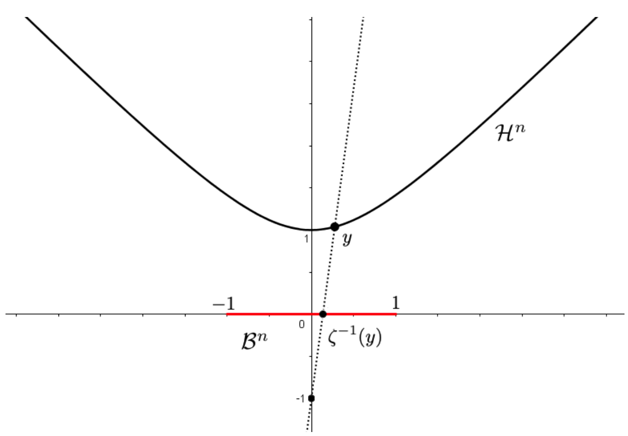


Figure 13.4: Graphic representation of the action of ζ^{-1} for $n = 1$.

Thence $\zeta^{-1}(y)$ must be of the form

$$\zeta^{-1}(y) = -e_{n+1} + s(y + e_{n+1}).$$

We just need to find the parameter s such that $\zeta^{-1}(y)$ belongs to $\mathcal{B}^n$, i.e. $\langle \zeta^{-1}(y), e_{n+1}\rangle = 0$ and $|\zeta^{-1}(y)| < 1$. The first condition implies that $-1 + sy_{n+1} + s = 0$, hence $s = \frac{1}{1+y_{n+1}}$. The second condition is verified after a straightforward computation, keeping in mind that $y \in \mathcal{H}^n$, i.e. $y_1^2 + \cdots + y_n^2 = y_{n+1}^2 - 1$. So the candidate inverse of ζ has the following expression:

$$\zeta^{-1}(y) = -e_{n+1} + \frac{1}{1+y_{n+1}}(y+e_{n+1}) = \left(\frac{y_1}{1+y_{n+1}}, \ldots, \frac{y_n}{1+y_{n+1}}, 0\right) \cong \left(\frac{y_1}{1+y_{n+1}}, \ldots, \frac{y_n}{1+y_{n+1}}\right).$$

Let us now show the injectivity of ζ. Let $x, y \in \mathcal{B}^n$ such that $\zeta(x) = \zeta(y)$, then

$$\left(\frac{2x_1}{1-|x|^2}, \ldots, \frac{2x_n}{1-|x|^2}, \frac{1+|x|^2}{1-|x|^2}\right) = \left(\frac{2y_1}{1-|y|^2}, \ldots, \frac{2y_n}{1-|y|^2}, \frac{1+|y|^2}{1-|y|^2}\right),$$

it is convenient to start observing that the equality of the last components implies $\frac{1+|x|^2}{1-|x|^2} = \frac{1+|y|^2}{1-|y|^2}$, so $|x| = |y|$. Hence, $\frac{2x_i}{1-|x|^2} = \frac{2y_i}{1-|y|^2}$ for all $i = 1, \ldots, n$, so $x_i = y_i$. Since $x, y \in \mathcal{B}^n$ we have that $x_{n+1} = y_{n+1} = 0$. Hence $x = y$, thus ζ is injective.
Let us show the surjectivity of ζ. To do that we will show that ζ^{-1} is its right inverse, i.e. $\zeta(\zeta^{-1}(y)) = y$, for all $y \in \mathcal{H}^n$. Before developing the computations note that, since $y \in \mathcal{H}^n$, $\|y\|^2 = -1$ i.e. $|y|^2 - y_{n+1}^2 = -1$, or $|y|^2 = y_{n+1}^2 - 1$, so

$$|\zeta^{-1}(y)|^2 = ||\zeta^{-1}(y)||^2 \underset{\text{(direct computation)}}{=} \frac{|\bar{y}|^2}{(y_{n+1}+1)^2} = \frac{y_{n+1}^2 - 1}{(y_{n+1}+1)^2} = \frac{y_{n+1}-1}{y_{n+1}+1}, \tag{13.25}$$

and so

$$1 - |\zeta^{-1}(y)|^2 = \frac{2}{1+y_{n+1}}. \tag{13.26}$$

Using (13.25) and (13.26) and the definitions of ζ and ζ^{-1} we obtain:

$$\begin{aligned}
\zeta(\zeta^{-1}(y)) &= \left(2\frac{y_1}{1+y_{n+1}}\frac{1}{1-|\zeta^{-1}(y)|^2}, \ldots, 2\frac{y_n}{1+y_{n+1}}\frac{1}{1-|\zeta^{-1}(y)|^2}, \frac{1+|\zeta^{-1}(y)|^2}{1-|\zeta^{-1}(y)|^2}\right) \\
&= \big(y_1, \ldots, y_n, y_{n+1}\big).
\end{aligned}$$

□

13.2.2 The hyperbolic metric of $\mathcal{B}^n$

The aim this section is to put a metric structure $d_{\mathcal{B}}$ onto $\mathcal{B}^n$, thus making $(\mathcal{B}^n, d_{\mathcal{B}})$ a hyperbolic space. The most natural way to do it is through the bijection ζ defined in the previous subsection. Indeed, we can transfer the metric $d_{\mathcal{H}}$ (see definition 13.5) of $\mathcal{H}^n$ onto $\mathcal{B}^n$, defining $d_{\mathcal{B}}$ in such a way that ζ becomes an isometry between $(\mathcal{B}^n, d_{\mathcal{B}})$ and $(\mathcal{H}^n, d_{\mathcal{H}})$.

Def. 13.2.1 *We define the hyperbolic metric, also called **Poincaré metric**, $d_{\mathcal{B}}$ on the open unit ball $\mathcal{B}^n$ as follows:*

$$d_{\mathcal{B}}(x, y) = d_{\mathcal{H}}(\zeta(x), \zeta(y)), \quad \textit{for all } x, y \in \mathcal{B}^n.$$

*The metric space $(\mathcal{B}^n, d_{\mathcal{B}})$ is called the conformal ball model of hyperbolic geometry and $(\mathcal{B}^2, d_{\mathcal{B}})$ is called the **Poincaré disk**.*

Once again, the hyperbolic cosine will give us an elegant reformulation of the metric.

Theorem 13.2.1 *For all $x, y \in \mathcal{B}^n$ the Poincaré metric $d_{\mathcal{B}}$ satisfies*

$$\cosh(d_{\mathcal{B}}(x,y)) = 1 + \frac{2|x-y|^2}{(1-|x|^2)(1-|y|^2)}.$$

Proof.

$$\begin{aligned}
\cosh(d_{\mathcal{B}}(x,y))) &\underset{\text{def.13.2.1}}{=} \cosh(d_{\mathcal{H}}(\zeta(x),\zeta(y))) \\
&\underset{(13.5)}{=} -\zeta(x)\circ\zeta(y) \\
&= -\sum_{i=1}^{n}\frac{4x_iy_i}{(1-|x|^2)(1-|y|^2)} + \frac{(1+|x|^2)(1+|y|^2)}{(1-|x|^2)(1-|y|^2)} \\
&= \frac{-4\langle x,y\rangle + 1 + |x|^2 + |y|^2 + |x|^2|y|^2}{(1-|x|^2)(1-|y|^2)} \\
&= \frac{(1-|x|^2)(1-|y|^2) + 2|x|^2 + 2|y|^2 - 4\langle x,y\rangle}{(1-|x|^2)(1-|y|^2)} \\
&= 1 + \frac{2|x-y|^2}{(1-|x|^2)(1-|y|^2).}
\end{aligned}$$

□

Let us now try to give a graphic interpretation of the behavior of this metric on the Poincaré disk. Let us consider the sets S_k, with $k = 1, 2, 3, \dots$ defined as follows:

$$S_k = \{x \in \mathcal{B}^2 : d_{\mathcal{B}}(0,x) = k\},$$

i.e. the set of all points in $\mathcal{B}^2$ having constant hyperbolic distance equal to k from the origin of the disk. After a straightforward computation, S_k can be rewritten as

$$S_k = \left\{x \in \mathcal{B}^2 : |x|^2 = \frac{\cosh(k)-1}{\cosh(k)+1}\right\}.$$

This means that the sets S_k are circles with center in the origin and radius $\sqrt{\frac{\cosh(k)-1}{\cosh(k)+1}}$. As shown in figure 13.5, while k increases of equal steps of length 1, the circles tend to accumulate rapidly on the border of $\mathcal{B}^n$.

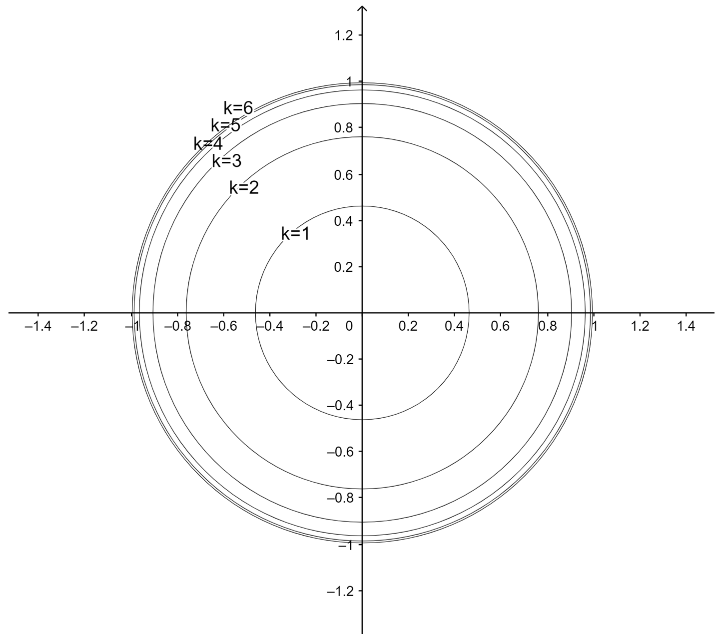


Figure 13.5: Depiction of S_k with $k = 1, \dots, 6$.

Corollary 13.2.1 *For all $x \in \mathcal{B}^n$ we have that*

$$d_{\mathcal{B}}(0, x) = \log\left(\frac{1 + |x|}{1 - |x|}\right).$$

Proof. Let $x \in \mathcal{B}^n$. As a direct computation from theorem 13.2.1 we obtain

$$\cosh(d_{\mathcal{B}}(0, x)) = 1 + \frac{2|x|^2}{1 - |x|^2} = \frac{1 + |x|^2}{1 - |x|^2}.$$

This formula can be rewritten in an elegant way using the fact that

$$\operatorname{arccosh}(y) = \log(y + \sqrt{y^2 - 1})$$

obtaining:

$$\begin{aligned} d_{\mathcal{B}}(0, x) &= \log\left(\frac{1 + |x|^2}{1 - |x|^2} + \sqrt{\frac{(1 + |x|^2)^2}{(1 - |x|^2)^2} - 1}\right) \\ &= \log\left(\frac{1 + |x|^2}{1 - |x|^2} + \frac{2|x|}{1 - |x|^2}\right) \\ &= \log\left(\frac{1 + |x|^2 + 2|x|}{1 - |x|^2}\right) \\ &= \log\left(\frac{1 + |x|}{1 - |x|}\right). \end{aligned}$$

□

As we did for $\mathcal{H}^n$ in definition 13.1.15, we are going to define now the element of hyperbolic arc length also for the conformal model $\mathcal{B}^n$. The idea is still to use the isometry ζ to link the element of hyperbolic arc length on $\mathcal{B}^n$ to the analogue on $\mathcal{H}^n$.

Theorem 13.2.2 *The element of hyperbolic arc length $\|dx\|_{\mathcal{B}}$ of the conformal model $(\mathcal{B}^n, d_{\mathcal{B}})$ is given by:*

$$\|dx\|_{\mathcal{B}} = \frac{2|dx|}{1-|x|^2}.$$

Proof. Let $x \in \mathcal{B}^n$ and $y = \zeta(x)$. Since ζ is an isometry between $(\mathcal{B}^n, d_{\mathcal{B}})$ and $(\mathcal{H}^n, d_{\mathcal{H}})$, it is natural to define the element of hyperbolic arc length on $\mathcal{B}^n$ as $||dx||_{\mathcal{B}} = ||dy||$. We recall that the element of hyperbolic arc length of $\mathcal{H}^n$ is

$$||dy|| = (dy_1^2 + \cdots + dy_n^2 - dy_{n+1}^2)^{\frac{1}{2}}.$$

Then, by definition of ζ, we have that

$$y_i = \frac{2x_i}{1-|x|^2} \qquad \text{for } i = 1, \ldots, n.$$

We want to compute now dy_i, for all $i = 1, \ldots, n$. Let us first recall that, by formula (3.6), we have that

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

Note that for all $j \neq i$ it holds that:

$$\frac{\partial y_i}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{2x_i}{1-|x|^2}\right) = \frac{4x_i x_j}{(1-|x|^2)^2},$$

while

$$\frac{\partial y_i}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{2x_i}{1-|x|^2}\right) = \frac{2(1-|x|^2)+4x_i^2}{(1-|x|^2)^2} = \frac{2}{1-|x|^2} + \frac{4x_i^2}{(1-|x|^2)^2}.$$

Thence

$$\begin{aligned}
dy_i &= \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_j} dx_j = \frac{\partial y_i}{\partial x_i} dx_i + \sum_{j \neq i} \frac{\partial y_i}{\partial x_j} dx_j \\
&= \left(\frac{2}{1-|x|^2} + \frac{4x_i^2}{(1-|x|^2)^2}\right) dx_i + \sum_{j \neq i} \frac{4x_i x_j}{(1-|x|^2)^2} dx_j \\
&= \frac{2dx_i}{1-|x|^2} + \frac{4x_i^2 dx_i}{(1-|x|^2)^2} + \sum_{j \neq i} \frac{4x_i x_j}{(1-|x|^2)^2} dx_j \\
&= \frac{2dx_i}{1-|x|^2} + \sum_{j=1}^{n} \frac{4x_i x_j}{(1-|x|^2)^2} dx_j \\
&= \frac{2dx_i}{1-|x|^2} + \frac{4x_i}{(1-|x|^2)^2} \sum_{j=1}^{n} x_j dx_j \\
&= \frac{2dx_i}{1-|x|^2} + \frac{4x_i \langle x, dx \rangle}{(1-|x|^2)^2}.
\end{aligned}$$

So

$$dy_i^2 = \frac{4}{(1-|x|^2)^2}\left(dx_i^2 + \frac{4x_i dx_i \langle x, dx \rangle}{1-|x|^2} + \frac{4x_i^2 \langle x, dx \rangle^2}{(1-|x|^2)^2}\right).$$

Moreover, from the definition of ζ we know that

$$y_{n+1} = \frac{1+|x|^2}{1-|x|^2}.$$

Again from formula (3.6) we obtain:

$$\begin{aligned} dy_{n+1} &= \sum_{j=1}^{n} \frac{\partial y_{n+1}}{\partial x_j} dx_j = \sum_j \frac{\partial}{\partial x_j}\left(\frac{1+|x|^2}{1-|x|^2}\right) dx_j \\ &= \sum_{j=1}^{n} \frac{2x_i(1-|x|^2)+2x_i(1+|x|^2)}{(1-|x|^2)^2} dx_j \\ &= \sum_{j=1}^{n} \frac{4x_j}{(1-|x|^2)^2} dx_j \\ &= \frac{4}{(1-|x|^2)^2}\sum_{j=1}^{n} x_j dx_j \\ &= \frac{4\langle x, dx\rangle}{(1-|x|^2)^2}. \end{aligned}$$

Hence

$$dy_{n+1}^2 = \frac{16\langle x, dx\rangle^2}{(1-|x|^2)^4}.$$

Combining the previous results we obtain

$$\|dx\|_{\mathcal{B}} = \|dy\| = \sqrt{\sum_{i=1}^{n} dy_i^2 - dy_{n+1}^2} = \sqrt{\frac{4|dx|^2}{(1-|x|^2)^2}} = \frac{2|dx|}{1-|x|^2}.$$

□

13.2.3 The isometry group of $\mathcal{B}^n$

As previously announced, Möbius transformations play a key role in the conformal model. In the following we will prove indeed that they act isometrically, w.r.t. $d_{\mathcal{B}}$, on $\mathcal{B}^n$. Let us start our analysis with a useful lemma.

Lemma 13.2.1 *Let ϕ be a Möbius transformation stable on $\mathcal{B}^n$, i.e. $\phi \in \mathcal{M}(\mathcal{B}^n)$, then for all $x, y \in \mathcal{B}^n$ it holds that:*

$$\frac{|\phi(x)-\phi(y)|^2}{(1-|\phi(x)|^2)(1-|\phi(y)|^2)} = \frac{|x-y|^2}{(1-|x|^2)(1-|y|^2)}. \tag{13.27}$$

Proof. In this proof we are going to use several results of section 12.5, in particular the definition of orthogonality between spheres 12.5.4 and theorem 12.5.6.

Let $\phi \in \mathcal{M}(\mathcal{B}^n)$ and let us recall that theorem 12.5.6 states that, if $\phi(\infty) = \infty$, then $\phi \in \mathrm{O}(n)$. This means that $|\phi(x)-\phi(y)| = |x-y|$,$|\phi(x)| = |x|$ and $|\phi(y)| = |y|$, hence equation (13.27) is clearly satisfied.

Theorem 12.5.6 also states that, if $\phi(\infty) \neq \infty$, then ϕ can be decomposed as $\phi = \psi \circ \sigma$, with $\psi \in \mathrm{O}(n)$ and $\sigma = \sigma_{a,r}$ inversion with respect to a sphere $\Sigma = S^{n-1}_{a,r}$ orthogonal to S^{n-1}. Since ψ satisfies equation (13.27), to conclude it is sufficient to prove that σ satisfies it too.

Hence, let us consider $\phi = \sigma_{a,r}$, with $S^{n-1}_{a,r}$ orthogonal to S^{n-1}. From the third possible scenario in definition 12.5.4, we know that $r^2 = |a|^2 - 1$. As an inversion ϕ has the following expression:

$$\phi(x) = a + \frac{r^2}{|x-a|^2}(x-a),$$

we have

$$|\phi(x)|^2 = |a|^2 + \frac{2r^2 \langle a, x\rangle - 2r^2 + r^4}{|x-a|^2},$$

so

$$\begin{aligned}
|\phi(x)|^2 - 1 &= |a|^2 + \frac{2r^2 \langle a, x\rangle - 2r^2 + r^4}{|x-a|^2} \\
&= \frac{|a|^2|x-a|^2 + 2r^2 \langle a, x-a\rangle + r^4 - |x-a|^2}{|x-a|^2} \\
&= \frac{\big(|a|^2 - 1\big)|x-a|^2 + 2r^2 \langle a, x-a\rangle + r^4}{|x-a|^2} \\
&\underset{|a|^2-1=r^2}{=} \frac{r^2\big(|x-a|^2 + 2\langle a, x-a\rangle + r^2\big)}{|x-a|^2} \\
&\underset{r^2=|a|^2-1}{=} \frac{r^2\big(|x-a|^2 + 2\langle a, x-a\rangle + |a|^2 - 1\big)}{|x-a|^2} \\
&= \frac{r^2(|x-a+a|^2 - 1)}{|x-a|^2} \\
&= \frac{r^2(|x|^2 - 1)}{|x-a|^2}.
\end{aligned}$$

Since ϕ is an inversion, from **3.** in theorem 12.2.2 we have that:

$$|\phi(x) - \phi(y)|^2 = \frac{r^4|x-y|^2}{|x-a||y-a|}.$$

Putting together the previous results we come to the conclusion as follows:

$$\begin{aligned}
\frac{|\phi(x) - \phi(y)|^2}{(1 - |\phi(x)|^2)(1 - |\phi(y)|^2)} &= \frac{r^4|x-y|^2}{|x-a|^2|y-a|^2} \frac{|x-a|^2|y-a|^2}{r^4(1-|x|^2)(1-|y|^2)} \\
&= \frac{|x-y|^2}{(1-|x|^2)(1-|y|^2)}.
\end{aligned}$$

□

A direct consequence of this lemma and the expression of $d_{\mathcal{B}}$ given in theorem 13.2.1 is the following theorem.

Theorem 13.2.3 *Let $\phi \in \mathcal{M}(\mathcal{B}^n)$, then ϕ acts as an isometry on $(\mathcal{B}^n, d_{\mathcal{B}})$:*

$$d_{\mathcal{B}}(x, y) = d_{\mathcal{B}}(\phi(x), \phi(y)),$$

for all $x, y \in \mathcal{B}^n$.

This theorem is a first step towards our aim to prove that the isometry group of $(\mathcal{B}^n, d_{\mathcal{B}})$ is linked to its Möbius group $\mathcal{M}(\mathcal{B}^n)$.

Lemma 13.2.2 *The action of $\mathcal{M}(\mathcal{B}^n)$ on $\mathcal{B}^n$ is transitive.*

Proof. To prove the transitivity it is sufficient to prove that for all $a \in \mathcal{B}^n$ it exists an element $\sigma_a \in \mathcal{M}(\mathcal{B}^n)$ such that $\sigma_a(0) = a$. Clearly, if $a = 0$, we can choose $\sigma_0 = id_{\mathcal{B}^n} \in \mathcal{M}(\mathcal{B}^n)$. Let us consider $a \in \mathcal{B}^n$, $a \neq 0$ and set $\sigma_a = \sigma_{\frac{a}{|a|^2},r}$ such that $r^2 = \frac{1}{|a|^2} - 1$, then $\sigma_a \in \mathcal{M}(\mathcal{B}^n)$ since σ_a is an inversion of a sphere orthogonal to S^{n-1}, by theorem, 12.5.5 $\sigma_a \in \mathcal{M}(\mathcal{B}^n)$. As an inversion σ_a has the following expression:

$$\sigma_a(x) = \sigma_{\frac{a}{|a|^2},r}(x) = \frac{a}{|a|^2} + \frac{r^2}{\left|x - \frac{a}{|a|^2}\right|^2}\left(x - \frac{a}{|a|^2}\right),$$

hence, using the fact that $\frac{1}{|a|^2} - r^2 = 1$, we obtain

$$\sigma_a(0) = \frac{a}{|a|^2} - r^2\frac{|a|^4}{|a|^2}\frac{a}{|a|^2} = a\left(\frac{1}{|a|^2} - r^2\right) = a.$$

$\square$

Theorem 13.2.4 *Every Möbius transformation of $\mathcal{B}^n$ restricts to an isometry of the conformal model $\mathcal{B}^n$, and every isometry of $\mathcal{B}^n$ extends to unique a Möbius transformation of $\mathcal{B}^n$.*

Proof. As seen in theorem 13.2.3, every Möbius transformation of $\mathcal{M}(\mathcal{B}^n)$, when restricted to $\mathcal{B}^n$, is an isometry of $(\mathcal{B}^n, d_{\mathcal{B}})$. Hence, it remains to prove that every isometry of $\mathcal{M}(\mathcal{B}^n)$ can be extendend to a unique element of $\mathcal{M}(\mathcal{B}^n)$.

Let ϕ be an isometric transformation on $\mathcal{B}^n$, w.r.t. the metric $d_{\mathcal{B}}$. Let us start by defining the function ψ as follows:

$$\psi = \begin{cases} \phi & \text{if } \phi(0) = 0 \\ \sigma|_{\mathcal{B}^n}\,\phi & \text{if } \phi(0) \neq 0 \end{cases},$$

where σ is an element of $\mathcal{M}(\mathcal{B}^n)$ such that $\sigma(\phi(0)) = 0$, where the existence of such σ is guaranteed by Lemma 13.2.2. Because of the first part of this theorem, since $\sigma \in \mathcal{M}(\mathcal{B}^n)$, then $\sigma|_{\mathcal{B}^n}$ is an isometry of $\mathcal{B}^n$. Thus, $\psi(0) = 0$ and ψ is an isometry of $\mathcal{B}^n$. We notice that for $x, y \in \mathcal{B}^n$, we have

$$\begin{aligned} d_{\mathcal{B}}(\psi(x), 0) = d_{\mathcal{B}}(x, 0) \quad &\Longleftrightarrow \quad \frac{|\psi(x)|^2}{1 - |\psi(x)|^2} = \frac{|x|^2}{1 - |x|^2} \\ &\Longleftrightarrow \quad |\psi(x)| = |x|. \end{aligned}$$

In the same way, $|\psi(y)| = |y|$. From this we deduce that ψ preserves Euclidean distances on $\mathcal{B}^n$ as follows:

$$\begin{aligned} d_{\mathcal{B}}(\psi(x), \psi(y)) = d_{\mathcal{B}}(x, y) \quad &\Longleftrightarrow \quad \frac{|\psi(x) - \psi(y)|^2}{(1 - |\psi(x)|^2)} = \frac{|x - y|^2}{(1 - |x|^2)(1 - |y|^2)} \\ &\Longleftrightarrow \quad |\psi(x) - \psi(y)| = |x - y|. \end{aligned}$$

Now we know that $\psi(0) = 0$ and ψ is a Euclidean isometry on $\mathcal{B}^n$, hence it is linear and continuous on $\mathcal{B}^n$. By continuity we can extend it to $\overline{\mathcal{B}^n}$ by setting its extension $\bar{\psi} : \overline{\mathcal{B}^n} \to \overline{\mathcal{B}^n}$ as $\bar{\psi}(x) := 2\psi\left(\frac{x}{2}\right)$. Clearly $\bar{\psi}$ is linear as well on $\mathcal{B}^n$ and it preserves Euclidean distances on $\mathcal{B}^n$. This means that it preserves the Euclidean scalar product and maps orthonormal bases into orthonormal bases. Thus by Lemma 11.1.2 there exists $A \in \mathrm{O}(n)$ such that $\bar{\psi} = A|_{\bar{\mathcal{B}}^n}$. Now we are allowed to define our candidate Φ, extension of ϕ, as follows:

$$\Phi = \begin{cases} A & \text{if } \phi(0) = 0 \\ \sigma^{-1}A & \text{if } \phi(0) \neq 0 \end{cases}.$$

We just need to check that it is a good candidate, i.e. that $\Phi \in \mathcal{M}(\mathcal{B}^n)$ and $\Phi|_{\mathcal{B}^n} = \phi$. As an element of $\mathrm{O}(n)$, by corollary 12.5.5, $A \in \mathcal{M}(\mathcal{B}^n)$, while $\sigma^{-1} \in \mathcal{M}(\mathcal{B}^n)$ by definition of σ, hence $\Phi \in \mathcal{M}(\mathcal{B}^n)$. If $\phi(0) = 0$, then

$$\Phi|_{\mathcal{B}^n} = A|_{\mathcal{B}^n} = \bar{\psi}\big|_{\mathcal{B}^n} = \psi = \phi,$$

while, if $\phi(0) \neq 0$, we have that

$$\Phi|_{\mathcal{B}^n} = (\sigma^{-1} \circ A)\big|_{\mathcal{B}^n} = \sigma^{-1}\big|_{\mathcal{B}^n} \circ \bar{\psi}\big|_{\mathcal{B}^n} = \sigma^{-1}\big|_{\mathcal{B}^n} \circ \psi = \sigma^{-1}\big|_{\mathcal{B}^n} \circ \sigma|_{\mathcal{B}^n} \circ \phi = \phi.$$

So $\Phi|_{\mathcal{B}^n} = \phi$ and Φ is an extension of ϕ.

We want to prove now the uniqueness of the extension Φ of ϕ. Let us suppose that there exist two extensions $\Phi_1, \Phi_2 \in \mathcal{M}(\mathcal{B}^n)$ of ϕ, i.e. $\Phi_1|_{\mathcal{B}^n} = \Phi_2|_{\mathcal{B}^n} = \phi$. By continuity it holds that $\Phi_1|_{S^{n-1}} = \Phi_2|_{S^{n-1}}$, i.e. for all $x \in S^{n-1}$ we have that $\Phi_1(x) = \Phi_2(x)$. This means that, for all $x \in S^{n-1}$ $\Phi_2^{-1} \circ \Phi_1(x) = x$, i.e. $\Phi_2^{-1} \circ \Phi_1 \in \mathcal{M}(\mathcal{B}^n) \subset \mathcal{M}(\hat{\mathbb{R}}^n)$ fixes the sphere S^{n-1} pointwise. This allows us to apply theorem 12.4.5, which states that $\Phi_2^{-1} \circ \Phi_1$ is either $id_{\hat{\mathbb{R}}^n}$ or the inversion w.r.t. S^{n-1}. This second case is not possible, because, if $\Phi_2^{-1} \circ \Phi_1$ was the inversion w.r.t. S^{n-1}, then it should map $\mathcal{B}^n$ into $\overline{\mathcal{B}^n}^c$, while both Φ_1 and Φ_2^{-1}, as elements of $\mathcal{M}(\mathcal{B}^n)$, are stable on $\mathcal{B}^n$. Hence $\Phi_2^{-1} \circ \Phi_1 = id_{\hat{\mathbb{R}}^n}$, so $\Phi_1 = \Phi_2$. □

Let us call here $\mathcal{I}(\mathcal{B}^n)$ the isometry group of $\mathcal{B}^n$ w.r.t. the metric $d_{\mathcal{B}}$. Warning: in section 12.5 we used the same notation to indicate the isometry group of $\mathcal{B}^n$ w.r.t. the Euclidean metric. A direct consequence of this theorem, analogous to theorem 13.1.2 for $\mathcal{H}^n$, is the following corollary.

Corollary 13.2.2 *$\mathcal{I}(\mathcal{B}^n)$ and $\mathcal{M}(\mathcal{B}^n)$ are isomorphic.*

The metric $d_{\mathcal{B}}$ of the conformal model $\mathcal{B}^n$ is classically expressed in two possible equivalent ways. We have already seen the first one in theorem 13.2.1. Now we are going to introduce the second one, which makes use of several results of chapter 12. The most relevant ones are theorem 12.4.1, theorem 12.4.3 and corollary 12.4.4.

While the first expression is a natural consequence of the link with the model $\mathcal{H}^n$, this second one involves the concept of cross-ratio, defined in subsection 12.4.1.

Theorem 13.2.5 *Let $x, y \in \mathcal{B}^n$ and let $u, v \in S^{n-1}$ be the two points obtained by intersecting the circle, or line, containing x and y and orthogonal to S^{n-1}, as depicted in figure 13.6, then:*

$$d_{\mathcal{B}}(x, y) = \log([x, y, u, v]) = \log\left(\frac{|x-u||y-v|}{|x-v||y-u|}\right).$$

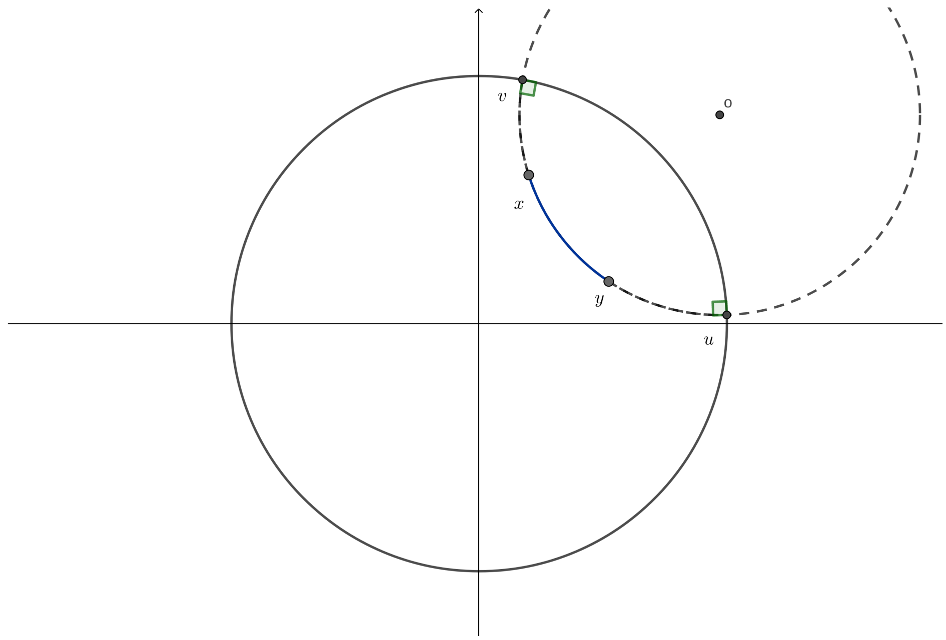


Figure 13.6: Depiction of the graphic construction to find points u and v starting from x and y.

Proof. First of all notice that, if $x = y$, we do not need to find the points u and v, because $d_{\mathcal{B}}(x,x) = \log([x,x,u,v]) = \log\left(\frac{|x-u||x-v|}{|x-v||x-u|}\right) = \log 1 = 0$.

Let $x, y \in \mathcal{B}^n$ be two distinct points. For simplicity, let us start by supposing that $y = 0$. In this particular case to prove the theorem it suffices to rewrite the statement of corollary 13.2.1 as follows:

$$\begin{aligned} d_{\mathcal{B}}(x,0) &= \log\left(\frac{1+|x|}{1-|x|}\right) \\ &= \log\left(\frac{\left|x+\frac{x}{|x|}\right|\left|0-\frac{x}{|x|}\right|}{\left|x-\frac{x}{|x|}\right|\left|0+\frac{x}{|x|}\right|}\right) \\ &= \log\left(\left[x, 0, -\frac{x}{|x|}, \frac{x}{|x|}\right]\right). \end{aligned}$$

Hence, for all $x \in \mathcal{B}^n$ and for $y = 0$ we obtain that the statement of the theorem is true, since $u = -\frac{x}{|x|}$ and $v = \frac{x}{|x|}$ are obtained by the intersection from the straight line $L_x = \{tx : t \in \mathbb{R}\}$ joining x and 0, orthogonal to S^{n-1}, as depicted in figure 13.7. Clearly L_x is orthogonal to S^{n-1} since it is a diameter of the unit disk.

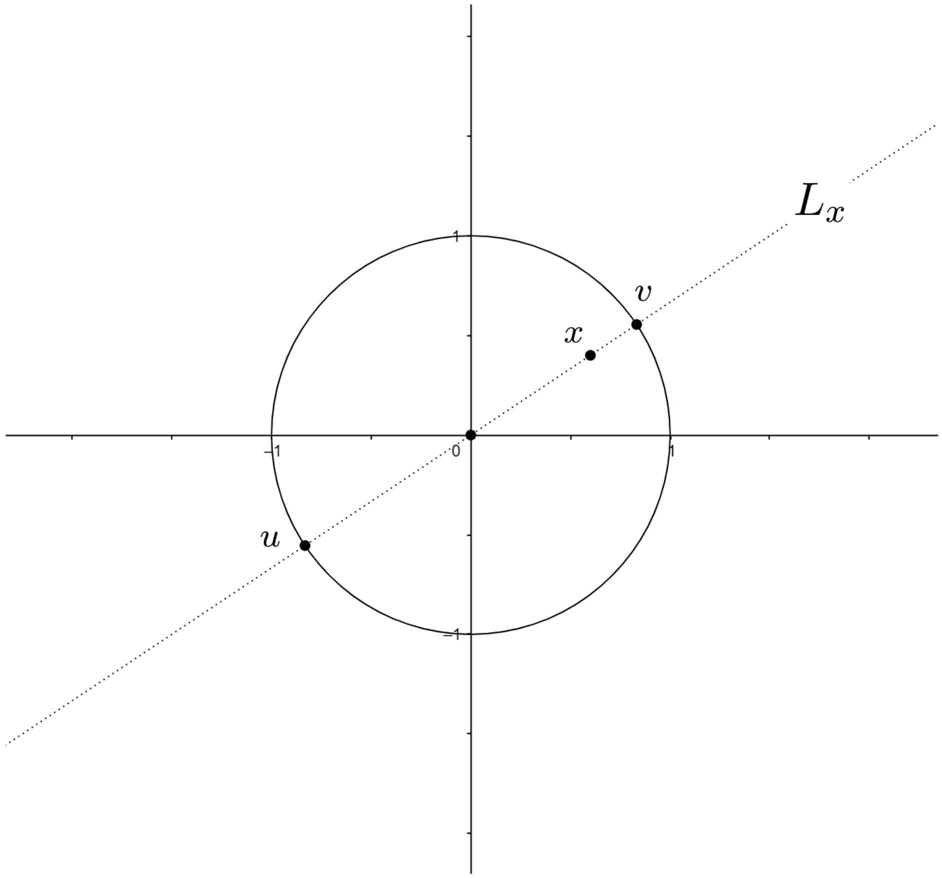


Figure 13.7: Depiction of the points $x, y = 0, u = -\frac{x}{|x|}, v = \frac{x}{|x|}$, aligned on L_x.

Let us consider now the case $y \neq 0$. Here we are going to use the same idea we used in the proof of lemma 13.2.2. Since $y \neq 0$, we can define $\tilde{y} = \frac{y}{|y|^2}$ and $r = \sqrt{|\tilde{y}|^2 - 1}$. Let us consider the inversion $\sigma := \sigma_{\tilde{y},r}$ w.r.t. the sphere $S^{n-1}_{\tilde{y},r}$. By the third possible scenario in definition 12.5.4 it is clear that $S_{\tilde{y},r}$ is orthogonal to S^{n-1}. Moreover by theorem 12.5.5 $\sigma \in \mathcal{M}(\mathcal{B}^n)$. Being an inversion, σ has the following expression:

$$\sigma(x) = \frac{y}{|y|^2} + \frac{\frac{1}{|y|^2} - 1}{\left|x - \frac{y}{|y|^2}\right|^2}\left(x - \frac{y}{|y|^2}\right),$$

hence

$$\sigma(y) = \frac{y}{|y|^2} + \frac{\frac{1}{|y|^2} - 1}{\left|y - \frac{y}{|y|^2}\right|}\left(y - \frac{y}{|y|^2}\right) = \frac{y}{|y|^2} + \frac{1}{|y|^2}\frac{1 - |y|^2}{||y|^2 - 1|^2} y(|y|^2 - 1) = 0. \tag{13.28}$$

Furthermore, since $\sigma \in \mathcal{M}(\mathcal{B}^n)$, by theorem 13.2.3 it is an isometry w.r.t. $d_{\mathcal{B}}$. Hence we can come back to the simplest case (with the straight line and $y = 0$) as follows:

$$\begin{aligned} d_{\mathcal{B}}(x,y) \underset{\text{th.13.2.3}}{=} & \; d_{\mathcal{B}}(\sigma(x),\sigma(y)) \underset{(13.28)}{=} d_{\mathcal{B}}(\sigma(x),0) \\ = & \; \log\left[\sigma(x),0,\tilde{u},\tilde{v}\right] \\ = & \; \log\left[\sigma(x),0,-\frac{\sigma(x)}{|\sigma(x)|},\frac{\sigma(x)}{|\sigma(x)|}\right], \end{aligned}$$

where $\tilde{u} = -\frac{\sigma(x)}{|\sigma(x)|}$ and $\tilde{v} = \frac{\sigma(x)}{|\sigma(x)|}$ are the two points of intersection of the line $L_{\sigma(x)} = \{t\sigma(x) : t \in \mathbb{R}\}$ with S^{n-1}. Let us call u and v the inverse points[2] of $\tilde{u}$ and $\tilde{v}$. By theorem 12.4.1, as a Möbius transformation, σ^{-1} preserves cross-ratios, thus the previous equation becomes

$$\begin{aligned} d_{\mathcal{B}}(x,y) &= \log\left[\sigma(x),0,\tilde{u},\tilde{v}\right] \\ &= \log\left[\sigma^{-1}(\sigma(x)),\sigma^{-1}(0),\sigma^{-1}(\tilde{u}),\sigma^{-1}(\tilde{v})\right] \\ &= \log\left[x,\sigma(0),\sigma(\tilde{u}),\sigma(\tilde{v})\right] \\ &= \log\left[x,y,u,v\right]. \end{aligned}$$

What is left to prove is that u and v are the points of intersection of a straight line or circle $C_{x,y}$ and S^{n-1}.

Let us consider the plane $\operatorname{span}(x,y)$. Of course, also $\tilde{y} = \frac{y}{|y|^2}$, belongs to $\operatorname{span}(x,y)$. Note that the action of σ is stable on $\operatorname{span}(x,y)$, indeed, for all $z \in \operatorname{span}(x,y)$ it holds that $\sigma(z) = \sigma^{-1}(z)$ belongs to the straight line joining y and z, which is contained in $\operatorname{span}(x,y)$, since both y and z belong to $\operatorname{span}(x,y)$.

As an element of $\mathcal{M}(\mathcal{B}^n)$, σ is stable on S^{n-1}. This means that σ is stable on $S^{n-1} \cap \operatorname{span}(x,y)$, thus, since $\tilde{u},\tilde{v} \in S^{n-1} \cap \operatorname{span}(x,y)$, the points u and v belong to $S^{n-1} \cap \operatorname{span}(x,y)$.

Note that $\sigma|_{\operatorname{span}(x,y)}$ is a Möbius transformation on $\operatorname{span}(x,y)$, in particular it is a circular inversion.

By theorem 12.4.3 we know that σ maps spheres of $\hat{\mathbb{R}}^n$ into spheres of $\hat{\mathbb{R}}^n$, the same holds for $\sigma|_{\operatorname{span}(x,y)}$ on $\operatorname{span}(x,y)$, i.e. it is stable on the set of circles and straight lines of $\operatorname{span}(x,y)$ (obtained by restriction of the hyperspheres of $\hat{\mathbb{R}}^n$).

By construction we know that the points $\sigma(x), 0, \tilde{u}$, and $\tilde{v}$ belong to the straight line $L_{\sigma(x)}$, which is mapped into a circle or straight line, hence the points x, y, u and v belong to the same circle or straight line $C_{x,y}$.

Moreover, by corollary 12.4.4, σ preserves orthogonality between spheres, and so does $\sigma|_{\operatorname{span}(x,y)}$. Since $L_{\sigma(x)}$ is orthogonal to S^{n-1}, also its image $C_{x,y}$ is orthogonal to S^{n-1}. □

Figure 13.8 provides a depiction of the geometric construction associated to this proof.

Let us now make some consideration concerning the geometric construction of $C_{x,y}$, which can be either a straight line or a circle, i.e. an sphere of $\hat{\mathbb{R}}^2$. Given $x, y \in \mathcal{B}^n$ we want to find $C_{x,y}$ orthogonal to S^{n-1}.

- If $x = y$, as we have seen at the beginning of the previous proof, there is no construction needed.

[2] By definition of 12.4.4 w.r.t. an inversion $u = \sigma(\tilde{u}) = \sigma^{-1}(\tilde{u})$ and $v = \sigma(\tilde{v}) = \sigma^{-1}(\tilde{v})$, where $\sigma = \sigma^{-1}$ by property **2.** in theorem 12.2.2.

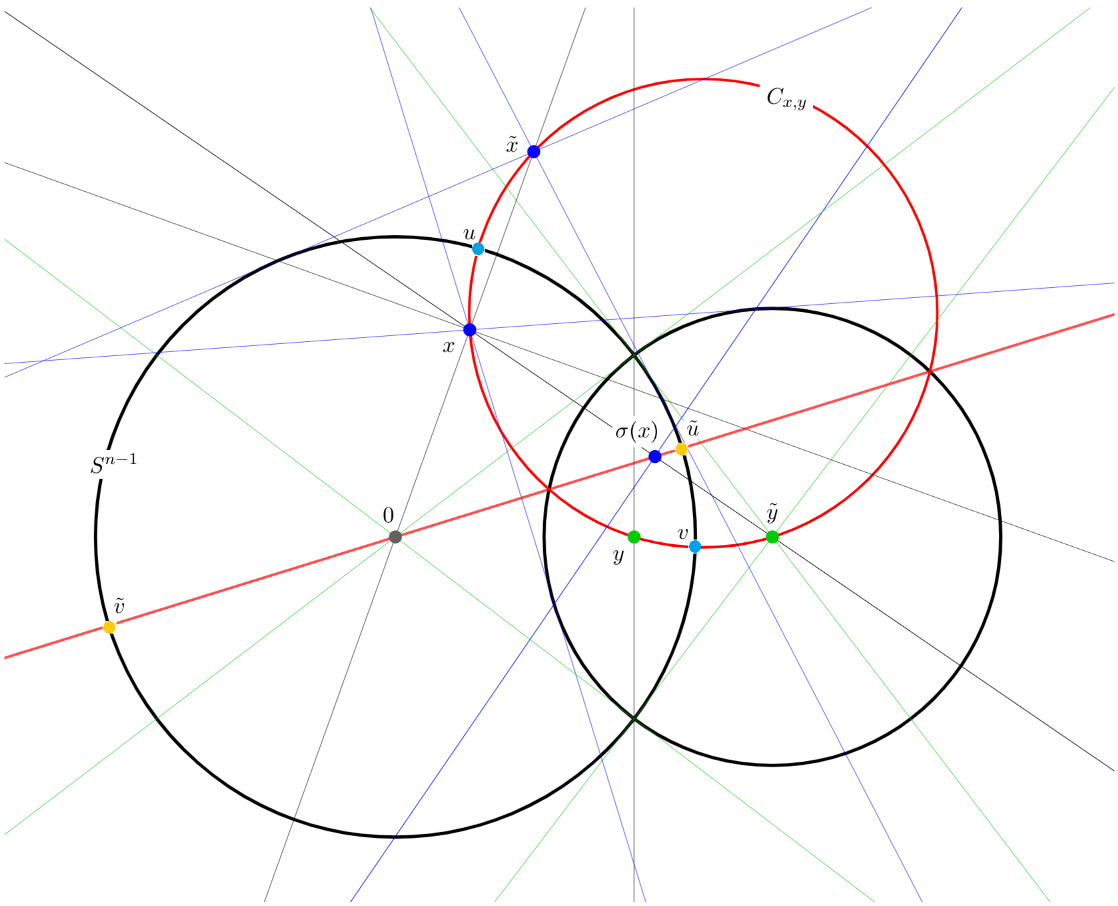


Figure 13.8: This picture can be useful to visualize all the objects that we have built during the proof of theorem 13.2.5.

- If $x \neq y$ there are two possible scenarios whether 0 belongs or not to the straight line joining x and y, let us call it $L_{x,y}$. If $0 \in L_{x,y}$ then $L_{x,y} \cap \mathcal{B}^n$ is a diameter of the unit ball, hence it is orthogonal to S^{n-1}, thus $L_{x,y} = C_{x,y}$.

  When $0 \notin L_{x,y}$, then, by condition **2.** in definition 12.5.4, $L_{x,y}$ is not orthogonal to S^{n-1}. Hence $C_{x,y}$ is not a straight line, but a circle. We are looking for a circle that satisfies condition **3.** in definition 12.5.4, i.e. $|a|^2 = r^2 + 1$. All the points x belonging to $C_{x,y}$ are such that $|x - a|^2 = r^2 = |a|^2 - 1$, after a straightforward computation this last equation becomes $|x|^2 - 2\langle a, x\rangle = -1$. It is easy to verify that, if x belongs to $C_{x,y}$ also $\tilde{x} = \frac{x}{|x|^2}$ belongs to it. Thence $C_{x,y}$ is the unique circle passing through the three points $x, y, \tilde{x}$.

  A second possible way of constructing the circle $C_{x,y}$ is noticing that, since $x, y \in C_{x,y}$, its center a is equidistant from x and y, i.e. a belongs to the perpendicular bisector of the segment joining x and y. This means that a satisfies a linear equation whose coefficients involve x and y ad constants. This equation, together with the equation $|x|^2 - 2\langle a, x\rangle = -1$ (which is linear in a), seen above is sufficient to determine uniquely a and thus also $r = \sqrt{|a|^2 - 1}$. So also in this case we have determined uniquely the circle $C_{x,y}$.

13.2.4 Geodesics of $\mathcal{B}^n$

From the last part of the previous subsection we might get the intuition that the geodesics of the conformal ball model are lines and circles orthogonal to S^{n-1}.

From corollary 13.2.2 we know that $\mathcal{I}(\mathcal{B}^n) \cong \mathcal{M}(\mathcal{B}^n)$, hence every element of the isometry group should map geodesics of $(\mathcal{B}^n, d_{\mathcal{B}})$ into other geodesics. As we have seen in the proof of theorem 13.2.5 the elements of $\mathcal{I}(\mathcal{B}^n) \cong \mathcal{M}(\mathcal{B}^n)$ are stable on the set of lines and circles orthogonal to S^{n-1} that is what suggests us that they might be a good candidate to be the geodesics of this hyperbolic model.

This intuition will be confirmed to be a particular instance ($m = 1$) of the more general result stated in theorem 13.2.6, which involves the concept of **m-spheres** and **m-planes** orthogonal to S^{n-1} as defined below.

Def. 13.2.2 *Let us consider $\mathbb{R}^n$ and $m \in \mathbb{N}$, with $m < n$.*

1. *A m-plane of $\mathbb{R}^n$ is a vector subspace of $\mathbb{R}^n$ of dimension m.*

2. *A m-sphere is the intersection between a sphere $S^{n-1}_{a,r}$ and a $m+1$-plane of $\mathbb{R}^n$ containing its center a, hence it has dimension m.*

Clearly lines and circles orthogonal to S^{n-1} are examples of 1-planes and 1-spheres, respectively. As we did in subsection 12.4.2, we can call the set of all the extendend m-planes of $\mathbb{R}^n$ (i.e. m-planes $\cup \{\infty\}$) together with the m-spheres of $\mathbb{R}^n$ the set of all the m-spheres of $\hat{\mathbb{R}}^n$.

Theorem 13.2.6 *The action of $\mathcal{M}(\mathcal{B}^n) \cong \mathcal{I}(\mathcal{B}^n)$ is transitive on the set of all m-spheres of $\hat{\mathbb{R}}^n$ orthogonal to S^{n-1}, with $m \in \{1, \dots, n-1\}$.*

Proof. Let us start by noticing that the action of $\mathcal{M}(\mathcal{B}^n)$, as a subset of $\mathcal{M}(\hat{\mathbb{R}}^n)$, is stable on the set of hypespheres of $\hat{\mathbb{R}}^n$, as a consequence of theorem 12.4.3. Moreover, as direct consequence of corollary 12.4.4, the action of $\mathcal{M}(\mathcal{B}^n) \subset \mathcal{M}(\hat{\mathbb{R}}^n)$, is stable on the set of hypespheres of $\hat{\mathbb{R}}^n$ orthogonal to S^{n-1}. Since the m-spheres of $\hat{\mathbb{R}}^n$ are obtained by intersection of hyperspheres of $\hat{\mathbb{R}}^n$ with vector subspaces of $\mathbb{R}^n$, we can conclude that the action of $\mathcal{M}(\mathcal{B}^n)$ is stable on the set of m-spheres of $\hat{\mathbb{R}}^n$ orthogonal to S^{n-1}.

What is left to prove is that for all any couple of m-spheres Σ_1, Σ_2 of $\hat{\mathbb{R}}^n$ orthogonal to S^{n-1}, it exists $\phi \in \mathcal{M}(\mathcal{B}^n)$ such that $\phi(\Sigma_1) = \Sigma_2$. Thence it is exhaustive to consider the following the three possible scenarios listed below.

1. Let Σ_1 and Σ_2 be two m-planes orthogonal to S^{n-1}, i.e. m-planes passing through the origin, it is sufficient to prove that a generic m-plane $V'_m = \operatorname{span}(v_1, \dots, v_m)$, where $(v_1, \dots, v_m)$ is an orthonormal basis of $\mathbb{R}^n$, can be mapped into the m-plane given by $V_m = \operatorname{span}(e_1, \dots, e_m)$. This is true thanks to the fact that we can always complete $(v_1, \dots, v_m)$ to a orthonormal basis of $\mathbb{R}^n$ $(v_1, \dots, v_m, v_{m+1}, \dots, v_n)$ through the Gram-Schmidt procedure, then, by setting $A := \begin{pmatrix} | & \dots & | \\ v_1 & \dots & v_n \\ | & \dots & | \end{pmatrix}$, we have that A^{-1} maps $V'_m \cup \{\infty\}$ into $V_m \cup \{\infty\}$. Finally, by point **1.** of theorem 12.5.6, $A^{-1} \in \mathrm{O}(n) \subset \mathcal{M}(\mathcal{B}^n)$.

2. Let us now consider the scenario in which $\Sigma_1 = S^{n-1}_{a,r} \cap V_{m+1}$ and $\Sigma_2 = P(b, 0) \cap V'_{m+1}$. Σ_1 is supposed to be orthogonal to S^{n-1}, so a and r are such that $|a|^2 = r^2 + 1$, and

Σ_2 is a m-plane, hence it is orthogonal to S^{n-1}. Furthermore, from point **1.** of this proof, it is not restrictive to suppose $V_{m+1} = V'_{m+1}$. Moreover, since point **1.** and point **3.** prove that $\mathcal{M}(\mathcal{B}^n)$ acts transitively on the set of m-planes and m-spheres of $\mathbb{R}^n$, respectively, it is not restrictive to choose $b = \frac{a}{|a|}$. Thence we have $\Sigma_1 = S^{n-1}_{a,r} \cap V_{m+1}$ and $\Sigma_2 = \left(P\left(\frac{a}{|a|}, 0\right) \cup \{\infty\}\right) \cap V_{m+1}$. Since Σ_1 is orthogonal to S^{n-1}, we have that $\Sigma_1 \cup S^{n-1} \neq \varnothing$. Hence it is possible to find a point[3] a' belonging to the line joining a and 0, i.e. $a' \in \operatorname{span}(a)$, such that $a' \in \mathcal{B}^n$. Condition $|a|^2 = r^2 + 1$ implies that $|a| > r$, hence we can set $a' = \frac{|a|-r}{|a|}a$ as depicted in figure 13.9.

As we did before the equation describing $S^{n-1}_{a,r}$, can be rewritten as $|x|^2 - 2\langle a, x\rangle + 1 = 0$. After a straightforward computation it is easy to verify that if $a' \in S^{n-1}_{a,r}$ also $\sigma_{0,1}(a') = \frac{a'}{|a'|^2} \in S^{n-1}_{a,r}$. Moreover, since $a' \in \mathcal{B}^n$, then $\frac{a'}{|a'|^2} \in \overline{\mathcal{B}^n}^c$ hence we can consider the sphere $S^{n-1}_{\frac{a'}{|a'|^2}, r'_a}$ centered in $\frac{a'}{|a'|^2}$, orthogonal to S^{n-1}, i.e. $r'_a = \sqrt{\frac{1}{|a'|^2} - 1}$. Let us call σ_a the circular inversion w.r.t. $S^{n-1}_{\frac{a'}{|a'|^2}, r'_a}$. Clearly $\sigma_a \in \mathcal{M}(\mathcal{B}^n)$.

We are going to prove now that $\sigma_a(S^{n-1}_{a,r}) = P\left(\frac{a}{|a|}, 0\right) \cup \{\infty\}$, hence $\sigma_a(S^{n-1}_{a,r} \cap V_{m+1}) = \left(P\left(\frac{a}{|a|}, 0\right) \cup \{\infty\}\right) \cap V_{m+1}$, so $\sigma_a(\Sigma_1) = \Sigma_2$. Since $S^{n-1}_{a,r}$ passes through $\frac{a'}{|a'|^2}$, the center of the sphere w.r.t. which we are performing an inversion, and since $\sigma_a(\frac{a'}{|a'|^2}) = \infty$, then $\infty \in \sigma_a(S^{n-1}_{a,r})$, hence $\sigma_a(S^{n-1}_{a,r}) = P(b, k) \cup \{\infty\}$. Moreover $k = 0$, indeed it is easy to verify that $\sigma(a') = 0$.

What is left to prove is just that $b = \frac{a}{|a|}$. Let us consider the straight line $\operatorname{span}(a)$, see again figure 13.9. Clearly it is orthogonal to $S^{n-1}_{\frac{a'}{|a'|^2}, r'_a}$, since circular inversions act radially, it holds that $\sigma_a(\operatorname{span}(a)) = \operatorname{span}(a)$. Furthermore $\operatorname{span}(a)$ is also orthogonal to $S^{n-1}_{a,r}$. By corollary 12.4.4 we have that the orthogonality between $S^{n-1}_{a,r}$ and $\operatorname{span}(a)$ holds also for their images, hence $\sigma_a(S^{n-1}_{a,r})$ and $\sigma_a(\operatorname{span}(a))$ are orthogonal, i.e. $P(b, 0) \cup \{\infty\}$ is orthogonal to $\operatorname{span}(a)$, so $b = \frac{a}{|a|}$.

3. Finally let Σ_1 and Σ_2 be two m-spheres in $\mathbb{R}^n$ orthogonal to S^{n-1}, i.e. $\Sigma_1 = S^{n-1}_{a,r} \cap V_{m+1}$ and $\Sigma_1 = S^{n-1}_{b,s} \cap V_{m+1}$, with $|a|^2 = r^2 + 1$ and $|b|^2 = s^2 + 1$. As we did in point **2.** it is not restrictive to consider the intersection with the same $(m+1)$-plane. Moreover it is not restrictive to consider the center b as an element of $\operatorname{span}(a)$, indeed this configuration is obtained after a rotation which is an element of $\mathrm{O}(n) \subset \mathcal{M}(\mathcal{B}^n)$. Let us repeat the same construction of point **2.** for both the spheres, hence we define $a' = \frac{|a|-r}{|a|}$, $b' = \frac{|b|-s}{|b|}$, $r'_a = \sqrt{\frac{1}{|a'|^2} - 1}$, $r'_b = \sqrt{\frac{1}{|b'|^2} - 1}$, $\sigma_a = \sigma_{\frac{a'}{|a'|^2}, r'_a}$ and $\sigma_b = \sigma_{\frac{b'}{|b'|^2}, r'_b}$. We know that $\sigma_a, \sigma_b \in \mathcal{M}(\mathcal{B}^n)$ and $\sigma_a(S^{n-1}_{a,r}) = \sigma_b(S^{n-1}_{b,s}) = P\left(\frac{a}{|a|}, 0\right) \cup \{\infty\}$, thus $\sigma_b^{-1} \circ \sigma_a(\Sigma_1) = \Sigma_2$.

$\square$

[3]Notice that it is not possible that $a \in \mathcal{B}^n$, indeed a is the center of a sphere $S^{n-1}_{a,r}$ orthogonal to S^{n-1}, thus the inversion $\sigma_{a,r}$, by theorem 12.5.5, belongs to $\mathcal{M}(\mathcal{B}^n)$. This means that $\sigma_{a,r}(\mathcal{B}^n) = \mathcal{B}^n$ and $\sigma(a) = \infty \notin \mathcal{B}^n$, hence $a \notin \mathcal{B}^n$.

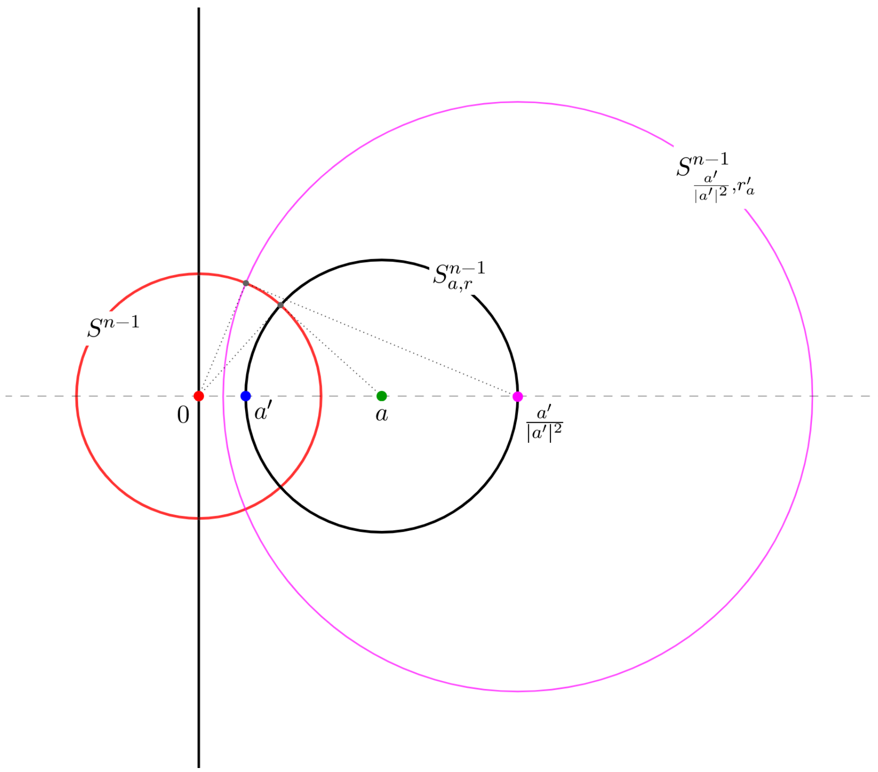


Figure 13.9: Bidimensional representation of the points and spheres involved in the proof of theorem 13.2.6.

We are going now to proceed similarly as we did for $\mathcal{H}^n$ in lemma 13.1.3. The following lemma will be functional to characterize the geodesics of this model.

Lemma 13.2.3 *Let $x, y \in \mathcal{B}^n$ be two linearly dependent distinct points. Then, $z \in \mathcal{B}^n$ verifies*

$$d_{\mathcal{B}}(x, y) = d_{\mathcal{B}}(x, z) + d_{\mathcal{B}}(z, y), \tag{13.29}$$

if and only if z lies between x and y, i.e. it exists $t \in [0, 1]$ such that $z = tx + (1 - t)y$.

Proof. First of all notice that $x, y \in \mathcal{B}^n$ are supposed to be linearly dependent, i.e. they are collinear with 0.

In the following we will always suppose 0 to belong to the segment joining x and y. Indeed, if 0 is not posed between x and y, there exists an element of $\mathcal{M}(\mathcal{B}^n)$, mapping 0 into a point w belonging to the segment joining x and y, and vice-versa. This map is defined as $\sigma_w = \sigma_{\frac{w}{|w|^2}, r}$, with $r = \sqrt{\frac{1}{|w|^2} - 1}$, we have already adopted this technique several times. After a straightforward computation, it is easy to verify that $\sigma_w(w) = 0$. Moreover σ_w is stable on the line joining x and y, by theorem 13.2.3 it is an isometry w.r.t. $d_{\mathcal{B}}$, hence it preserves all the distances between x, y, z.

$\boxed{\Longleftarrow}$: Let us suppose $z \in \mathcal{B}^n$ to lie between x and y. We start from the simplest case $z = 0$. It is evident, see figure 13.10, that $\frac{x}{|x|} = -\frac{y}{|y|}$. Recalling theorem 13.2.5, the following chain of equalities holds:

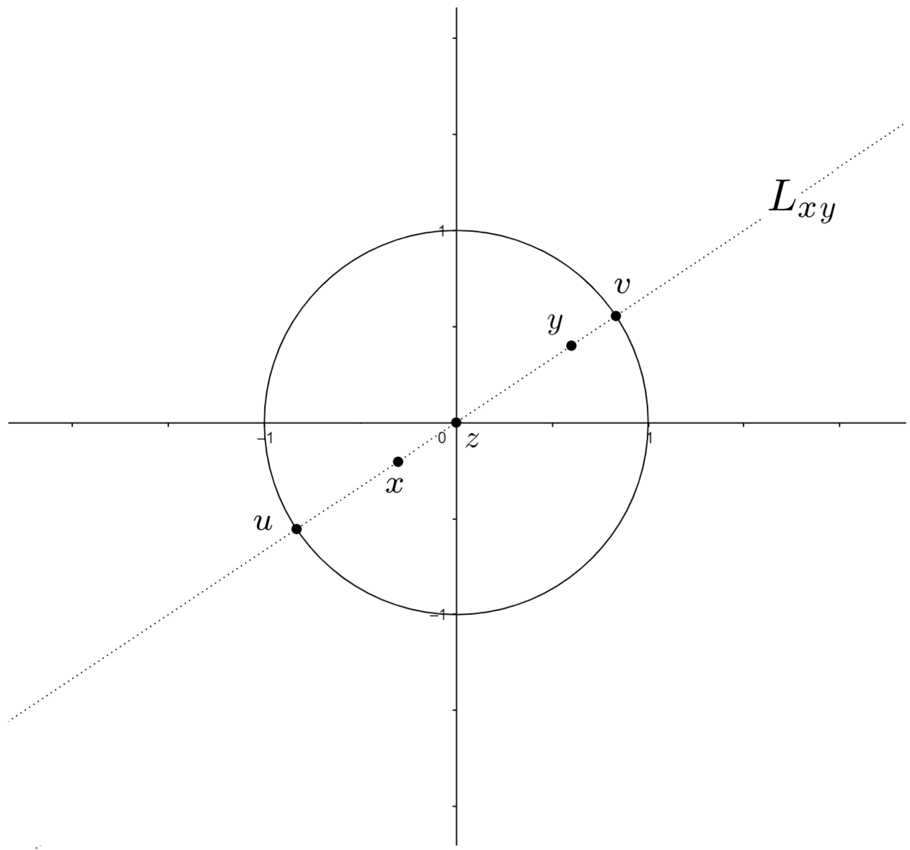


Figure 13.10: Depiction of the points $x, y, u = \frac{x}{|x|} = -\frac{y}{|y|}, v = -\frac{x}{|x|} = \frac{y}{|y|}$, aligned on $L_{x,y}$.

$$
\begin{aligned}
d_{\mathcal{B}}(x,y) &= \log[x,y,u,v] = \log\left(\left[x, y, \frac{x}{|x|}, -\frac{x}{|x|}\right]\right) \\
&= \log\left(\frac{\left|x - \frac{x}{|x|}\right|}{\left|x + \frac{x}{|x|}\right|}\frac{\left|y + \frac{x}{|x|}\right|}{\left|y - \frac{x}{|x|}\right|}\right) \\
&= \log\left(\frac{\left|x - \frac{x}{|x|}\right|}{\left|x + \frac{x}{|x|}\right|}\frac{\left|y - \frac{y}{|y|}\right|}{\left|y + \frac{y}{|y|}\right|}\right) \\
&= \log\left(\frac{\left|x - \frac{x}{|x|}\right|}{\left|x + \frac{x}{|x|}\right|}\right) + \log\left(\frac{\left|y - \frac{y}{|y|}\right|}{\left|y + \frac{y}{|y|}\right|}\right) \\
&= \log\left(\frac{\left|x - \frac{x}{|x|}\right|}{\left|x + \frac{x}{|x|}\right|}\frac{\left|0 + \frac{x}{|x|}\right|}{\left|0 - \frac{x}{|x|}\right|}\right) + \log\left(\frac{\left|0 + \frac{y}{|y|}\right|}{\left|0 - \frac{y}{|y|}\right|}\frac{\left|y - \frac{y}{|y|}\right|}{\left|y + \frac{y}{|y|}\right|}\right) \\
&= \log\left(\left[x, 0, \frac{x}{|x|}, -\frac{x}{|x|}\right]\right) + \log\left(\left[0, y, -\frac{y}{|y|}, \frac{y}{|y|}\right]\right) \\
&= d_{\mathcal{B}}(x,0) + d_{\mathcal{B}}(0,y).
\end{aligned}
$$

Let us consider now $z \neq 0$ lying between x and y. As we did in the proof of lemma 13.2.2, we use the circular inversion $\sigma_z = \sigma_{\frac{z}{|z|^2},r} \in \mathcal{M}(\mathcal{B}^n)$, with $r = \sqrt{\frac{1}{|z|^2} - 1}$, to map z in 0 and preserve all the hyperbolic distances. Therefore σ_z is stable on the line joining x and y.

Hence the following equalities hold:

$$\begin{aligned} d_{\mathcal{B}}(x,y) &= d_{\mathcal{B}}(\sigma_z(x), \sigma_z(y)) = d_{\mathcal{B}}(\sigma_z(x), 0) + d_{\mathcal{B}}(0, \sigma_z(y)) \\ &= d_{\mathcal{B}}(\sigma_z(x), \sigma_z(z)) + d_{\mathcal{B}}(\sigma_z(z), \sigma_z(y)) \\ &= d_{\mathcal{B}}(x,z) + d_{\mathcal{B}}(z,y). \end{aligned}$$

$\boxed{\Longrightarrow}$: Conversely, let us suppose that z does not belong to the segment joining x and y. Our aim is to prove that in this configuration equation (13.2.3) does not hold.

A first, trivial, case is when z is linearly dependent with x and y, but lies outside the segment joining x and y. It is not restrictive to consider e.g. y lying between x and z. Then, by the first implication of this theorem we have that $d_{\mathcal{B}}(x,z) = d_{\mathcal{B}}(x,y) + d_{\mathcal{B}}(y,z)$, hence

$$d_{\mathcal{B}}(x,y) = d_{\mathcal{B}}(x,z) - d_{\mathcal{B}}(z,y) < d_{\mathcal{B}}(x,z) + d_{\mathcal{B}}(z,y),$$

where the last inequality holds because $d_{\mathcal{B}}$ is positive definite.

Let us suppose that z does not belong to the line joining x and y. Let us consider the Euclidean orthogonal projection $\tilde{z}$ of z onto the line joining x and y. Let us suppose that $\tilde{z}$ belongs to the segment joining x and y. By the first part of this theorem we know that $d_{\mathcal{B}}(x,y) = d_{\mathcal{B}}(x,\tilde{z}) + d_{\mathcal{B}}(\tilde{z},y)$. Our aim now is to prove that $d_{\mathcal{B}}(x,\tilde{z}) < d_{\mathcal{B}}(x,z)$ and, analogously $d_{\mathcal{B}}(\tilde{z},y) < d_{\mathcal{B}}(z,y)$.

Our strategy is to lead us back to the Euclidean case. From the considerations we made after theorem 13.2.1, see also figure 13.5, we know that, in the conformal model $\mathcal{B}^n$, hyperbolic spheres or balls centered in the origin are Euclidean spheres or balls centered in the origin. This brings us to the intuition that the assertion still holds if we drop off the condition of being centered in the origin, indeed hyperbolic spheres or balls are Euclidean spheres or balls. This is easy to verify. Let us start by defining the hyperbolic sphere as:

$$S_{x,z} := S^{n-1}_{x,d_{\mathcal{B}}(x,z)} = \{w \in \mathcal{B}^n : d_{\mathcal{B}}(w,x) = d_{\mathcal{B}}(x,z)\}.$$

Since $d_{\mathcal{B}}(x,z)$ is a positive constant, we can set it to k. From 13.2.1 we have that

$$\begin{aligned} S_{x,z} = S^{n-1}_{x,k} &= \{w \in \mathcal{B}^n : d_{\mathcal{B}}(w,x) = k\} \\ &= \left\{ w \in \mathcal{B}^n : 1 + \frac{1|w-x|}{(1-|w|^2)(1-|x|^2)} = \cosh(k) \right\} \\ &= \{w \in \mathcal{B}^n : (h+1)|w|^2 - 2\langle x, w\rangle + |x|^2 - h = 0\}, \end{aligned}$$

where we set to h the constant given by $h = \frac{(1-|x|^2)(\cosh(k)-1)}{2}$ to simplify the computation. Since $h+1 \neq 0$, equation $(h+1)|w|^2 - 2\langle x, w\rangle + |x|^2 - h = 0$ represents a Euclidean sphere. Hence the hyperbolic spheres are Euclidean spheres. The same holds for the open balls, it can be proven analogously. Thence we can define the open ball centered in x of radius $d_{\mathcal{B}}(x,z)$ as

$$B_{x,z} = \{w \in \mathcal{B}^n : d_{\mathcal{B}}(w,x) < d_{\mathcal{B}}(x,z)\}.$$

Since $B_{x,z}$ is a Euclidean ball, it holds that $\tilde{z} \in B_{x,z}$, see figure 13.11. Hence $d_{\mathcal{B}}(\sigma_x(\tilde{z}), 0) = d_{\mathcal{B}}(\tilde{z}, x) < d_{\mathcal{B}}(x,z) = d_{\mathcal{B}}(0, \sigma_x(z))$. Repeating the procedure analogously for y, hence defining in the same way $S_{y,z}$ and $B_{y,z}$ and so on, we obtain that $\tilde{z} \in B_{y,z}$. $d_{\mathcal{B}}(\tilde{z}, y) < d_{\mathcal{B}}(y, z)$. Hence $d_{\mathcal{B}}(x, y) = d_{\mathcal{B}}(x, \tilde{z}) + d_{\mathcal{B}}(\tilde{z}, y) < d_{\mathcal{B}}(x, z) + d_{\mathcal{B}}(z, y)$.

For the case in which $\tilde{z}$ does not belong to the segment joining x and y, the conclusion is evident from some consideration we made above in this proof. In fact $d_{\mathcal{B}}(x, y) < d_{\mathcal{B}}(x, \tilde{z}) + d_{\mathcal{B}}(\tilde{z}, y) < d_{\mathcal{B}}(x, z) + d_{\mathcal{B}}(z, y)$. □

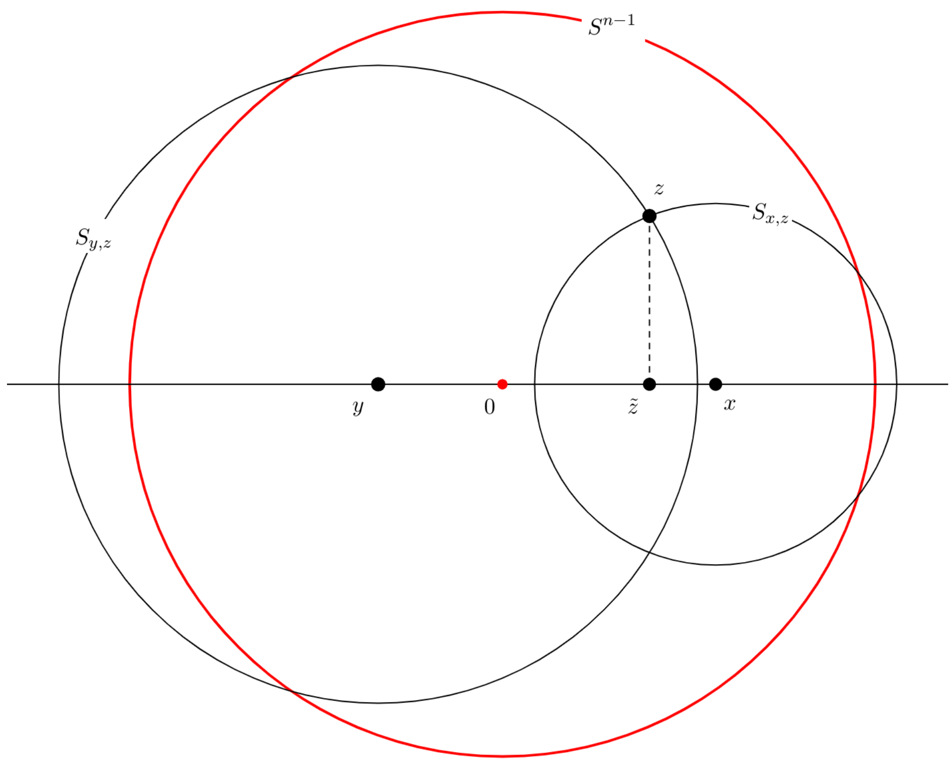


Figure 13.11: Graphic representation in dimension 2 of the construction performed in the proof of lemma 13.2.3.

As we have seen in lemma 13.1.3 concerning the hyperboloid model $\mathcal{H}^n$, if three x, y, z points belonging to a certain hyperbolic space, satisfy equation given in 13.1.3 (or in lemma 13.2.3 as well) then they are hyperbolically collinear, so they belong to the same geodesic[4] of the model.

This means that the diameters of $\mathcal{B}^n$, i.e. the 1-planes orthogonal to S^{n-1}, are geodesics for $(\mathcal{B}^n, d_{\mathcal{B}})$. Together with the statement of lemma 13.2.3 this implies that we can obtain exhaustively all the other possible geodesics of $(\mathcal{B}^n, d_{\mathcal{B}})$, through the action of $\mathcal{M}(\mathcal{B}^n)$. Indeed the elements of $\mathcal{M}(\mathcal{B}^n)$, acting isometrically, map geodesics into geodesics, hence the other only other possible kinds geodesics are the 1-spheres orthogonal to S^{n-1}.

Theorem 13.2.7 *The geodesics arcs of the conformal ball model are the* 1*-planes and* 1*-spheres orthogonal to* S^{n-1}.

[4]Being hyperbolically collinear by definition 13.1.7 means belonging to the same hyperbolic lines, furthermore corollary 13.1.6 states that the hyperbolic lines of $\mathcal{H}^n$ coincide with the geodesics of the model.

We could have shown this theorem and the previous lemma as direct consequences of the definition of $d_{\mathcal{B}}$ as related to $d_{\mathcal{H}}$ through the isometry ζ. For instance, lemma 13.2.3 has a one line proof as a consequence of theorem 13.1.4, and the statement of theorem 13.2.7 descends from the fact that, as an isometry, ζ maps geodesics of $\mathcal{B}^n$ into geodesics of $\mathcal{H}^n$, hence the proof in this case could have been very simple, consisting in showing that the image through ζ of circles and diameters orthogonal to S^{n-1} are geodesics of $\mathcal{H}^n$.

However we preferred to adopt the equivalent expression for the metric $d_{\mathcal{B}}$ given in theorem 13.2.1, which makes $\mathcal{B}^n$ a model of hyperbolic geometry that stands by itself, almost independently from $\mathcal{H}^n$.

13.2.5 The upper-half space $\mathcal{U}^n$

This section is about the metric structure of the second conformal model of hyperbolic geometry: the upper-half space $\mathcal{U}^n$. In section 12.5 we analyzed the link between the two conformal models $\mathcal{B}^n$ and $\mathcal{U}^n$. In particular let us recall the definition of the so-called standard transformation η, see 12.5.1, i.e. the bijection between $\mathcal{U}^n$ and $\mathcal{B}^n$:

$$\begin{array}{rcccl} \eta: & \mathcal{U}^n & \xrightarrow{\sim} & \mathcal{B}^n & \\ & x & \longmapsto & \eta(x) & := \sigma_{e_n,\sqrt{2}} \circ \rho_{e_n,0}(x) \\ & & & & = e_n + \frac{2}{|\rho_{e_n,0}(x)-e_n|^2}(\rho_{e_n,0}(x) - e_n) \\ & & & & = e_n + \frac{2}{|x+e_n|^2}(\rho_{e_n,0}(x) - e_n) \\ & & & & = e_n + \frac{2}{|x+e_n|^2}(x - \langle x, e_n\rangle e_n - e_n) \\ & & & & = e_n + \frac{2}{|x+e_n|^2}(x - \langle x, e_n\rangle e_n - \langle e_n, e_n\rangle e_n) \\ & & & & = e_n + \frac{2}{|x+e_n|^2}(x - \langle x + e_n, e_n\rangle e_n) \\ & & & & = \left(\frac{2x_1}{|x+e_n|^2}, \ldots, \frac{2x_{n-1}}{|x+e_n|^2}, 1 - \frac{2\langle x+e_n,e_n\rangle}{|x+e_n|^2}\right) \\ & & & & = \left(\frac{2x_1}{|x+e_n|^2}, \ldots, \frac{2x_{n-1}}{|x+e_n|^2}, 1 - \frac{2(x_n+1)}{|x+e_n|^2}\right). \end{array}$$

Notice that, to rewrite explicitly the expression of η, we used the fact that, as a reflection, $\rho_{e_n,0}$ is a Euclidean isometry, and that $e_n = \rho_{e_n,0}(-e_n)$, hence

$$|\rho_{e_n,0}(x) - e_n| = |\rho_{e_n,0}(x) - \rho_{e_n,0}(-e_n)| = |x + e_n|. \tag{13.30}$$

For later purposes, we will need the following equality, obtained from the definition above after a straightforward computation:

$$1 - |\eta(x)|^2 = \frac{4x_n}{|x + e_n|^2}. \tag{13.31}$$

In the same way as $\mathcal{B}^n$ inherited naturally its metric structure from $(\mathcal{H}^n, d_{\mathcal{H}})$**,** $\mathcal{U}^n$ **inherits all the metric properties from** $(\mathcal{B}^n, d_{\mathcal{B}})$. Indeed the metric $d_{\mathcal{U}}$ is naturally defined forcing η to be an isometry between $(\mathcal{B}^n, d_{\mathcal{B}})$ and $(\mathcal{U}^n, d_{\mathcal{U}})$.

Def. 13.2.3 *Let $x, y \in \mathcal{U}^n$, the hyperbolic metric $d_{\mathcal{U}}$ on $\mathcal{U}^n$ is given by*

$$d_{\mathcal{U}}(x, y) = d_{\mathcal{B}}(\eta(x), \eta(y)).$$

Similarly to the case of $\mathcal{B}^n$, see theorem 13.2.1, there is an equivalent, elegant reformulation of $d_{\mathcal{U}}$.

Theorem 13.2.8 *For all $x, y \in \mathcal{U}^n$ the hyperbolic metric $d_{\mathcal{U}}$ of the upper-half space satisfies*

$$\cosh(d_{\mathcal{U}}(x,y)) = 1 + \frac{|x-y|^2}{2x_n y_n}.$$

Proof. From the definition of η and equation (13.30) we obtain the following preliminary result:

$$|\eta(x) - \eta(y)| = \frac{2|\rho_{e_n,0}(x) - \rho_{e_n,0}(y)|}{|\rho_{e_n,0}(x) - e_n||\rho_{e_n,0}(y) - e_n|} = \frac{2|x-y|}{|x+e_n||y+e_n|}. \tag{13.32}$$

Thence,

$$\begin{aligned}
\cosh(d_{\mathcal{U}}(x,y)) \underset{\text{def.13.2.3}}{=} & \cosh(d_{\mathcal{B}}(\eta(x), \eta(y))) \\
\underset{\text{thm.13.2.1}}{=} & 1 + \frac{2|\eta(x)-\eta(y)|^2}{(1-|\eta(x)|^2)(1-|\eta(y)|^2)} \\
\underset{\text{eq.(13.31)}}{=} & 1 + \frac{2|\eta(x)-\eta(y)|^2}{\frac{4x_n}{|x+e_n|^2}\frac{4y_n}{|y+e_n|^2}} = 1 + \frac{|\eta(x)-\eta(y)|^2|x+e_n|^2|y+e_n|^2}{8x_n y_n} \\
\underset{\text{eq.(13.32)}}{=} & 1 + \frac{4|x-y|^2|x+e_n|^2|y+e_n|^2}{8x_n y_n|x+e_n|^2|y+e_n|^2} \\
= & 1 + \frac{|x-y|^2}{2x_n y_n}.
\end{aligned}$$

□

Since η is a Möbius transformation, all the results concerning the unit ball model can be transported to $(\mathcal{U}^n, d_{\mathcal{U}})$. The most relevant ones are listed below. The proof of those assertions is a direct consequence of the analogous results for $\mathcal{B}^n$ after a composition with η, which acts isometrically between $(\mathcal{U}^n, d_{\mathcal{U}})$ and $(\mathcal{B}^n, d_{\mathcal{B}})$, hence translates all the metric properties from one model to the other.

Theorem 13.2.9 *The following properties of the hyperbolic model $(\mathcal{U}^n, d_{\mathcal{U}})$ hold:*

1. *The isometry group of $\mathcal{U}^n$ is isomorphic to its Möbius group $\mathcal{M}(\mathcal{U}^n)$.*
2. *The geodesics of $\mathcal{U}^n$ are lines and circles orthogonal to its border $\partial\mathcal{U}^n \cong \hat{\mathbb{R}}^{n-1}$.*
3. *Let $x, y \in \mathcal{U}^n$, then we have the following equivalent expression for the metric $d_{\mathcal{U}}$*

$$d_{\mathcal{U}}(x,y) = \log([x,y,u,v]),$$

where u and v are the points of intersection between the geodesic $\gamma_{x,y}$, passing through x and y, and $\hat{\mathbb{R}}^{n-1}$.

Notice that since the circles must be orthogonal to $\hat{\mathbb{R}}^{n-1}$, by condition **2.** in definition 12.5.4, their center is supposed to belong to $\hat{\mathbb{R}}^{n-1}$, hence their are actually half-circles.

Furthermore, considering the formula given in point **3.** of the previous theorem, we must stress that it holds also when the geodesic $\gamma_{x,y}$ is a straight half-line, indeed in that case u is the point of intersection between $\gamma_{x,y}$ and $\hat{\mathbb{R}}^{n-1}$ and $v = \infty$. In that case we will need to use the definition of cross-ratio involving the chordal metric, see 12.4.2.

As we did for both $\mathcal{B}^n$ and $\mathcal{H}^n$, also here we are going to introduce the concept of element of hyperbolic arc length w.r.t. $d_{\mathcal{U}}$.

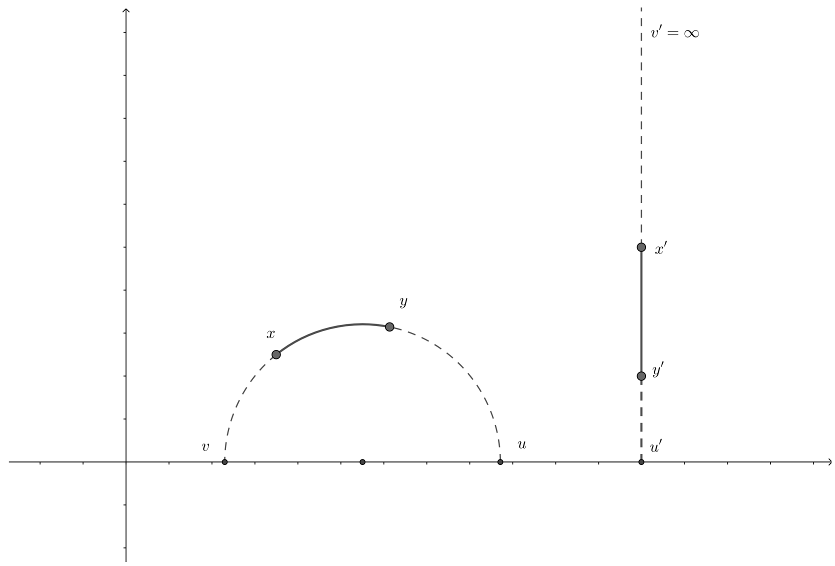


Figure 13.12: The two types of geodesic lines of the upper-half plane: half-circles and Euclidean half-lines orthogonal to $\hat{\mathbb{R}}^{n-1}$.

Theorem 13.2.10 *The element of hyperbolic arc length $||dx||_{\mathcal{U}}$ of the conformal model $(\mathcal{U}^n, d_{\mathcal{U}})$ is given by:*

$$||dx||_{\mathcal{U}} = \frac{|dx|}{x_n}.$$

Proof. In this proof we are going to proceed analogously as we did in the proof of theorem 13.2.2 for $\mathcal{B}^n$.

Let $x \in \mathcal{U}^n$ and $y = \eta(x)$. Since η acts isometrically between $(\mathcal{U}^n, d_{\mathcal{U}})$ and $(\mathcal{B}^n, d_{\mathcal{B}})$, the natural way to define the element of hyperbolic arc length on $\mathcal{U}^n$ is $||dx||_{\mathcal{U}} = ||dy||_{\mathcal{B}}$.

By theorem 13.2.2 we already know that

$$||dy||_{\mathcal{B}} = \frac{2|dy|}{1 - |y|^2}. \tag{13.33}$$

Furthermore from the definition of η, we have that

$$y_i = \frac{2x_i}{|x + e_n|^2} \qquad \text{for } i = 1, \ldots, n-1,$$

and

$$y_n = 1 - \frac{2(x_n + 1)}{|x + e_n|^2}\ .$$

Our aim is to compute dy_i, for all $i = 1, \ldots, n$. To do that let us first recall that, from formula (3.6), we have

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

Let $i = 1, \ldots, n-1$, then if $j \neq i$ and $j \neq n$, we obtain

$$\frac{\partial y_i}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{2x_i}{|x + e_n|^2}\right) = -\frac{4x_i x_j}{|x + e_n|^4}.$$

If $j = i$ we have

$$\frac{\partial y_i}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{2x_i}{|x+e_n|^2}\right) = 2\frac{|x+e_n|^2 - 2x_i^2}{|x+e_n|^4} = \frac{2}{|x+e_n|^2} - \frac{4x_i^2}{|x+e_n|^4}.$$

while, when $j = n$, it holds

$$\frac{\partial y_i}{\partial x_n} = \frac{\partial}{\partial x_n}\left(\frac{2x_i}{|x+e_n|^2}\right) = -\frac{4x_i(x_n+1)}{|x+e_n|^4}.$$

So for $i = 1, \ldots, n$

$$\begin{aligned}
dy_i &= \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_j} dx_j = \frac{\partial y_i}{\partial x_i} dx_i + \sum_{i\neq j=1}^{n-1} \frac{\partial y_i}{\partial x_j} dx_j + \frac{\partial y_n}{\partial x_n} dx_n \\
&= \left(\frac{2}{|x+e_n|^2} - \frac{4x_i^2}{|x+e_n|^4}\right) dx_i - \sum_{i\neq j=1}^{n-1} \frac{4x_i x_j}{|x+e_n|^4} dx_j - \frac{4x_i(x_n+1)}{|x+e_n|^4} dx_n \\
&= \frac{2dx_i}{|x+e_n|^2} - \frac{4x_i}{|x+e_n|^4} \sum_{j=1}^{n-1} x_j dx_j - \frac{4x_i(x_n+1)}{|x+e_n|^4} dx_n \\
&= \frac{2dx_i}{|x+e_n|^2} - \frac{4x_i \langle x+e_n, dx\rangle}{|x+e_n|^4}.
\end{aligned}$$

Let $i = n$. For $j = 1, \ldots, n-1$, we have

$$\frac{\partial y_n}{\partial x_j} = \frac{\partial}{\partial x_i}\left(1 - \frac{2(x_n+1)}{|x+e_n|^2}\right) = \frac{4(x_n+1)x_j}{|x+e_n|^4}.$$

For $j = n$

$$\frac{\partial y_n}{\partial x_n} = \frac{\partial}{\partial x_n}\left(1 - \frac{2(x_n+1)}{|x+e_n|^2}\right) = -\frac{2}{|x+e_n|^2} + \frac{4(x_n+1)^2}{|x+e_n|^4}.$$

Again from formula (3.6) we obtain:

$$\begin{aligned}
dy_n &= \sum_{j=1}^{n} \frac{\partial y_n}{\partial x_j} dx_j = \sum_{j\neq n} \frac{\partial y_n}{\partial x_j} dx_j + \frac{\partial y_n}{\partial x_n} dx_n \\
&= \sum_{j\neq n} \frac{4(x_n+1)x_j dx_j}{|x+e_n|^4} + \left(-\frac{2}{|x+e_n|^2} + \frac{4(x_n+1)^2}{|x+e_n|^4}\right) dx_n \\
&= \frac{4(x_n+1)\langle x+e_n, dx\rangle}{|x+e_n|^4} - \frac{2dx_n}{|x+e_n|^2}.
\end{aligned}$$

Hence

$$\begin{aligned}
|dy|^2 &= dy_n^2 + \sum_{i\neq n}^{d} y_i^2 \\
&= \left(-\frac{2dx_n}{|x+e_n|^2} + \frac{4(x_n+1)\langle dx, x+e_n\rangle}{|x+e_n|^4}\right)^2 + \sum_{i=1}^{n-1}\left(\frac{2dx_i}{|x+e_n|^2} - \frac{4x_i\langle dx, x+e_n\rangle}{|x+e_n|^4}\right)^2 \\
&= \frac{4}{|x+e_n|^4}\left(|dx|^2 - \frac{4\langle dx, x+e_n\rangle^2}{|x+e_n|^2} + \frac{4|x+e_n|^2\langle dx, x+e_n\rangle^2}{|x+e_n|^4}\right) \\
&= \frac{4|dx|^2}{|x+e_n|^4}.
\end{aligned}$$

Thus

$$|dy| = \frac{2|dx|}{|x+e_n|^2}.$$

Furthermore, recalling that equation 13.31 stated that $1-|y|^2 = \frac{4x_n}{|x+e_n|^2}$, we can conclude the proof as follows:

$$\|dx\|_{\mathcal{U}} = \|dy\|_{\mathcal{B}} = \frac{2|dy|}{1-|y|^2} = \frac{4|dx|}{|x+e_n|^2}\frac{1}{1-|y|^2} = \frac{4|dx|}{|x+e_n|^2}\frac{|x+e_n|^2}{4x_n} = \frac{|dx|}{x_n}.$$

□

Notice that the arc length of the upper half space model has the easiest analytical form among the hyperbolic models.

13.3 The projective hyperbolic model $\mathcal{K}^n$

The projective model $\mathcal{K}^n$, as its name suggests, is obtained **mapping the hyperboloid model $\mathcal{H}^n$ in the projective space** $\mathbb{RP}^n$. The 2-dimensional version of this model is the so-called **Klein disk** (in fact the $\mathcal{K}$ is for Klein). Actually we have already implicitly seen this model in section 10.4 as an example of homogeneous space. In this section we will equip it with a hyperbolic metric.

13.3.1 Link between $\mathcal{K}^n$ and $\mathcal{H}^n$

Let us start by recalling some notions we will need in the following, see section 10.4 for further details.

The n-dimensional projective space is defined as $\mathbb{RP}^n = \mathbb{R}^{n+1} \diagup_{\mathbb{R}^\times}$. Given$x \in \mathbb{R}^{n+1}$ and $\lambda \in \mathbb{R}^\times$, the elements of $\mathbb{RP}^n$ are the equivalence classes $x \cdot \mathbb{R}^\times = [x] = [\lambda x]$. The following map is the canonical projection from $\mathbb{R}^{n+1}$ onto $\mathbb{RP}^n$:

$$\begin{array}{rccc} p: & \mathbb{R}^{n+1} & \longrightarrow & \mathbb{RP}^n \\ & x & \longmapsto & [x]. \end{array}$$

The so-called projective linear group, defined as $\mathrm{PGL}(n+1,\mathbb{R}) = \mathrm{GL}(n+1,\mathbb{R}) \diagup_{\mathbb{R}^\times}$ acts transitively on it $\mathbb{RP}^n$.

Analogously to what we did with the bijection ζ used for the conformal ball model, we will proceed in our intent by searching for a map κ that connects $\mathcal{K}^n$ with $\mathcal{H}^n$ bijectively. The hyperbolic metric $d_{\mathcal{K}}$ will be obtained by forcing κ to be an isometry.

First of all, it must be clearly stated that, **as a set, the projective model $\mathcal{K}^n$ is equal to the conformal ball model $\mathcal{B}^n$**, i.e.

$$\mathcal{K}^n = \{x \in \mathbb{R}^n : |x| < 1\} \subset \mathbb{R}^n,$$

however, we will see that **the metric spaces** $(\mathcal{K}^n, d_{\mathcal{K}})$ and $(\mathcal{B}^n, d_{\mathcal{B}})$, **though isometric, they are different**. As we did for $\mathcal{B}^n$, to define the map that links $\mathcal{K}^n$ to $\mathcal{H}^n$, it is convenient to describe $\mathcal{K}^n$ as a subset of $\mathbb{R}^{n+1}$.

$$\mathcal{K}^n \cong \mathcal{K}^n_0 \left\{x \in \mathbb{R}^{n+1} : |\bar{x}| < 1 \text{ and } x_{n+1} = 0\right\} \subset \mathbb{R}^{n+1},$$

where we recall that, for all $x \in \mathbb{R}^{n+1}$, $\bar{x} = (x_1, \ldots, x_n)^t$ and that the hyperboloid $\mathcal{H}^n$ is given by $\mathcal{H}^n = \{x \in \mathbb{R}^{n+1} : ||x||^2 = -1 \text{ and } x_{n+1} > 0\} \subset \mathbb{R}^{n+1}$, $\|\ \|$ being the Lorentz norm.

As we have already seen in equation (10.7), it is natural define the injection ι of $\mathbb{R}^n$ into $\mathbb{RP}^n$ as:

$$\begin{array}{rccc} \iota : & \mathbb{R}^n & \hookrightarrow & \mathbb{RP}^n \\ & u & \longmapsto & \begin{bmatrix} u \\ 1 \end{bmatrix}. \end{array} \tag{13.34}$$

If we restricts ι to $\mathcal{K}$ and identify it with its image via ι, i.e. $\mathcal{K}^n \cong \iota(\mathcal{K}^n)$, we can interpret $\mathcal{K}^n$ as a projective space. Through this identification the elements of $\mathcal{K}^n$ can be seen either as points of $\mathbb{R}^n$, or as an equivalence class in $\mathbb{RP}^n$, made up by elements of $\mathbb{R}^{n+1}$.

Let us derive the expression of the map κ that permits to identify $\mathcal{H}^n$ and $\mathcal{K}^n$. The geometrical construction is depicted in figure 13.13 for $n = 1$: κ is obtained by connecting with a straight line each point of the hyperboloid to the origin 0 and then by considering the intersection between this line and the unit ball placed at height 1 along the $(n+1)$-axis, that we denote with $\mathcal{K}_1^n$ and which is clearly isomorphic to $\mathcal{K}^n$:

$$\mathcal{K}^n \cong \mathcal{K}_0^n \cong \mathcal{K}_1^n = \left\{ x \in \mathbb{R}^{n+1} : |\bar{x}| < 1 \text{ and } x_{n+1} = 1 \right\}.$$

Theorem 10.4.1 of section 10.4 implies that every Euclidean line in time-like cone passing through 0 intersects a unique point of $\mathcal{H}^n$, thus guaranteeing that κ constructed as above is a bijection.

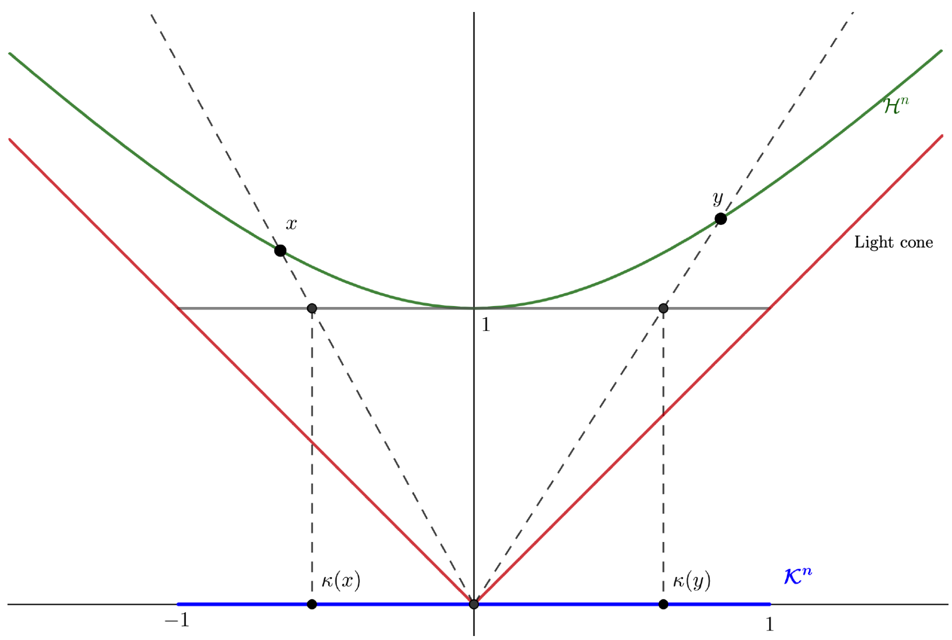


Figure 13.13: The projection κ between $\mathcal{H}^n$ and $\mathcal{K}^n$ for $n = 1$.

To obtain the analytical form of κ, consider a point $x \in \mathcal{H}^n$, then the points lying on the straight line that connects x to 0 are of the form αx, $\alpha \in \mathbb{R}$, thus κ will be defined by the real scalar $\bar{\alpha}$ such that $\kappa(x) = \bar{\alpha} x \in \mathcal{K}_1^n$ which is equivalent to requiring both the conditions $\alpha x_{n+1} = 1$ and $|\bar{\alpha}\bar{x}| < 1 \iff \bar{\alpha}^2 |\bar{x}|^2 < 1$. The first condition implies $\bar{\alpha} = \frac{1}{x_{n+1}}$, well-defined because $x \in \mathcal{H}^n$, so $x_{n+1} > 0$, if we insert this value of $\bar{\alpha}$ in the explicit expression of $\|x\|^2$, which is equal to -1 because $x \in \mathcal{H}^n$, we get: $\|x\|^2 = \bar{\alpha}^2 |\bar{x}|^2 - \frac{1}{x_{n+1}^2} x_{n+1}^2 = -1 \iff \bar{\alpha}^2 |\bar{x}|^2 = 0 < 1$, so the second condition is verified too for $\bar{\alpha} = \frac{1}{x_{n+1}}$.

To resume, the bijection κ from $\mathcal{H}^n$ to $\mathcal{K}^n$ is given by:

$$\begin{array}{rccl} \kappa: & \mathcal{H}^n & \xrightarrow{\sim} & \mathcal{K}^n \\ & x & \longmapsto & \boxed{\kappa(x) = \left(\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right)}. \end{array}$$

Let us now derive the analytical expression of κ^{-1}. By definition of $\mathcal{K}_0^n$ and $\mathcal{K}_1^n$ it is immediate to see that $y \in \mathcal{K}_0^n \iff y + e_{n+1} \in \mathcal{K}_1^n$. $\kappa^{-1}(y)$ is the element of the straight line passing through 0 and $y + e_{n+1}$ that belongs to $\mathcal{H}^n$. To determine it we simply have to find the real constant $\bar{\beta}$ such that $\kappa^{-1}(y) = \bar{\beta}(y+e_{n+1}) \in \mathcal{H}^n$, i.e. $\|\bar{\beta}(y+e_{n+1})\|^2 = \bar{\beta}^2\|y+e_{n+1}\|^2 = -1$, thus $\bar{\beta}^2 = -\frac{1}{\|y+e_{n+1}\|^2}$. Now, since $y + e_{n+1}$ is a time-like vector, $\|y + e_{n+1}\|^2 < 0$, so $\beta = \frac{1}{|\|y+e_{n+1}\||}$.

To resume, the inverse of κ is given by:

$$\begin{array}{rccl} \kappa^{-1}: & \mathcal{K}^n & \xrightarrow{\sim} & \mathcal{H}^n \\ & y & \longmapsto & \kappa^{-1}(y) = \frac{y+e_{n+1}}{|\|y+e_{n+1}\||}. \end{array}$$

13.3.2 The hyperbolic metric of $\mathcal{K}^n$ and its isometry group

Consistently with what we have done with the conformal model, we define the hyperbolic metric on $\mathcal{K}^n$ by setting κ, the bijection between $\mathcal{H}^n$ and $\mathcal{K}^n$, as an hyperbolic isometry between the two spaces.

Def. 13.3.1 *For all $x, y \in \mathcal{K}^n$ the hyperbolic metric $d_{\mathcal{K}}$ on $\mathcal{K}^n$ for is defined as*

$$d_{\mathcal{K}}(x, y) = d_{\mathcal{H}}(\kappa^{-1}(x), \kappa^{-1}(y)).$$

Similarly to the other hyperbolic models, there is a simple reformulation of $d_{\mathcal{K}}$ involving the hyperbolic cosine.

Theorem 13.3.1 *For all $x, y \in \mathcal{K}^n$, the hyperbolic metric $d_{\mathcal{K}}$ satisfies the following equation*

$$\cosh(d_{\mathcal{K}}(x, y)) = \frac{1 - \langle x, y\rangle}{\sqrt{1 - |x|^2}\sqrt{1 - |y|^2}}.$$

Proof. First of all let us recall that for all $x, y \in \mathcal{H}^n$ we have that $-x \circ y = x_n y_n - \sum_{i=1}^{n-1} x_i y_i$:

$$\begin{aligned} \cosh(d_{\mathcal{K}}(x, y)) &\underset{\text{def. 13.3.1}}{=} \cosh(d_{\mathcal{H}}(\kappa^{-1}(x), \kappa^{-1}(y))) \\ &\underset{\text{eq. (13.5)}}{=} -\left(\frac{x + e_{n+1}}{|\|x + e_{n+1}\||}\right) \circ \left(\frac{y + e_{n+1}}{|\|y + e_{n+1}\||}\right) \\ &= \frac{1 - \langle x, y\rangle}{\sqrt{1 - |x|^2}\sqrt{1 - |y|^2}} \end{aligned}$$

□

It is important to stress that this last metric looks like the most complicated one, if compared with the others presented in this chapter. Indeed, the presence of the Euclidean inner product $\langle x, y\rangle$ implies that **the Klein metric depends both on the Euclidean**

distance between points and also on the angle between the two vectors identified by these points.

This is the reason why the projective model is often said to be the model that does not preserve the notion of Euclidean angles, while the Euclidean notion of line is preserved, as we will see later. On the other hand the conformal models do exactly the opposite: they preserve Euclidean angles by conformality, but they have distorted lines (the 1-spheres orthogonal to the border).

In the same way as we did for the conformal models, it is possible to reformulate the expression of the metric using the cross-ratio.

Corollary 13.3.1 *For all $x \in \mathcal{K}^n$ the hyperbolic metric $d_{\mathcal{K}}$ is such that*

$$d_{\mathcal{K}}(0,x) = \frac{1}{2}\log\left(\frac{1+|x|}{1-|x|}\right) = \frac{1}{2}\log\left(\left[x, 0, -\frac{x}{|x|}, \frac{x}{|x|}\right]\right).$$

Proof. For the following computation we need to recall that $\operatorname{arccosh}(x) = \log(x + \sqrt{x^2-1})$. Then

$$\begin{aligned} d_{\mathcal{K}}(0,x) \underset{\text{th. } 13.3.1}{=} & \operatorname{arccosh}\left(\frac{1}{\sqrt{1-|x|^2}}\right) \\ = & \log\left(\frac{1}{\sqrt{1-|x|^2}} + \sqrt{\frac{1}{1-|x|^2} - 1}\right) \\ = & \log\left(\frac{1+|x|}{\sqrt{1-|x|^2}}\right) = \frac{1}{2}\log\left(\frac{1+|x|}{1-|x|}\right) \\ = & \frac{1}{2}\log\left(\left[x, 0, -\frac{x}{|x|}, \frac{x}{|x|}\right]\right), \end{aligned}$$

where the last equality is given by a computation we did in the proof of theorem 13.2.5. □

Recalling corollary 13.2.1, we had that $d_{\mathcal{B}}(x,0) = \log\left(\frac{1+|x|}{\sqrt{1-|x|^2}}\right)$. Hence, for all x in the unit ball as a set, it holds the following relation between the two metrics $d_{\mathcal{B}}$ and $d_{\mathcal{K}}$:

$$d_{\mathcal{B}}(0,x) = \log\left(\left[x, 0, -\frac{x}{|x|}, \frac{x}{|x|}\right]\right) = 2d_{\mathcal{K}}(0,x).$$

Moreover notice that $-\frac{x}{|x|}$ and $\frac{x}{|x|}$ are the two points of intersection between the Euclidean line joining 0 and x, and S^{n-1}. As we will show later, this construction holds for any generic couple of points $x, y \in \mathcal{K}^n$.

Theorem 13.3.2 *The element of hyperbolic arc length on the projective model $\mathcal{K}^n$ is given by*

$$\|dx\|_{\mathcal{K}} = \frac{\sqrt{(1-|x|^2)|dx|^2 + \langle x, dx\rangle^2}}{1-|x|^2}.$$

The proof of this theorem is totally analogous to the proofs of theorem 13.2.2 for $\mathcal{B}^n$ and theorem 13.2.10 for $\mathcal{U}^n$.

Def. 13.3.2 *The action of a projective transformation $\phi \in \mathrm{PGL}(n+1,\mathbb{R})$ on $\mathbb{R}^n$ is defined by*

$$\begin{array}{cccc} \phi : & \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ & x = \begin{bmatrix} x \\ 1 \end{bmatrix} & \longmapsto & \left[\phi\begin{pmatrix} x \\ 1 \end{pmatrix}\right] = \begin{bmatrix} y \\ y_{n+1} \end{bmatrix} = \frac{y}{y_{n+1}}. \end{array}$$

Note that projective transformations are not always well defined on all $\mathbb{R}^n$ since $y_{n+1} = \left\langle \phi\begin{pmatrix} x \\ 1 \end{pmatrix}, e_{n+1} \right\rangle$ can be zero, indeed ι defined in equation (13.34) is not bijective, e.g. $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \notin \iota(\mathbb{R}^n)$.

To characterize the isometry group of $\mathcal{K}^n$ we need to look for a set of projective transformations stable on $\mathcal{K}^n$. To do so we will need the following lemma.

Lemma 13.3.1 *Let $\phi \in \mathrm{GL}(n+1,\mathbb{R})$. Then ϕ leaves the union of the light cone with the future light cone, i.e. the set $\{x \in \mathbb{R}^n : \|x\| \leqslant 1\}$, invariant if and only if there exists a scalar $\lambda > 0$ such that $\lambda\phi$ is a Lorentz transformation.*

Proof.

$\boxed{\Longrightarrow}$: it is clear, by corollary 11.3.1, that if $\phi \in \mathrm{GL}(n+1,\mathbb{R})$, $\lambda > 0$ and $\lambda\phi$ is a Lorentz transformation, then ϕ leaves the light cone invariant.

$\boxed{\Longleftarrow}$: let us now suppose that $\phi \in \mathrm{GL}(n+1,\mathbb{R})$ is such that it leaves the light cone invariant. By continuity, ϕ leaves also the inside of the light cone (i.e. the set time like vectors) invariant and, by the same argument, it also leaves the set of light-like vectors invariant, i.e. the border of the light cone. Hence, since e_{n+1} is time-like, $\phi(e_{n+1})$ is time-like. Furthermore, since $\mathrm{span}(\phi(e_{n+1}))$ is a 1-dimensional time-like vector subspace, by the transitivity of $\mathrm{O}(n,1)$ on the set of 1-dimensional time-like vector subspaces, there is a Lorentz transformation $A \in \mathrm{O}(n,1)$ such that $A(\mathrm{span}(\phi(e_{n+1}))) = \mathrm{span}(e_{n+1})$, i.e.

$$A\phi(e_{n+1}) = \lambda e_{n+1},$$

with $\lambda > 0$. It remains to show that $\lambda^{-1}A\phi \in \mathrm{O}(n+1) \cap \mathrm{PO}(n,1)$, see also corollary 11.3.3. Let $x \in \mathbb{R}^{n+1}$ linearly independent to e_{n+1} and $B_x \in \mathrm{O}(n+1) \cap \mathrm{O}(n,1)$ such that $\tilde{\phi} = \lambda^{-1}B_x A\phi$ leaves $\mathrm{span}(x, e_{n+1})$ invariant and fixes e_{n+1}. Consequently, we may assume $n = 1$ and, because $\tilde{\phi}$ leaves e_{n+1} unchanged, it is of the form

$$\tilde{\phi} = \begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix}.$$

Since $\tilde{\phi}$ is stable on the set of light-like vectors, we have

$$\begin{aligned} \left\|\tilde{\phi}\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\|^2 = 0 = \left\|\tilde{\phi}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\|^2 \quad &\Longleftrightarrow \quad a^2 - (b-1)^2 = a^2 - (b+1)^2 \\ &\Longleftrightarrow \quad b = 0 \text{ and } a = \pm 1. \end{aligned}$$

Hence, $\tilde{\phi} \in \mathrm{O}(n+1) \cap \mathrm{O}(n,1)$, which implies $\lambda^{-1}A\phi \in \mathrm{O}(n+1) \cap \mathrm{O}(n)$ and that $\lambda^{-1}\phi$ is a Lorentz transformation. □

Lemma 13.3.2 *A projective transformation* $[\phi] \in \mathrm{PGL}(n+1, \mathbb{R})$ *leaves* $\mathcal{K}^n$ *invariant if and only if anyone of its elements, as an equivalence class (i.e.* $\lambda\phi$ *with* $|\lambda| > 0$*) leaves the light cone invariant.*

Proof. This proof is a direct consequence of theorem 10.4.1. □

Hence by combining both lemmas we come to the conclusion that every projective transformation that leaves $\mathcal{K}^n$ invariant is the equivalence class of a unique positive Lorentz transformation.

Theorem 13.3.3 *Every isometry of* $\mathcal{K}^n$ *extends to a unique projective transformation that leaves* $\mathcal{K}^n$ *invariant and every projective transformation that leaves* $\mathcal{K}^n$ *invariant can be restricted to an isometry of* $\mathcal{K}^n$*.*

Proof. The isometries of $\mathcal{H}^n$ are its Lorentz transformation and, via the isometry $\kappa : \mathcal{H}^n \to \mathcal{K}^n$, they correspond to the isometries of $\mathcal{K}^n$. Hence, by applying lemma 13.3.1 and lemma 13.3.2, we obtain the result of this theorem. □

Corollary 13.3.2 $\mathcal{I}(\mathcal{K}^n) = \mathrm{PO}(n,1) \diagup_{\mathbb{R}^\times}$.

Corollary 13.3.3 *An isometry of* $\mathcal{K}^n$ *fixes* 0 *if and only if it is the restriction to* $\mathcal{K}^n$ *of an orthogonal transformation of* $\mathbb{R}^n$*.*

While the projective model does not maintain the Euclidean notion of angles, it offers a big advantage compared to other models: the hyperbolic lines of the projective model are exactly the Euclidean lines. This makes this model very useful for convexity arguments.
It is worth to mention that we cannot have a hyperbolic model that retains both the Euclidean lines and the Euclidean angles (i.e. it is conformal). Intuitively, otherwise, i.e. if we preserve both notions of Euclidean lines and angles, we have a model of Euclidean geometry.

Theorem 13.3.4 *The hyperbolic lines of* $\mathcal{K}^n$ *are the Euclidean lines restricted to* $\mathcal{K}^n$*.*

As announced previously, we can give a version of the hyperbolic metric on $\mathcal{K}^n$ using the cross-ratio. In literature, it is also known as the Cayley-Klein metric, see also figure 13.14. This formulation of the hyperbolic metric can also be extended to a general bounded convex set, in that case it is called the Hilbert metric, we will treat it in detail in subsection 13.3.3.

Theorem 13.3.5 *For all* $x, y \in \mathcal{K}^n$ *the hyperbolic metric on* $\mathcal{K}^n$ *is given by:*

$$d_{\mathcal{K}}(x,y) = \frac{1}{2}\log\big([x,y,u,v]\big),$$

where u, v *are the two points of intersection of the Euclidean line* $L_{x,y}$ *and* S^{n-1} *such that* $|x-v| > |y-v|$ *and* $|y-u| > |x-u|$*.*

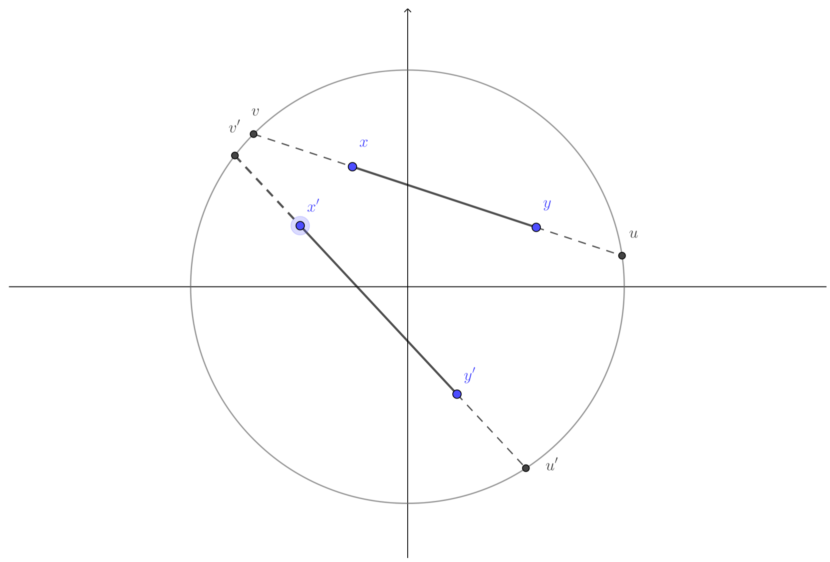


Figure 13.14: Example, in the 2-dimensional of the construction needed to calculate the Cayley-Klein distances between couples of distinct points in $\mathcal{K}^n$.

Notice that $|x - v| > |y - v|$ and $|y - u| > |x - u|$ this last condition is given just to state that u is the one on the side of x and v the one on the side of y.

13.3.3 Birkhoff version of the Hilbert metric on convex sets

In this section we will study a generalisation of the Cayley-Klein metric to any convex set: the Hilbert metric. To do so, we will introduce a second metric Birkhoff metric that will allow us to extend the easily prove that the Hilbert metric is well defined on any convex set. This section is mainly based on [**?**].

Def. 13.3.3 *Let $\Omega \subset \mathbb{R}^n$ be a bounded open convex set, non-empty. We define the Hilbert metric on Ω as*

$$\delta(x, y) = \begin{cases} \log\big([x, y, u, v]\big) & \textit{if } x \neq y \\ 0 & \textit{if } x = y \end{cases}$$

for all $x, y \in \Omega$ and with $u, v \in \partial\Omega$ defined as the points of intersection of the Euclidean line passing through x and y and the border of Ω such that $|x - u| > |y - u|$ and $|y - v| > |x - v|$.

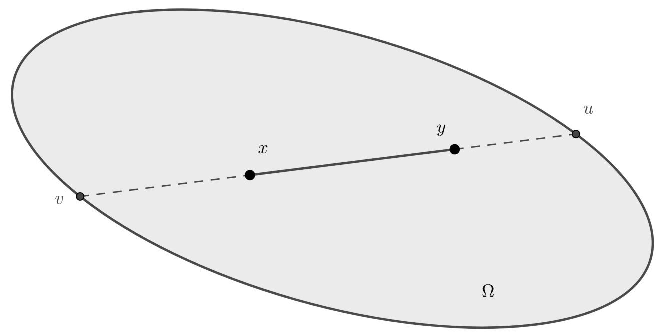


Figure 13.15: Example of the Hilbert metric on a 2−dimensional convex set

To show that the Hilbert metric is well defined we will be using a path different from how the Hilbert metric is usually introduced by introducing Birkhoff's version of the Hilbert metric on cones.

Def. 13.3.4 *Let V be a vector space and $\mathcal{C} \subset V$ a subset. We say that $\mathcal{C}$ is a cone if $\mathcal{C}$ verifies*

1. *$\mathcal{C}$ is convex: $\forall x, y \in \mathcal{C}$, $\lambda \in [0,1]$, $\lambda x + (1-\lambda)y \in \mathcal{C}$*
2. *$\lambda\mathcal{C} \subseteq \mathcal{C}$, for all $\lambda > 0$.*
3. *$\bar{\mathcal{C}} \cap (-\bar{\mathcal{C}}) = \{0\}$*

Mathematical tradition dictates us to make the following remark: by combining 1. and 2. of the definition of a cone we obtain that a cone is stable under addition. In what follows, we shall consider V to be a vector space and $\mathcal{C} \subset V$ a cone. In order to define Birkhoff's metric, we introduce a partial ordering on cone.

Def. 13.3.5 *We define the partial ordering on the $\mathcal{C}$ for $x, y \in \mathcal{C}$ as*

$$x \leqslant_c y \iff y - x \in \mathcal{C}.$$

*Furthermore, we say that y **dominates** x if there exists $\alpha, \beta \in \mathbb{R}$ such that*

$$\alpha y \leqslant_c x \leqslant_c \beta y$$

and the equivalence relationship given by this partial ordering as

$$x \sim_c y \iff y \textbf{ dominates } x \text{ and } x \textbf{ dominates } y.$$

In the case that y dominates x, we note the following quantities:

$$\begin{aligned} M\Big(\frac{x}{y}\Big) &= \inf\{\beta \in \mathbb{R} : x \leqslant_c \beta y\} \\ m\Big(\frac{x}{y}\Big) &= \sup\{\alpha \in \mathbb{R} : \alpha_c \leqslant_c x\} \end{aligned}$$

Lemma 13.3.3 *If $x, y \in \mathcal{C}\backslash\{0\}$, then $x \sim_c y$ if and only if there is $0 < \alpha \leqslant \beta$ such that*

$$\alpha y \leqslant_c x \leqslant \beta y.$$

Moreover, if $x \sim_c y$ we have

$$m\Big(\frac{x}{y}\Big) = \sup\{\alpha > 0 : y \leqslant \frac{1}{\alpha} x\} = M\Big(\frac{y}{x}\Big)^{-1}$$

Def. 13.3.6 *We define the Birkhoff metric on the cone $\mathcal{C}$ as*

$$d(x,y)=\begin{cases}\log\left(\frac{M(\frac{x}{y})}{m(\frac{x}{y})}\right) & \text{if } x\sim_c y \text{ and } y,x\neq 0\\ 0 & \text{if } x=y=0\\ \infty & \text{otherwise}\end{cases}$$

Theorem 13.3.6 *Let $x,y,z\in\mathcal{C}\backslash\{0\}$ such that $x\sim_c y\sim_c z$. Then,*

1. $d(x,y)\geqslant 0$
2. $d(x,y)=d(y,x)$
3. $d(x,z)\leqslant d(x,y)+d(y,z)$
4. $d(x,y)=d(\lambda x,\mu y)$ *for all* $\lambda,\mu>0$

Moreover, if V is a Banach space, then $d(x,y)=0$ if and only if $x=\lambda y$ for some $\lambda\geqslant 0$.

Proof. Let $x,y,z\in\mathcal{C}\backslash 0$ such that $x\sim_c y\sim_c z$.

1. We take note that if $0<\alpha<m(\frac{x}{y})$ and $0<M(\frac{x}{y})<\beta$, we have $\alpha y\leqslant_c x\leqslant_c \beta y$ and $y\leqslant_c \frac{\beta}{\alpha}y$ since

$$\begin{aligned}\frac{\beta}{\alpha}y-y\in\mathcal{C} &\iff \beta y-\alpha y\in\mathcal{C}\\ &\iff (\underset{\in\mathcal{C}}{\beta y-x})+(\underset{(\in\mathcal{C}}{x-\alpha y)}\in\mathcal{C})\in\mathcal{C}.\end{aligned}$$

Hence, $(\frac{\beta}{\alpha}-1)y\in\mathcal{C}$ and consequently $\frac{\beta}{\alpha}-1\leqslant 0$ since $\bar{\mathcal{C}}\cap(-\bar{\mathcal{C}})=\{0\}$. To conclude, if we note $(\alpha_n)_{n\geqslant 0}$ and $(\beta_n)_{n\geqslant 0}$ two real sequences such that $0<\alpha_n<m(\frac{x}{y})$, $0<M(\frac{x}{y})<\beta_n$,

$$\lim_{n\to\infty}\alpha_n=m(\frac{x}{y})\quad\text{and}\quad\lim_{n\to\infty}\beta_n=M(\frac{x}{y}),$$

then we obtain through the limit

$$\frac{M(\frac{x}{y})}{m(\frac{x}{y})}=\lim_{n\to\infty}\frac{\alpha_n}{\beta_n}\geqslant 1.$$

2. To prove the second point, we simply need to use $m(\frac{x}{y})=M(\frac{y}{x})^{-1}$:

$$\begin{aligned}d(x,y) &= \log\big(\frac{M(\frac{x}{y})}{m(\frac{x}{y})}\big)=\log\big(M(\frac{x}{y})M(\frac{y}{x})\big)\\ &= \log\big(\frac{M(\frac{y}{x})}{m(\frac{y}{x})}\big)=d(y,x)\ .\end{aligned}$$

3. Let α,β as before, $0<\lambda<m\big(\frac{y}{z}\big)$ and $0<M\big(\frac{y}{z}\big)<\mu$. Then, we have $\alpha y\leqslant_c x$ and $\lambda z\leqslant_c y$, which when combined gives us $\alpha\lambda z\leqslant x$, thus

$$0<\alpha\lambda<m(\frac{x}{y}).$$

Similarly by combining $x \leqslant_c \beta y$ and $y \leqslant_c \mu z$, we obtain $x \leqslant_c \beta\mu z$ and so

$$M\big(\frac{x}{z}\big) \leqslant \beta\mu$$

and by pushing α,β,λ and μ to their respective *sup* or *inf* limits, we obtain

$$m\big(\frac{y}{z}\big)m\big(\frac{x}{y}\big) \leqslant m\big(\frac{x}{z}\big) \quad \text{and} \quad M\big(\frac{x}{z}\big) \leqslant M\big(\frac{x}{y}\big)M\big(\frac{y}{z}\big).$$

This allows us to directly conclude:

$$\begin{aligned} d(x,z) &= \log\left(\frac{M(\frac{x}{z})}{m(\frac{x}{z})}\right) \\ &\leqslant \log\left(\frac{M(\frac{x}{y})}{m(\frac{x}{y})}\frac{M(\frac{y}{z})}{m(\frac{y}{z})}\right) \\ &\leqslant d(x,y) + d(y,z)\ . \end{aligned}$$

4. Let $\lambda, \mu > 0$. Then,

$$M\big(\frac{\lambda x}{\mu y}\big) = \frac{\lambda}{\mu}M\big(\frac{x}{y}\big) \quad \text{and} \quad m\big(\frac{\lambda x}{\mu y}\big) = \frac{\lambda}{\mu}m\big(\frac{x}{y}\big),$$

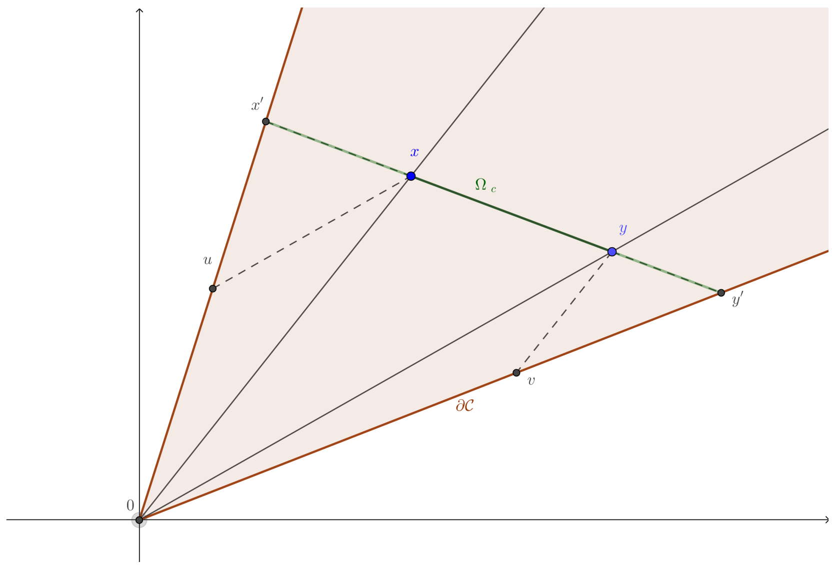
and so

$$d(x,y) = d(\lambda x, \mu y).$$

$\square$

Theorem 13.3.7 *Let $\mathcal{C} \subset V$ a closed cone in a $(n+1)-$dimensional vector space with a non-empty interior (ie $\overset{o}{\mathcal{C}} \neq \{\emptyset\}$) and $H \subset V$ be a n-dimensional affine hyperplane such that $\Omega_c = H \cap \overset{o}{\mathcal{C}}$ is a open, bounded and convex set. Then, the restriction of the Birkhoff metric d on Ω_c coincides with the Hilbert metric δ.*

Proof. Let $x, y \in \Omega_c$ distinct, $\alpha = m\big(\frac{x}{y}\big) = M\big(\frac{y}{x}\big)^{-1}$ and $\beta = M\big(\frac{x}{y}\big)$. We remark that because $\mathcal{C}$ is closed, we have αy and $x \leqslant \beta y$. We set $u = x - \alpha y \in \partial\mathcal{C}$, $v = y - \frac{1}{\beta}x \in \partial\mathcal{C}$, $\ell_{x,y}$ the Euclidean line passing through x and y and $x', y' \in \partial\Omega_c$ the two points of intersection between $\ell_{x,y}$ and $\partial\Omega_c$ such that $|y - x'| > |x - x'|$ and $|x - y'| > |y - y'|$.



Since x' and y' do not lie between x and y, we have $\lambda, \mu > 1$ such that

$$x' = y + \lambda(x - y) \quad \text{and} \quad y' = x + \mu(y - x).$$

Now let $\phi \in V^*$ be a linear functional such that[5]

$$H = \{z \in V : \phi(z) = 1\}$$

and we remark that

$$y + \lambda(x - y) = x' = \frac{u}{\phi(u)} = \frac{x - \alpha y}{1 - \alpha} \quad \Longrightarrow \quad \alpha = \frac{\lambda - 1}{\lambda}$$

$$x + \mu(y - x) = y' = \frac{v}{\phi(v)} = \frac{y - \beta^{-1}x}{1 - \beta^{-1}} \quad \Longrightarrow \quad \beta = \frac{\mu}{\mu - 1}$$

which leads us to

$$\frac{|y - x'|}{|x - x'|} = \frac{\lambda}{1 - \lambda} = \frac{1}{\alpha} = M\left(\frac{y}{x}\right) = m\left(\frac{x}{y}^{-1}\right)$$

and

$$\frac{|x - y'|}{|y - y'|} = \frac{\mu}{1 - \mu} = \beta = M\left(\frac{x}{y}\right).$$

Finally we may conclude:

$$d(x, y) = \log\left(\frac{M(\frac{x}{y})}{m(\frac{x}{y})}\right) = \log\left(\frac{|y - x'|}{|x - x'|}\frac{|x - y'|}{|y - y'|}\right) = \delta(x, y)$$

$\square$

[5] If a is a normal vector of H and $t \in \mathbb{R}$ such that $ta \in H$, then we have

$$H = \{x \in V : \langle x, a\rangle = t\} = \{x \in V : \underbrace{\left\langle \frac{a}{t}\right\rangle, x}_{=\phi(x)} = 1\}$$

Corollary 13.3.4 *The Hilbert metric δ is well defined on any open bounded convex set.*

Proof. If $\Omega \subset \mathbb{R}^n$ is a convex, bounded and open set, then we embed it in $\mathbb{R}^{n+1}$ as $\Omega' = \{\begin{pmatrix} x \\ 1 \end{pmatrix} : x \in \Omega\}$, set $\mathcal{C} = \{\lambda x : x \in \bar{\Omega}', \lambda \geqslant 0\}$ and $H = \{\begin{pmatrix} x \\ 1 \end{pmatrix} : x \in \mathbb{R}^{n+1}\}$. $\square$

This corollary is the weak version to a much stronger result: we can extend the Hilbert metric to any open, convex, possibly unbounded, subset of an infinite-dimensional Banach space.

Part III:

Applications

The only simple notions whose specialisations form a multiply extended manifoldness are the positions of perceived objects and colors.

B. Riemann, 1854

Chapter 14

The kinematics of special relativity

In this chapter we will introduce the standard formulation of special relativity kinematics, linking it to the concepts and nomenclature of hyperbolic geometry that we have developed in the previous chapters. The main reference is the classical and still unparalleled masterpiece of Landau and Lifshitz [8].

The idea of an absolute, immutable, space where the events of nature take place has been discarded since the era of Galileo, who introduced the first (spatial) relativistic theory in physics. Formally, Galilean relativity is based on the following two postulates, the first of which is very often considered as implicit:

Postulates of Galilean relativity

1. there is no privileged instant of time or position in space, i.e. time is homogeneous and space is homogeneous and isotropic, so, in particular, there is no privileged direction in space. Moreover, the geometry of space is Euclidean;

2. the laws of mechanics have the same form in all inertial (not accelerated) reference frames, i.e. no inertial reference frame is privileged: when the laws of mechanics are written in terms of the coordinates of any inertial frame, their form is exactly the same.

The so-called **standard configuration of relative motion** is described as follows: if (X, Y, Z) and (X', Y', Z') are the spatial coordinate axes of two reference frames[1] $\mathscr{R}$ and $\mathscr{R}'$, respectively, then we set $X' = X$ and Y' parallel to Y, Z' parallel to Z at any time $t' = t$. So, in the standard configuration, the relative motion takes place along the coincident $X' = X$ axes. The velocity of this relative motion is constant and denoted with v.

It has to be noticed that time is considered as absolute: in Galilean relativity any reference frame provides the same description of time passing.

If the time origin is chosen to be the moment in which the reference frames $\mathscr{R}$ and $\mathscr{R}'$ coincide, then, from the laws of mechanics, it follows immediately that the equations connecting

[1] In physics, a reference frame is identified by a set of coordinates that permit to indicate the spatial position and the instant of time.

the coordinate description of the two frames, known as **Galilean transformations**, are:

$$\begin{cases} t' = t \\ x' = x - vt \\ y' = y \\ z' = z \end{cases} \iff \begin{cases} t = t' \\ x = x' + vt \\ y = y' \\ z = z' \end{cases} .$$

Galilean relativity was incorporated in Newtonian mechanics, in which interaction among particles is described by means of potential energy, which appears in Newton equations as a function of the coordinates of the interacting particles. This implicitly contains a concept that Newton himself always considered paradoxical, but that he never managed to overcome, the *instantaneous propagation of interactions*. In the third letter to Richard Bentley (1692) Newton wrote: '*It is inconceivable that inanimate brute matter should, without the mediation of something else, which is not material, operate upon, and affect other matter without mutual contact*, [...] *That gravity should be innate, inherent, and essential to matter, so that one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man who has in philosophical matters a competent faculty of thinking, can ever fall into it. Gravity must be caused by an agent acting constantly according to certain laws; but whether this agent be material or immaterial, I have left to the consideration of my readers*'.

In fact, according to Newtonian mechanics as formulated in the Principia (Philosophiæ Naturalis Principia Mathematica, 1687), the force exerted by a particle on another at a particular instant of time depends only on its position at that precise instant, so a change in its position affects the other immediately.

Based on the discovery of the Jupiter moons by Galileo, astronomer Ole Rømer revealed that the speed of light is not instantaneous and many other experiments that followed showed that instantaneous interaction does not exist in nature: any change in a particle starts to have an influence on another particle only after a certain lapse of time, which establishes the existence of a *maximal velocity of propagation of the interaction*, obtained simply through the division of the spatial separation of the two particles by this lapse of time.

This existence of such a maximum velocity implies immediately that the motion of particles with a speed larger than that is impossible, for if such a motion occurred, it would permit to realize an interaction with a velocity exceeding the maximal admissible. In physics, the interaction propagating from a particle to another is called the **signal** that the first one sent out to 'inform' the other about the changes that it has experienced. For this reason, the velocity of propagation of interaction is more commonly called *signal velocity*.

In retrospective, the Michelson-Morley's empirical observation **that the speed of light in vacuum is constant in every inertial reference frame is a natural consequence of** the previous observation and **the Galilean postulates of relativity**: why should a reference frame be so privileged to be granted a faster maximum velocity or be so unfortunate to have a slower one? However, in a physical society dominated by the idea of 'luminiferous aether', it took a struck of genius an independent thinker, Albert Einstein, to come up with this 'brilliantly simple' conclusion and accept the constancy of light speed as a law of nature.

There is an interesting analogy between the roles played by Poincaré and Galileo, on one side, and Newton and Einstein, on the other side. Galileo was the first one to grasp the concept of spatial relativity, but Newton, through differential calculus, put it into formulae and developed his eponymous mechanics. Analogously, Poincaré already in 1889, see [17], claimed that 'the existence of the luminiferous aether was a question for metaphysicians to deal with' and embraced the constancy of the speed of light, however, he failed to produce Einstein's splendid achievements because he did not think to gain any new knowledge combining the constancy of the speed of light with Newtonian time and Euclidean space (in this sense, anticipating the philosophy behind Einstein's theory of general relativity!), see e.g. [23] for more information.

As Lorentz and Poincaré, Einstein believed that not only the laws of mechanics had to be the same in all inertial reference frame, but also those of electromagnetism. For this reason, he reinforced the postulates of Galilean relativity to incorporate Maxwell's equations and he elevated the constancy of the speed of light as a fundamental postulate of physics.

This resulted in the following:

Postulates of Einstein's theory of special relativity

1. same as in Galilean relativity;
2. the laws of nature (mechanics and electromagnetism) have the same form in all inertial reference frames;
3. the speed of light in vacuum has the same constant value c when measured in all inertial reference frames.

The large value of c compared to the typical speed v of particles explains why Newtonian mechanics works so well in most situations: if $v \ll c$, then c can be considered as infinite without generating measurable errors.

As we will show, considering these new set of postulates imposes, among other surprising consequences, to give up the concept of an absolute time and to build a 4-dimensional 'spacetime' where both space and time are relative to an observer and whose geometry is not Euclidean anymore, but Lorentzian.

To avoid misleading interpretations, from now on the term relativistic will always be referred to Einstein's postulates an not the Galilean ones.

14.1 Events, world-lines and spacetime intervals

Using a standard nomenclature, we call **event** e a point in $\mathbb{R}^4$ written in coordinates as[2] $e = (x^0 \equiv ct, x^1, x^2, x^3)^t = (ct, \mathbf{x})^t = (x^\mu)^t$, $\mu = 0, 1, 2, 3$, where t and $\mathbf{x} = (x^i)$, $i = 1, 2, 3$, are, respectively, the time instant and the spatial position of a particle or a ray of light as measured by an inertial observer with respect to a given inertial reference frame $\mathscr{R}$. A **world-line** is the trajectory in $\mathbb{R}^4$ described by the evolution in time of the spatial coordinates of a particle or a ray of light. A uniform rectilinear motion describes a straight world-line.

Let us now consider the following two events: the first, $e_1 = (ct_1, x_1^i)$, consists in a light signal emanating at the time t_1 from the spatial position (x_1^i); the second, $e_2 = (ct_2, x_2^i)$, consists in the same light signal arriving at the time t_2 in the spatial position (x_2^i). Since the

[2] Using ct instead of t is customary in special relativity: physically, this amounts at replacing the time t with the corresponding space ct traveled by a ray of light during t, in this way, the coordinates (ct, x, y, z) have the same dimensions.

signal propagates with constant speed c, the distance that is traveled is $c(t_2 - t_1)$, however, by the first postulate, the description of space is Euclidean, so this same distance in $\mathbb{R}^3$ verifies the following equality:

$$c(t_2 - t_1) = \left(\sum_{i=1}^{3} (x_2^i - x_1^i)^2 \right)^{1/2},$$

In special relativity, the quantity

$$s_{1,2} := \left(c^2(t_2 - t_1)^2 - \sum_{i=1}^{3} (x_2^i - x_1^i)^2 \right)^{1/2}$$

is called **spacetime interval** between the two events, which can be recognized to be the **Minkowski pseudo-norm** of $e_2 - e_1$. It is more common to work with the squared interval to avoid writing the square root.

For the events e_1 and e_2 considered above we have:

$$c^2(t_2 - t_1)^2 - \|\mathbf{x}_2 - \mathbf{x}_1\|^2 = 0, \tag{14.1}$$

$\|x_2 - x_1\|$ being the Euclidean distance in $\mathbb{R}^3$ between $\mathbf{x}_1$ and $\mathbf{x}_2$. From eq. (14.1), it follows that the Minkowski distance is null for any two events connected by a signal traveling at the speed of light, no matter how fare in space or time the two events are. Otherwise stated, *from the perspective of 4-dimensional distances*, measured with the spacetime interval (alias the Minkowski norm), *a ray of light stays perfectly still.*

Now, since we have not imposed any constraint on the events, eq. (14.1) must remain valid also when their difference is infinitesimal, thus we can write the differential version of the squared interval as

$$\boxed{ds^2 = c^2 dt^2 - \|d\mathbf{x}\|^2}. \tag{14.2}$$

The infinitesimal square spacetime interval can be written as the indefinite quadratic form

$$ds^2 = (dx^\mu)^t \eta_{\mu\nu} (dx^\nu),$$

where $\eta_{\mu\nu}$ is the matrix $\operatorname{diag}(1, -1, -1, -1)$. $\mathcal{M} = (\mathbb{R}^4, ds^2)$ is the Minkowski spacetime and, as we have already recalled, $\|e\|_{\mathcal{M}}^2 = (x^0)^2 - [(x^1)^2 + (x^2)^2 + (x^3)^2]$ coincides with the square of the Minkowski pseudo-norm of $e \in \mathcal{M}$.

Noticeable subsets of $\mathcal{M}$ are defined by the values taken by ds^2:

- $ds^2 = 0$, the events e_1, e_2 are connected by a signal traveling at the light speed, they belong to the lightcone $\mathcal{L} = \{e = (ct, x, y, z)^t \in \mathbb{R}^4 \ : \ \|e\|_{\mathcal{M}}^2 = 0 \iff c^2t^2 - x^2 - y^2 - z^2 = 0\}$;

- $ds^2 > 0$, or $\|d\mathbf{x}\|^2 < c^2 dt^2$, i.e. the spatial separation between the events e_1, e_2 is less that the distance traveled by a light ray, which implies that they are connected by a world-line with speed less than c, they lie in the interior of the lightcone, the *time-like* zone of the Minkowski space. It is also called *causality region*, because changes in the event e_1 cause changes in the event e_2;

- $ds^2 < 0$, or $\|d\mathbf{x}\|^2 > c^2 dt^2$, i.e. the spatial separation between the events e_1, e_2 is larger that the distance traveled by a light ray, i.e. the events e_1, e_2 cannot be physically connected, they lie outside the lightcone, the *space-like* zone of the Minkowski space, also called *non-causal region.*

The wording time-like and space-like is now explained, the choice between time or space depends on the part of the spacetime interval that prevails over the other.

Now we come to a crucial result: postulates 1 and 3 of special relativity imply the following key invariance property of ds^2.

Theorem 14.1.1 (Invariance of spacetime interval in any inertial reference frame) *The spacetime interval ds^2 between two events described in the inertial reference frame $\mathscr{R}$ and the spacetime interval ds'^2 between the same couple of events described in any other inertial reference frame $\mathscr{R}'$ is exactly the same:*

$$ds'^2 = ds^2.$$

As a consequence, also any finite spacetime interval between two events is invariant under changes of the inertial reference frame.

Proof. We follow the very elegant proof in [8], see also [9], page 7, for an alternative demonstration.

By the postulate of constancy of the speed of light, it follows immediately that if $ds^2 = 0$ in $\mathscr{R}$, then also $ds'^2 = 0$. Moreover, ds^2 and ds'^2 are infinitesimals of the same order (quadratic). These two observations imply that ds^2 and ds'^2 must be proportional to each other[3], i.e. $ds'^2 = \alpha ds^2$, $\alpha \in \mathbb{R}$. The intuitive argument goes as follows: any non-linear function applied to ds^2 would modify the infinitesimal order, so only affine functions can relate ds'^2 to ds^2. However, if we had $ds'^2 = \alpha ds^2 + \beta$, with $\beta \neq 0$, then $ds^2 = 0$ would not imply $ds'^2 = 0$, so we are left with $ds'^2 = \alpha ds^2$.

The coefficient α cannot depend on time or space coordinates, otherwise different points in space and different moments in time would not be equivalent, which is against the first postulate. Since the only difference between $\mathscr{R}$ and $\mathscr{R}'$ is their relative speed $\mathbf{v}$, we have that α can only be a function of $\mathbf{v}$, however, due to the postulate of isotropy of space, α cannot depend on the direction of $\mathbf{v}$, but only on its magnitude v, hence we are left with the function $\alpha(v)$.

To prove that $\alpha(v)$ is a constant function equal to 1, let us consider three inertial reference frames $\mathscr{R}$, $\mathscr{R}_1$ and $\mathscr{R}_2$, with v_1 and v_2 indicating the magnitude of the relative velocity of $\mathscr{R}_1$ and $\mathscr{R}_2$ w.r.t. $\mathscr{R}$, respectively, and $v_{1,2}$ the relative velocity between $\mathscr{R}_1$ and $\mathscr{R}_2$. Thanks to the arguments discussed above, we have:

$$ds_1^2 = \alpha(v_1) ds^2, \quad ds_2^2 = \alpha(v_2) ds^2, \quad ds_2^2 = \alpha(v_{1,2}) ds_1^2, \tag{14.3}$$

the equality of the second and third equations implies

$$\alpha(v_2) ds^2 = \alpha(v_{1,2}) ds_1^2 \iff ds_1^2 = \frac{\alpha(v_2)}{\alpha(v_{1,2})} ds^2,$$

[3] The rigorous statement about the proportionality of ds^2 and ds'^2 is the following: let $n, p \geqslant 1$ be integers, $d := n + p$ and V a d-dimensional real vector space. Let h be a pseudo-scalar product on V with signature of type (n, p) and g be a symmetric bilinear form on V such that the null set of the associated quadratic form of h is contained in that of g, i.e for every $v \in V$, if $h(v, v) = 0$ then $g(v, v) = 0$. Then, there exists a constant $\alpha \in \mathbb{R}$ such that $g = \alpha h$. Furthermore, if $n \neq p$ and the signature of g is (n, p) too, then $\alpha > 0$. See for example the url https://en.wikipedia.org/wiki/Derivations_of_the_Lorentz_transformations#Using_group_theory for the proof.

so, comparing with the first equation in formula (14.3) we find

$$\alpha(v_{1,2}) = \frac{\alpha(v_2)}{\alpha(v_1)}.$$

But $v_{1,2}$ depends (and so does $\alpha(v_{1,2})$) on the relative position, i.e. the angle, between $\mathbf{v}_1$ and $\mathbf{v}_2$, which is not the case for $\alpha(v_1)$ and $\alpha(v_2)$, hence the previous formula can be valid only if $\alpha(v)$ is a constant function: $\alpha(v) = k$ for all v, which implies $\alpha(v_{1,2}) = \alpha(v_1) = \alpha(v_2) = k$, so $k = k/k = 1$. $\square$

14.1.1 The issue of simultaneity of events in the 4-dimensional Minkowski spacetime

One of the most important consequences of the postulates of special relativity is the *impossibility to find two inertial reference frames in which observers can describe events occurring simultaneously both in space and time.*

In order to understand why, let us consider two events described by $e_1 = (ct_1, x_1, y_1, z_1)^t$ and $e_2 = (ct_2, x_2, y_2, z_2)^t$ in the inertial frame $\mathscr{R}$ and search for an inertial frame $\mathscr{R}'$ in which these events occur at the same point in space. If $t_{1,2} := t_2 - t_1$ and $t'_{1,2} := t'_2 - t'_1$, then the squared spacetime intervals between e_1 and e_2 in $\mathscr{R}$ and $\mathscr{R}'$ are, respectively,

$$s^2_{1,2} = c^2 t^2_{1,2} - \|\mathbf{x}_2 - \mathbf{x}_1\|^2, \quad s'^2_{1,2} = c^2 t'^2_{1,2} - \|\mathbf{x}'_2 - \mathbf{x}'_1\|^2,$$

but the invariance of spacetime intervals in all inertial frames implies

$$c^2 t^2_{1,2} - \|\mathbf{x}_2 - \mathbf{x}_1\|^2 = c^2 t'^2_{1,2} - \|\mathbf{x}'_2 - \mathbf{x}'_1\|^2. \tag{14.4}$$

Now, if in $\mathscr{R}'$ the events e_1 and e_2 occur in the same place, then $\|\mathbf{x}'_2 - \mathbf{x}'_1\| = 0$, so eq. (14.4) becomes

$$c^2 t^2_{1,2} - \|\mathbf{x}_2 - \mathbf{x}_1\|^2 = c^2 t'^2_{1,2} > 0,$$

this inequality imposes a precise constraint on the possibility of observers belonging to the two frames to **describe two events as occurring at the same point in space**: this is **possible only if the events belong to the time-like region** of the Minkowski spacetime.

If we repeat the same analysis conducted above, but searching for a reference frame $\mathscr{R}'$ in which the **events** e_1 and e_2 can be described as **occurring at the same time**, i.e. $t'_{1,2} = 0$, then eq. (14.4) becomes

$$c^2 t^2_{1,2} - \|\mathbf{x}_2 - \mathbf{x}_1\|^2 = -\|\mathbf{x}'_2 - \mathbf{x}'_1\|^2 < 0,$$

i.e. e_1 and e_2 **must belong to the space-like region** of the Minkowski spacetime.

Due to the fact that the time-like or space-like character of an interval is independent of the inertial frame, the results that we have just found have an absolute nature and establish that it is impossible to find two inertial frames that allow observers to describe two events as occurring at the same location and in the same instant of time. Otherwise stated, **spacetime simultaneity is impossible**.

14.2 Proper time and proper length, time dilation and space contraction

Suppose to observe from an inertial frame $\mathscr{R}$ a clock moving uniformly with velocity v and suppose that the inertial frame $\mathscr{R}'$ is moving with the clock, or *co-moving*. Since the clock is at rest in $\mathscr{R}'$, the space displacement read in $\mathscr{R}'$ will be 0 for all instant of time t'. However, from the perspective of $\mathscr{R}$, in an infinitesimal lapse of time dt the clock moved covering a distance $\|d\mathbf{x}\| = \sqrt{(dx)^2 + (dy)^2 + (dz)^2}$. We can ask ourselves what is the corresponding time interval indicated by the moving clock in $\mathscr{R}'$. Because of the invariance of spacetime intervals we have the equality:

$$ds^2 = c^2(dt)^2 - \|d\mathbf{x}\|^2 = c^2(dt')^2 \iff (dt')^2 = (dt)^2 - (dt)^2 \frac{\|d\mathbf{x}\|^2}{c^2(dt)^2}$$

so

$$dt' = dt\sqrt{1 - \frac{\|d\mathbf{x}\|^2}{c^2(dt)^2}} = dt\sqrt{1 - \frac{v^2}{c^2}}.$$

Defining the **Lorentz factor** as

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \in [1, +\infty), \quad \gamma \underset{v \to 0}{\to} 1, \quad \gamma \underset{v \to c_-}{\to} +\infty,$$

we obtain $dt' = \frac{dt}{\gamma}$, integrating dt' between two instants of time t_1 and t_2 measured in $\mathscr{R}$, we get that the time elapsed in the inertial frame $\mathscr{R}'$ moving with the clock is

$$\Delta t' = t_2' - t_1' = \int_{t_1}^{t_2} dt' = \int_{t_1}^{t_2} \frac{dt}{\gamma} = \frac{t_2 - t_1}{\gamma} = \frac{\Delta t}{\gamma}.$$

So, to resume:

$$\Delta t' = \frac{\Delta t}{\gamma}. \tag{14.5}$$

The time read by a clock co-moving with a particle is called **proper time**. Eq. (14.5) allows us to relate the proper time with the time read in another inertial frame $\mathscr{R}$ moving with an absolute velocity v w.r.t. the clock. We can see that the two intervals of time approach to each other when $v \to 0$, but, if v is a non-negligible fraction of the speed of light, the proper time is considerably smaller than the time measured in $\mathscr{R}$.

This fact is often popularized by saying that '*the time measured by objects in movement flows slower that for those which are at rest*'. However, this sentence must be taken carefully: since the measure of time is relative to the inertial frame where a clock is, from the perspective of the moving object the time flows in exactly the same way, it is when its measure of time is compared to that of a clock moving with a different velocity that the we find a difference.

Notice that, if $v \to c_-$, then $\gamma \to +\infty$ and $\Delta t' \to 0$, which shows that not only the spacetime interval for a ray of light is null, but also its proper time! From eq. (14.5) we get

$$\boxed{\Delta t = \gamma \Delta t'}, \tag{14.6}$$

which is known as the **time dilation** formula, due to the fact that $\gamma > 1$ for all $v > 0$.

Let us now pass to the contraction of length, also called *Lorentz contraction* because it was first proven by Lorentz before the development of Einstein's theory of special relativity.

If treated superficially, the measure of length in special relativity can lead to paradoxes, misleading interpretations or even incorrect formulae. The reason is that there are several equivalent ways to derive the formula for length contraction and the majority of them, in order to be correctly applied, need the concept of *clock synchronization*, first discussed in philosophical terms by Poincaré, then written in a physical language by Einstein and further formalized mathematically by other scientists in the twentieth century, as Max von Laue, Ludwik Silberstein, Hans Reichenbach and Alan Macdonald.

There is however a clever derivation which is not founded on clock synchronization but on the already proven time dilation and on the fact that the *magnitude* of the relative velocity between two inertial frames is the same for both of them. To avoid discussing the delicate and time-consuming issue of clock synchronization, we shall derive Lorentz contraction of length using this method.

Just like the time measured by an observer at rest in an inertial frame is called proper time, we define the **proper length** to be the length of an object, e.g. a rod, at rest in an inertial frame, e.g. by superposing a ruler on the rod without taking care of time passing, since it is at rest.

Let us consider again the standard configuration of relative motion along the x-coordinates at velocity v of two inertial frames $\mathscr{R}$ and $\mathscr{R}'$. Imagine that a rod is placed along the X axis at rest in $\mathscr{R}$ and its proper length is measured to be L_0. Suppose, on the contrary, that a clock is at rest in $\mathscr{R}'$, then we can use it to measure the proper time of the frame $\mathscr{R}'$. Let T_0' be the proper time measured by this clock between the passage of the extreme points of the rod in a fixed spatial position in $\mathscr{R}'$. Since for an observer in $\mathscr{R}'$ the rod is moving with (absolute) velocity v, the length L' of the rod in $\mathscr{R}'$ can be *defined* to be

$$L' := T_0' v. \tag{14.7}$$

Now, the time dilation formula (14.6) allows us to relate the measure of time in $\mathscr{R}$ with that of $\mathscr{R}'$ simply by multiplying it by the Lorentz factor γ, hence

$$T = \gamma T_0'$$

is the time that has passed in $\mathscr{R}$ during the measure of the rod length in $\mathscr{R}'$. Thanks to the fact that we know the proper length of the rod L_0 in $\mathscr{R}$, they must be related as follows:

$$L_0 = Tv.$$

Solving this last equation and eq. (14.7) w.r.t. v and equating we find:

$$\frac{L'}{T_0'} = \frac{L_0}{T},$$

so

$$L' = \frac{L_0 T_0'}{T} = \frac{L_0 T_0'}{\gamma T_0'}$$

but $T = \gamma T_0'$, so we obtain

$$\boxed{L' = \frac{L_0}{\gamma}}, \tag{14.8}$$

which is known as the **length contraction** formula, because $\gamma > 1$ for all $v > 0$, so the length measured in the frame $\mathscr{R}'$ which moves at a constant velocity v w.r.t. to the rod is shorter than the one measured in the frame $\mathscr{R}$, where the rod is at rest.

14.3 Lorentz transformations in special relativity

We now discuss the problem of relating the coordinate description of events in two inertial frames $\mathscr{R}$, $\mathscr{R}'$. First of all, it is simple to deduce **from postulate 1** that **the coordinate transformation from $\mathscr{R}$ to $\mathscr{R}'$, $T : \mathbb{R}^4 \to \mathbb{R}^4$, $x^\mu \mapsto x'^\mu = T(x^\mu)$ must be affine** (under the reasonable hypothesis of being differentiable).

In fact, by postulate 1, there are no special instants and positions in $\mathbb{R}^4$, so the Euclidean distance between two events remains the same when these are rigidly shifted by a fixed vector $b \in \mathbb{R}^4$.

This is true independently of the coordinate system used to write the events in two arbitrary inertial reference frames $\mathscr{R}$ and $\mathscr{R}'$. Let $x = x^\mu$ and $y = y^\mu$ be the coordinates of the two events in $\mathscr{R}$ and $T^\mu(x)$ and $T^\mu(y)$ the coordinates of the same events in $\mathscr{R}'$. Since $(x^\mu + b^\mu) - (y^\mu + b^\mu) = x^\mu - y^\mu$, we must have $T^\mu(x + b) - T^\mu(y + b) = T^\mu(x) - T^\mu(y)$. If we derive the two sides of the last equation with respect to x^ν, $\nu = 1, 2, 3, 4$, we obtain $\frac{\partial T^\mu}{\partial x^\nu}(x + b) = \frac{\partial T^\mu}{\partial x^\nu}(x)$, for all $b \in \mathbb{R}^4$, since y does not depend on x.

Thanks to the fact that b is arbitrary, $x + b$ represents any vector in $\mathbb{R}^4$, so the function $\frac{\partial T^\mu}{\partial x^\nu}$ is constant, i.e. $\frac{\partial T^\mu}{\partial x^\nu}(x) = \Lambda^\mu_{\ \nu} \in \mathbb{R}$ for all $x \in \mathbb{R}^4$, $\mu, \nu = 1, 2, 3, 4$, so, integrating w.r.t. x we obtain the affine form of the coordinate transformation:

$$x'^\mu = T^\mu(x) = \Lambda^\mu_{\ \nu} x^\nu + a^\mu. \tag{14.9}$$

Let us now impose the preservation of the spacetime interval, i.e. of the Minkowski norm. The term a^μ does not modify the interval because it produces only a shift of the coordinate origin. So, the requirement of preservation of the Minkowski norm falls only on the matrix Λ, the linear part of T. Thanks to the results of chapter 11, this implies that Λ must belong to the Lorentz group

$$\mathscr{L} = O(1, 3) = \{\Lambda \in GL(4, \mathbb{R}) \ : \ \Lambda^t \eta \Lambda = \eta = \mathrm{diag}(1, -1, -1, -1)\}.$$

So, postulate 1 and the constancy of the speed of light in inertial reference frames imply that the coordinates used to describe the same event in two inertial frames are related by a non-homogeneous linear transformation of the type $x' = \Lambda x + a$, $\Lambda \in O(1, 3)$, $a \in \mathbb{R}^4$. The set of these transformations forms the so-called the *Poincaré group* defined by

$$\mathscr{P} = \{(\Lambda, a) \ : \ \Lambda \in O(1, 3), \ a \in \mathbb{R}^4\}$$

under the composition law given by $(\Lambda_1, a_1) \cdot (\Lambda_2, a_2) = (\Lambda_1\Lambda_2, a_1 + \Lambda_1 a_2)$.

The Poincaré group contains the Lorentz group and the group of translations as subgroups: $O(1, 3)$ is isomorphic to the subgroup of $\mathscr{P}$ given by the elements $(\Lambda, \mathbf{0})$, while the elements of $\mathscr{P}$ of the form (I_4, a) are translations by the constant vector $a \in \mathbb{R}^4$.

The affine coordinate transformations $x' = \Lambda x + a$ are called *Poincaré transformations*, and those corresponding to $a = 0$, i.e. the linear functions $x' = \Lambda x$, are of course the *Lorentz transformations*.

The most interesting part of these maps is clearly provided by the Lorentz transformation encoded in the matrix Λ. In the following section we will analyze the easiest non-trivial type of Lorentz transformation, called boost.

14.4 Lorentz boosts

Lorentz boosts are Lorentz transformations that take place only in a given spatial direction, thus leaving the coordinates relative to the orthogonal directions unchanged. In the standard configuration recalled before, this direction is taken to be the X axis. Hence, the coordinate modification of a Lorentz boost between two frames that move at relative velocity v in the x-direction will be defined by the action of the linear function $B(v) : \mathcal{M} \to \mathcal{M}$,

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \underset{B(v)}{\longmapsto} \begin{pmatrix} ct' \\ x' \\ y \\ z \end{pmatrix},$$

since, as we already know, also the time coordinate will be affected by this transformation.

There are several ways to determine the form of the matrix associated to $B(v)$, see e.g. [11] page 119 for a proof based on an extension of Galilean transformation or [3] page 38 for an elegant proof that relies on linear algebra techniques. Here we give the details of the most straightforward method based simply on writing the linear system

$$\begin{cases} ct = Act' + Bx' \\ x = Cct' + Dx' \end{cases} \tag{14.10}$$

and then imposing the preservation of the spacetime interval, i.e.

$$(ct)^2 - x^2 = (ct')^2 - x'^2.$$

Replacing ct and x written as in the linear system above we obtain:

$$A^2(ct')^2 + B^2x'^2 + 2ABct'x' - C^2(ct')^2 - D^2x'^2 - 2CDct'x' = (ct')^2 - x'^2,$$

or

$$(A^2 - C^2)(ct')^2 + (B^2 - D^2)x'^2 + (2AB - 2CD)ct'x' = (ct')^2 - x'^2$$

which is verified by the solutions of this system:

$$\begin{cases} A^2 - C^2 = 1 \\ D^2 - B^2 = 1 \\ AB - CD = 0 \end{cases}.$$

We now need to recall an elementary result.

Proposition 14.4.1 *Given any point $(x, y) \in \mathbb{R}^2$, with $x > 0$, on the hyperbola of equation $x^2 - y^2 = 1$, there exists a unique $u \in \mathbb{R}$ such that $x = \cosh(u)$, $y = \sinh(u)$.*

Proof. Since $\sinh : \mathbb{R} \to \mathbb{R}$ bijective, it is enough to set $u := \sinh^{-1}(y)$, in this way $y = \sinh(u)$ and so $x^2 - y^2 = 1$ can be re-written as $x^2 - \sinh^2(u) = 1$. This, together with the identity $\cosh^2(u) - \sinh^2(u) = 1$, implies that that $x = \pm\cosh(u)$, but since we have supposed that $x > 0$, we remain with $x = \cosh(u)$. □

By proposition 14.4.1, we have that there exist two parameters $\zeta, \xi \in \mathbb{R}$ such that:

$$\begin{cases} A = \cosh(\zeta),\ C = \sinh(\zeta) \\ D = \cosh(\xi),\ B = \sinh(\xi) \\ \cosh(\zeta)\sinh(\xi) - \cosh(\xi)\sinh(\zeta) = 0 \end{cases},$$

the third equation can be re-written as $\sinh(\zeta - \xi) = 0$, which implies $\xi = \zeta$, so we are left with the transformation

$$\begin{cases} ct = \cosh(\zeta)ct' + \sinh(\zeta)x' \\ x = \sinh(\zeta)ct' + \cosh(\zeta)x' \end{cases}, \tag{14.11}$$

or, in matrix notation

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} \cosh(\zeta) & \sinh(\zeta) \\ \sinh(\zeta) & \cosh(\zeta) \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix}.$$

To obtain the boost directed as in eq. (14.10) we just have to invert the matrix appearing above:

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh(\zeta) & -\sinh(\zeta) \\ -\sinh(\zeta) & \cosh(\zeta) \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} \cosh(-\zeta) & \sinh(-\zeta) \\ \sinh(-\zeta) & \cosh(-\zeta) \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix},$$

or

$$\begin{cases} ct' = \cosh(\zeta)ct - \sinh(\zeta)x \\ x' = -\sinh(\zeta)ct + \cosh(\zeta)x \end{cases}.$$

In order to relate the parameter ζ to the relative velocity v that defines the boost (together with the direction chosen, of course), let us consider the motion, w.r.t. the unprimed coordinates, of the origin of the primed coordinates. This is accomplished by replacing $x' = 0$ in eqs. (14.11), which gives:

$$ct = \cosh(\zeta)ct',\ x = \sinh(\zeta)ct',$$

dividing the two expressions we find

$$\frac{1}{c}\frac{x}{t} = \frac{\sinh(\zeta)}{\cosh(\zeta)} \iff \frac{v}{c} = \frac{\sinh(\zeta)}{\cosh(\zeta)} \tag{14.12}$$

or

$$\frac{v}{c} = \tanh(\zeta) \iff \zeta = \tanh^{-1}(v/c),$$

the parameter ζ is called the **rapidity** of the boost. It is custom to denote with

$$\beta := \frac{v}{c} \in [0, 1), \quad \beta \underset{v \to 0}{\to} 0,\ \beta \underset{v \to c}{\to} 1,$$

the **fraction of speed of light** at which the inertial frames are moving w.r.t. each other, so

$$\beta = \tanh(\zeta).$$

Thanks to the hyperbolic identity

$$\cosh(\zeta) = \frac{1}{\sqrt{1 - \tanh^2(\zeta)}}, \quad \forall \zeta \in \mathbb{R},$$

we obtain $\cosh(\zeta) = 1/\sqrt{1-\beta^2} = \gamma$ and so, using (14.12), we find $\sinh(\zeta) = \beta\gamma$. We resume below the relationships that we have just found:

$$\begin{cases} \beta = \frac{v}{c} = \tanh(\zeta) \\ \gamma = \frac{1}{\sqrt{1-\beta^2}} = \cosh(\zeta) \\ \beta\gamma = \sinh(\zeta) \end{cases} . \tag{14.13}$$

Re-introducing the orthogonal spatial coordinates which do not experience the transformation we get:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh(\zeta) & -\sinh(\zeta) & 0 & 0 \\ -\sinh(\zeta) & \cosh(\zeta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv B(\zeta) \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \qquad (x-\text{boost}),$$

or

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv B(\beta) \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \qquad (x-\text{boost}).$$

Written as a system of transformations, the x-boost gives:

$$\begin{cases} ct' = \gamma(ct - \beta x) = \gamma\left(ct - \frac{v}{c}x\right) \\ x' = \gamma(x - \beta ct) = \gamma(x - vt) \\ y' = y, \ z' = z \end{cases} \quad \text{and} \quad \begin{cases} ct = \gamma(ct + \beta x') = \gamma\left(ct' + \frac{v}{c}x'\right) \\ x = \gamma(x' + \beta ct') = \gamma(x' + vt') \\ y = y', \ z = z' \end{cases} . \tag{14.14}$$

Notice that, if $v \ll c$, then $\gamma \simeq 1$ and $v/c \simeq 0$, which implies $t' = t$ and $x' \simeq x - vt$, i.e. **for relative velocities negligible w.r.t. to speed of light, the Lorentz boosts reduce to the Galilean transformations**.

For the sake of completeness, the boosts in the y and z-directions are written below:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh(\zeta) & 0 & -\sinh(\zeta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sinh(\zeta) & 0 & \cosh(\zeta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv B(\zeta) \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \qquad (y-\text{boost}),$$

or

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & -\beta\gamma & 0 \\ 0 & 1 & 0 & 0 \\ -\beta\gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv B(\beta) \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \qquad (y-\text{boost}),$$

and

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh(\zeta) & 0 & 0 & -\sinh(\zeta) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh(\zeta) & 0 & 0 & \cosh(\zeta) \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv B(\zeta) \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \qquad (z-\text{boost}),$$

or

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv B(\zeta) \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \qquad (z-\text{boost}). \tag{14.15}$$

Notice that the matrices representing the **boosts are real and symmetric**.

The similarities between these matrices and the rotation matrices explain why Lorentz boost are also called **hyperbolic rotations**, however, it should be clear that **Lorentz boost are not rotations in the Euclidean sense**. Instead, being Lorentz transformations, *boosts take a vector in $\mathcal{M}$ representing an event e and transform it into another vector belonging to the same hyperboloid defined by the preserved Minkowski norm of e*, see figure 14.3 for a graphical depiction.

Finally, a simple direct computation shows that **two boost matrices along the same axis commute, but this is never the case for boost along different axes**. An important consequence of this fact is that, if the boosted frame $\mathscr{R}'$ is further boosted to a third frame $\mathscr{R}''$ in a different direction, then the transformation from $\mathscr{R}$ to $\mathscr{R}''$ cannot be described by a simple boost anymore, but by the composition of a boost and a rotation around the axis orthogonal to the plane defined by the two boost directions. This rotation is called either '**Wigner** or **Thomas** rotation', the details can be seen at the Wikipedia url: https://en.wikipedia.org/wiki/Thomas_precession#Mathematical_explanation.

14.4.1 Boost in the general configuration

Let us consider an event e with coordinates $(ct, x, y, z)^t$ in the reference frame $\mathscr{R}$ and $(ct', x', y', z')^t$ in the reference frame $\mathscr{R}'$, which is moving with a relative velocity $\mathbf{v}$ w.r.t. $\mathscr{R}$. In the standard configuration, the only non-null component of $\mathbf{v}$ is along the x-coordinates, while here we allow $\mathbf{v}$ to be a, fixed, vector of $\mathbb{R}^3$ with arbitrary direction. Our aim is to express the Lorentz boost w.r.t. $\mathbf{v}$. For that it will be useful to write the *spatial coordinates* of e in the two reference frames as the vectors $\mathbf{r} = (x, y, z)^t$ and $\mathbf{r}' = (x', y', z')^t$.

The main idea of the procedure is to decompose $\mathbf{r}$ in the components parallel and orthogonal to $\mathbf{v}$, denoted $\mathbf{r}_\parallel$ and $\mathbf{r}_\perp$, respectively, so that $\mathbf{r} = \mathbf{r}_\parallel + \mathbf{r}_\perp$.

A Lorentz transformation acts as a boost on $\mathbf{r}_\parallel$ and does not affect $\mathbf{r}_\perp$, thus we can write

$$\mathbf{r}' = \gamma(\mathbf{r}_\parallel - \mathbf{v}t) + \mathbf{r}_\perp = \gamma(\mathbf{r}_\parallel - \mathbf{v}t) + \mathbf{r} - \mathbf{r}_\parallel = -\gamma\mathbf{v}t + \mathbf{r} + (\gamma - 1)\mathbf{r}_\parallel,$$

but

$$\mathbf{r}_\parallel = \frac{\mathbf{r} \cdot \mathbf{v}}{|\mathbf{v}|^2}\mathbf{v}$$

and

$$\vec{\beta} := (\beta_x, \beta_y, \beta_z)^t = \frac{\mathbf{v}}{c} = \left(\frac{v_x}{c}, \frac{v_y}{c}, \frac{v_z}{c}\right)^t \quad \text{and} \quad |\vec{\beta}|^2 = \frac{|\mathbf{v}|^2}{c^2},$$

so

$$\begin{cases} x' = -\gamma\beta_x ct + x + (\gamma - 1)\left[x\beta_x \frac{\beta_x}{|\vec{\beta}|^2} + y\beta_y \frac{\beta_x}{|\vec{\beta}|^2} + z\beta_z \frac{\beta_x}{|\vec{\beta}|^2}\right] \\ \\ y' = -\gamma\beta_y ct + y + (\gamma - 1)\left[x\beta_x \frac{\beta_y}{|\vec{\beta}|^2} + y\beta_y \frac{\beta_y}{|\vec{\beta}|^2} + z\beta_z \frac{\beta_y}{|\vec{\beta}|^2}\right] \\ \\ z' = -\gamma\beta_z ct + z + (\gamma - 1)\left[x\beta_x \frac{\beta_z}{|\vec{\beta}|^2} + y\beta_y \frac{\beta_z}{|\vec{\beta}|^2} + z\beta_z \frac{\beta_z}{|\vec{\beta}|^2}\right] \end{cases} .$$

It remains only the time coordinate transformation: in the standard configuration, ct transforms as $ct' = \gamma(ct - (v/c)x)$, in the present situation we must replace x by the component of $\mathbf{r}$ in the $\mathbf{v}$ direction, i.e. $\mathbf{r}\cdot\mathbf{v}$, so the transformation is

$$ct' = \gamma\left(ct - \frac{\mathbf{r}\cdot\mathbf{v}}{c}\right) = \gamma\left(ct - \frac{xv_x}{c} - \frac{yv_y}{c} - \frac{zv_z}{c}\right),$$

or, in terms of $\vec{\beta}$,

$$ct' = \gamma(ct - x\beta_x - y\beta_y - z\beta_z).$$

The matrix expression of the boost $B(\mathbf{v})$ is then:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta_x\gamma & -\beta_y\gamma & -\beta_z\gamma \\ -\beta_x\gamma & 1+(\gamma-1)\frac{\beta_x^2}{|\vec{\beta}|^2} & (\gamma-1)\frac{\beta_x\beta_y}{|\vec{\beta}|^2} & (\gamma-1)\frac{\beta_x\beta_z}{|\vec{\beta}|^2} \\ -\beta_y\gamma & (\gamma-1)\frac{\beta_x\beta_y}{|\vec{\beta}|^2} & 1+(\gamma-1)\frac{\beta_y^2}{|\vec{\beta}|^2} & (\gamma-1)\frac{\beta_y\beta_z}{|\vec{\beta}|^2} \\ -\beta_z\gamma & (\gamma-1)\frac{\beta_x\beta_z}{|\vec{\beta}|^2} & (\gamma-1)\frac{\beta_y\beta_z}{|\vec{\beta}|^2} & 1+(\gamma-1)\frac{\beta_z^2}{|\vec{\beta}|^2} \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \equiv B(\mathbf{v}) \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \tag{14.16}$$

Notice that if $v_y = v_z = 0$, then $\beta_y = \beta_z = 0$ and $|\vec{\beta}|^2 = \beta_x^2$, so, in this case $B(\mathbf{v})$ becomes an x-boost, and similarly for the other situations in which the velocity has a non-null component in only one of the coordinate axes.

It is possible to re-write the general boost matrix appearing in eq. (14.16) in a more compact form, see e.g. [21]. First of all, let us use (14.13) to prove the following identity:

$$\frac{\gamma-1}{\beta^2} = \frac{\gamma^2}{\gamma+1}. \tag{14.17}$$

In fact, since $\gamma = \cosh\zeta = \sinh\zeta/\beta$, we have:

$$\begin{aligned}\frac{\gamma-1}{\beta^2} &= \frac{\cosh\zeta-1}{\beta^2} = \frac{(\cosh\zeta-1)(\cosh\zeta+1)}{\beta^2}(\cosh\zeta+1) = \frac{\cosh^2\zeta-1}{\beta^2(\cosh\zeta+1)} \\ &= \frac{\sinh^2\zeta}{\beta^2(\cosh\zeta+1)} = \frac{\gamma^2}{\gamma+1}.\end{aligned}$$

Then, let us notice that

$$\vec{\beta}\vec{\beta}^t = \frac{1}{c^2}\mathbf{v}\mathbf{v}^t = \frac{1}{c^2}\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}\begin{pmatrix} v_x & v_y & v_z \end{pmatrix} = \frac{1}{c^2}\begin{pmatrix} v_x^2 & v_xv_y & v_xv_z \\ v_xv_y & v_y^2 & v_yv_z \\ v_xv_z & v_yv_z & v_z^2 \end{pmatrix},$$

so the general boost matrix appearing in eq. (14.16) can be re-written as

$$B(\mathbf{v}) = \begin{pmatrix} \gamma & -\gamma\vec{\beta}^t \\ -\gamma\vec{\beta} & I_3 + \frac{\gamma^2}{1+\gamma}\vec{\beta}\vec{\beta}^t \end{pmatrix} = \begin{pmatrix} \gamma & -\frac{\gamma}{c}\mathbf{v}^t \\ -\frac{\gamma}{c}\mathbf{v} & I_3 + \frac{\gamma^2}{c^2(1+\gamma)}\mathbf{v}\mathbf{v}^t \end{pmatrix}. \tag{14.18}$$

The matrix associated to the inverse boost is simply $B(\mathbf{v})^{-1} = B(-\mathbf{v})$, i.e.

$$B(\mathbf{v})^{-1} = B(-\mathbf{v}) = \begin{pmatrix} \gamma & \frac{\gamma}{c}\mathbf{v}^t \\ \frac{\gamma}{c}\mathbf{v} & I_3 + \frac{\gamma^2}{c^2(1+\gamma)}\mathbf{v}\mathbf{v}^t \end{pmatrix}. \tag{14.19}$$

14.4.2 Decomposition of proper orthochronous Lorentz transformations

In this section we shall prove important formulae that establish the possibility of writing in a unique way a generic proper orthochronous Lorentz transformation, i.e. an element of the restricted Lorentz group $\mathscr{L}_+^\uparrow$, as the product of Lorentz boost and a spatial rotation that fixes time, in both orders.

For the sake of a simpler formulation of the theorem, it is better to represent $\Lambda \in \mathscr{L}_+^\uparrow$ as follows:

$$\Lambda = \begin{pmatrix} \Lambda_{00} & \Lambda_{01} & \Lambda_{02} & \Lambda_{03} \\ \Lambda_{10} & \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{20} & \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{30} & \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix}.$$

Theorem 14.4.1 *Every proper orthochronous Lorentz transformation $\Lambda \in \mathscr{L}_+^\uparrow$ can be uniquely decomposed as*

$$\Lambda = B(\mathbf{v})\mathcal{R}, \tag{14.20}$$

*where $B(\mathbf{v})$ represents a Lorentz boost in the $\mathbf{v}$-direction and $\mathcal{R}$ is an **extended spatial rotation**, i.e. a block matrix that fixes the time coordinate and rotates the spatial ones, i.e.*

$$\mathcal{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & R_{11} & R_{12} & R_{13} \\ 0 & R_{21} & R_{22} & R_{23} \\ 0 & R_{31} & R_{32} & R_{33} \end{pmatrix} \equiv \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & R \end{array}\right), \tag{14.21}$$

$\mathbf{v}$ is fully determined by the first column of Λ as follows:

$$\mathbf{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = -\frac{c}{\Lambda_{00}} \begin{pmatrix} \Lambda_{10} \\ \Lambda_{20} \\ \Lambda_{30} \end{pmatrix}, \tag{14.22}$$

the Lorentz factor associated to $\mathbf{v}$ is the first matrix element of Λ

$$\gamma = \frac{1}{\sqrt{1 + \frac{|\mathbf{v}|^2}{c^2}}} = \Lambda_{00}, \tag{14.23}$$

finally $R \in SO(3)$ and its entries depend on the matrix elements of Λ in this way:

$$R_{ij} = \Lambda_{ij} - \frac{\Lambda_{i0}\Lambda_{0j}}{\Lambda_{00}+1}, \quad i,j = 1,2,3. \tag{14.24}$$

Proof. First of all we notice that, thanks to theorem 11.3.3 and corollary 11.3.3, if Λ is written as in (14.20) with $B(\mathbf{v})$ a boost and $R \in SO(3)$, then it is a proper orthogonal Lorentz transformation, as a composition of two such transformations[4]. Hence the proof will be accomplished if, by knowing the matrix $\Lambda \in \mathscr{L}_+^\uparrow$ we can express $B(\mathbf{v})$ and $\mathcal{R}$ in terms of its matrix elements and prove the uniqueness of these two matrices.

[4]Notice that $\mathcal{R}$ leaves the temporal coordinate unchanged and preserves the norm of the spatial vector, thus it surely preserves the Minkowski norm.

Thanks to eq. (14.18), the boost matrix is completely specified in terms of the elements of Λ if we know how to write the velocity vector $\mathbf{v}$ and the Lorentz factor γ in terms of the entries of Λ. To this aim let us write explicitly:

$$\begin{aligned}\Lambda = B(\mathbf{v})\mathcal{R} &= \begin{pmatrix} \gamma & -\frac{\gamma}{c}\mathbf{v}^t \\ -\frac{\gamma}{c}\mathbf{v} & I_3 + \frac{\gamma^2}{c^2(1+\gamma)}\mathbf{v}\mathbf{v}^t \end{pmatrix} \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & R \end{array}\right) = \begin{pmatrix} \gamma & -\frac{\gamma}{c}\mathbf{v}^t R \\ -\frac{\gamma}{c}\mathbf{v} & R + \frac{\gamma^2}{c^2(\gamma+1)}\mathbf{v}\mathbf{v}^t R \end{pmatrix} \\ &= \begin{pmatrix} \gamma & -\frac{\gamma}{c}(R^t\mathbf{v})^t \\ -\frac{\gamma}{c}\mathbf{v} & R + \frac{\gamma^2}{c^2(\gamma+1)}\mathbf{v}(R^t\mathbf{v})^t \end{pmatrix}.\end{aligned} \tag{14.25}$$

The equality between the first column of Λ and the first column of the matrix on the right-hand-side, i.e.

$$\begin{pmatrix} \Lambda_{00} \\ \Lambda_{01} \\ \Lambda_{02} \\ \Lambda_{03} \end{pmatrix} = \begin{pmatrix} \gamma \\ -\frac{\gamma}{c}v_x \\ -\frac{\gamma}{c}v_y \\ -\frac{\gamma}{c}v_z \end{pmatrix},$$

implies eqs. (14.23) and (14.22), hence $B(\mathbf{v})$ is completely determined from Λ.

Thanks to this result, we can determine R in terms of Λ simply by observing that $\mathcal{R} = B(\mathbf{v})^{-1}\Lambda = B(-\mathbf{v})\Lambda$. We first notice that

$$\frac{\gamma^2}{c^2(\gamma+1)}\mathbf{v}\mathbf{v}^t = \frac{\cancel{\Lambda_{00}^2}}{\cancel{c^2}(\Lambda_{00}+1)}\frac{\cancel{c^2}}{\cancel{\Lambda_{00}^2}}\begin{pmatrix} \Lambda_{10} \\ \Lambda_{20} \\ \Lambda_{30} \end{pmatrix}(\Lambda_{10}\ \Lambda_{20}\ \Lambda_{30}) = \frac{1}{\Lambda_{00}+1}\begin{pmatrix} \Lambda_{10}^2 & \Lambda_{10}\Lambda_{20} & \Lambda_{10}\Lambda_{30} \\ \Lambda_{10}\Lambda_{20} & \Lambda_{20}^2 & \Lambda_{20}\Lambda_{30} \\ \Lambda_{10}\Lambda_{30} & \Lambda_{20}\Lambda_{30} & \Lambda_{30}^2 \end{pmatrix}$$

and that

$$\frac{\gamma}{c}\mathbf{v} = \frac{\cancel{\Lambda_{00}}}{\cancel{c}}\left(-\frac{\cancel{c}}{\cancel{\Lambda_{00}}}\right)\begin{pmatrix} \Lambda_{10} \\ \Lambda_{20} \\ \Lambda_{30} \end{pmatrix} = -\begin{pmatrix} \Lambda_{10} \\ \Lambda_{20} \\ \Lambda_{30} \end{pmatrix},$$

hence eq. (14.19) gives:

$$B(-\mathbf{v}) = \begin{pmatrix} \Lambda_{00} & -\Lambda_{10} & -\Lambda_{20} & -\Lambda_{30} \\ -\Lambda_{10} & 1 + \frac{\Lambda_{10}^2}{\Lambda_{00}+1} & \frac{\Lambda_{10}\Lambda_{20}}{\Lambda_{00}+1} & \frac{\Lambda_{10}\Lambda_{30}}{\Lambda_{00}+1} \\ -\Lambda_{20} & \frac{\Lambda_{10}\Lambda_{20}}{\Lambda_{00}+1} & 1 + \frac{\Lambda_{20}^2}{\Lambda_{00}+1} & \frac{\Lambda_{20}\Lambda_{30}}{\Lambda_{00}+1} \\ -\Lambda_{30} & \frac{\Lambda_{10}\Lambda_{30}}{\Lambda_{00}+1} & \frac{\Lambda_{20}\Lambda_{30}}{\Lambda_{00}+1} & 1 + \frac{\Lambda_{30}^2}{\Lambda_{00}+1} \end{pmatrix}$$

and so

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & R_{11} & R_{12} & R_{13} \\ 0 & R_{21} & R_{22} & R_{23} \\ 0 & R_{31} & R_{32} & R_{33} \end{pmatrix} = \begin{pmatrix} \Lambda_{00} & -\Lambda_{10} & -\Lambda_{20} & -\Lambda_{30} \\ -\Lambda_{10} & 1 + \frac{\Lambda_{10}^2}{\Lambda_{00}+1} & \frac{\Lambda_{10}\Lambda_{20}}{\Lambda_{00}+1} & \frac{\Lambda_{10}\Lambda_{30}}{\Lambda_{00}+1} \\ -\Lambda_{20} & \frac{\Lambda_{10}\Lambda_{20}}{\Lambda_{00}+1} & 1 + \frac{\Lambda_{20}^2}{\Lambda_{00}+1} & \frac{\Lambda_{20}\Lambda_{30}}{\Lambda_{00}+1} \\ -\Lambda_{30} & \frac{\Lambda_{10}\Lambda_{30}}{\Lambda_{00}+1} & \frac{\Lambda_{20}\Lambda_{30}}{\Lambda_{00}+1} & 1 + \frac{\Lambda_{30}^2}{\Lambda_{00}+1} \end{pmatrix}\begin{pmatrix} \Lambda_{00} & \Lambda_{01} & \Lambda_{02} & \Lambda_{03} \\ \Lambda_{10} & \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{20} & \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{30} & \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix}.$$

The explicit computation of R_{11} gives:

$$\begin{aligned} R_{11} &= -\Lambda_{10}\Lambda_{01} + \left(1 + \frac{\Lambda_{10}^2}{\Lambda_{00}+1}\right)\Lambda_{11} + \frac{\Lambda_{10}\Lambda_{20}}{\Lambda_{00}+1}\Lambda_{21} + \frac{\Lambda_{10}\Lambda_{30}}{\Lambda_{00}+1}\Lambda_{31} \\ &= \Lambda_{11} - \Lambda_{10}\Lambda_{01} + \frac{\Lambda_{10}}{\Lambda_{00}+1}\left(\Lambda_{10}\Lambda_{11} + \Lambda_{20}\Lambda_{21} + \Lambda_{30}\Lambda_{31}\right), \end{aligned}$$

but, using the Lorentz-orthogonality of the first two columns of Λ, it holds that

$$-\Lambda_{00}\Lambda_{01} + \Lambda_{10}\Lambda_{11} + \Lambda_{20}\Lambda_{21} + \Lambda_{30}\Lambda_{31} = 0 \iff \Lambda_{10}\Lambda_{11} + \Lambda_{20}\Lambda_{21} + \Lambda_{30}\Lambda_{31} = \Lambda_{00}\Lambda_{01},$$

so

$$\begin{aligned} R_{11} &= \Lambda_{11} - \Lambda_{10}\Lambda_{01} + \frac{\Lambda_{10}}{\Lambda_{00}+1}\Lambda_{00}\Lambda_{01} = \Lambda_{11} + \frac{\cancel{-\Lambda_{10}\Lambda_{01}\Lambda_{00}} - \Lambda_{10}\Lambda_{01} + \cancel{\Lambda_{10}\Lambda_{00}\Lambda_{01}}}{\Lambda_{00}+1} \\ &= \Lambda_{11} - \frac{\Lambda_{10}\Lambda_{01}}{\Lambda_{00}+1}. \end{aligned}$$

Analogous calculations for the other matrix elements of R give

$$R = \begin{pmatrix} \Lambda_{11} - \frac{\Lambda_{10}\Lambda_{01}}{\Lambda_{00}+1} & \Lambda_{12} - \frac{\Lambda_{10}\Lambda_{02}}{\Lambda_{00}+1} & \Lambda_{13} - \frac{\Lambda_{10}\Lambda_{03}}{\Lambda_{00}+1} \\ \Lambda_{21} - \frac{\Lambda_{20}\Lambda_{01}}{\Lambda_{00}+1} & \Lambda_{22} - \frac{\Lambda_{20}\Lambda_{02}}{\Lambda_{00}+1} & \Lambda_{23} - \frac{\Lambda_{20}\Lambda_{03}}{\Lambda_{00}+1} \\ \Lambda_{31} - \frac{\Lambda_{30}\Lambda_{01}}{\Lambda_{00}+1} & \Lambda_{32} - \frac{\Lambda_{30}\Lambda_{02}}{\Lambda_{00}+1} & \Lambda_{33} - \frac{\Lambda_{30}\Lambda_{03}}{\Lambda_{00}+1} \end{pmatrix},$$

thus confirming eq. (14.24).

Having obtained the explicit form of R, it can be verified by direct computation, using the properties satisfied by the matrix elements of the Lorentzian matrix Λ, that R belongs to $SO(3)$, in fact $\det(R) = 1$ and, if C_i, $i = 1, 2, 3$, are the columns of R, then $\|C_i\|^2 = 1$ and $\langle C_i, C_j\rangle = 0$ for $i \neq j \in 1, 2, 3$, thus they form an orthonormal basis of $\mathbb{R}^3$ and so $R \in SO(3)$.

Finally, we prove the uniqueness of the decomposition: suppose that there exists another couple of Lorentz-boost $B(\mathbf{v}')$ and time-fixing spatial rotation $\mathcal{R}'$ such that

$$\Lambda = B(\mathbf{v})\mathcal{R} = B(\mathbf{v}')\mathcal{R}'.$$

Then, using eq. (14.25) we can write

$$\begin{pmatrix} \gamma_v & -\frac{\gamma_v}{c}(R^t\mathbf{v})^t \\ -\frac{\gamma_v}{c}\mathbf{v} & R + \frac{\gamma_v^2}{c^2(\gamma_v+1)}\mathbf{v}(R^t\mathbf{v})^t \end{pmatrix} = \begin{pmatrix} \gamma_{v'} & -\frac{\gamma_{v'}}{c}((R')^t\mathbf{v}')^t \\ -\frac{\gamma_{v'}}{c}\mathbf{v}' & R' + \frac{\gamma_{v'}^2}{c^2(\gamma_{v'}+1)}\mathbf{v}'((R')^t\mathbf{v}')^t \end{pmatrix},$$

which implies that $\gamma_v = \gamma_{v'}$ and so, from the equality of the matrix element of position $(2, 1)$, it follows that $\mathbf{v} = \mathbf{v}'$. Since a boost matrix is completely determined by the expression of the Lorentz factor and the velocity, we have that $B(\mathbf{v}) = B(\mathbf{v}')$. Moreover, from the equality of the matrix element of position $(1, 2)$ we get $R^t\mathbf{v} = (R')^t\mathbf{v}'$ and this, together with the equality

$$R + \frac{\gamma_v^2}{c^2(\gamma_v+1)}\mathbf{v}(R^t\mathbf{v})^t = R' + \frac{\gamma_{v'}^2}{c^2(\gamma_{v'}+1)}\mathbf{v}'((R')^t\mathbf{v}')^t$$

implies immediately that $R = R'$ and so $\mathcal{R} = \mathcal{R}'$. Thus, the decomposition $\Lambda = B(\mathbf{v})\mathcal{R}$ is unique for all $\Lambda \in \mathscr{L}_+^\uparrow$. □

Other proofs are available, see e.g. [14] theorem 1.3.5 page 28 or [13] for a proof that uses the polar decomposition, further simplified in [24].

An analogous result holds for the decomposition of a proper orthochronous Lorentz transformation with the boost applied first.

Theorem 14.4.2 *Every proper orthochronous Lorentz transformation $\Lambda \in \mathscr{L}_+^\uparrow$ can be uniquely decomposed as*

$$\Lambda = \mathcal{R}B(\mathbf{u}), \tag{14.26}$$

where $\mathcal{R}$ is the same matrix as in theorem 14.4.1 *and $B(\mathbf{u})$ represents a Lorentz boost in the $\mathbf{u}$-direction. Moreover, the velocities $\mathbf{u}$ and $\mathbf{v}$ of theorem* 14.4.1 *are related as follows:*

$$\mathbf{v} = R\mathbf{u} \iff \mathbf{u} = R^t\mathbf{v}.$$

Proof. The proof of the first part of the theorem is identical to the one of theorem 14.4.1. The only thing that remains to be proven is the relationship between $\mathbf{u}$ and $\mathbf{v}$. To this aim let us consider again eq. (14.25):

$$\Lambda = B(\mathbf{v})\mathcal{R} = \begin{pmatrix} \gamma & -\frac{\gamma}{c}(R^t\mathbf{v})^t \\ -\frac{\gamma}{c}\mathbf{v} & R + \frac{\gamma^2}{c^2(\gamma+1)}\mathbf{v}(R^t\mathbf{v})^t \end{pmatrix},$$

but, by direct computation (using the fact that $R^tR = I_3$), it can be verified that

$$\begin{pmatrix} \gamma & -\frac{\gamma}{c}(R^t\mathbf{v})^t \\ -\frac{\gamma}{c}\mathbf{v} & R + \frac{\gamma^2}{c^2(\gamma+1)}\mathbf{v}(R^t\mathbf{v})^t \end{pmatrix} = \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & R \end{array}\right) \begin{pmatrix} \gamma & -\frac{\gamma}{c}(R^t\mathbf{v})^t \\ -\frac{\gamma}{c}R^t\mathbf{v} & I_3 + \frac{\gamma^2}{c^2(1+\gamma)}(R^t\mathbf{v})(R^t\mathbf{v})^t \end{pmatrix} = \mathcal{R}B(R^t\mathbf{v}).$$

As a consequence, $\Lambda = B(\mathbf{v})\mathcal{R} = \mathcal{R}B(R^t\mathbf{v})$, and so, by the uniqueness of the decomposition, it follows that $\mathbf{u} = R^t\mathbf{v}$. □

Corollary 14.4.1 *If two vectors $\mathbf{v}, \mathbf{u} \in \mathbb{R}^3$ with identical magnitude are related by $\mathbf{v} = R\mathbf{u}$, with $R \in SO(3)$, then the boosts with velocities given by the previous vectors are related as follows:*

$$B(\mathbf{u}) = \mathcal{R}^tB(\mathbf{v})\mathcal{R} \iff B(\mathbf{v}) = \mathcal{R}B(\mathbf{u})\mathcal{R}^t,$$

where $\mathcal{R}$ is the extended spatial rotation associated to R.

Proof. From the previous theorem we know that there is a unique $\Lambda \in \mathscr{L}_+^\uparrow$ such that $\Lambda = B(\mathbf{v})\mathcal{R} = \mathcal{R}B(R^t\mathbf{v}) = \mathcal{R}B(\mathbf{u})$, so $B(\mathbf{v})\mathcal{R} = \mathcal{R}B(\mathbf{u})$. Since $\mathcal{R}^{-1} = \mathcal{R}^t$, it follows that

$$B(\mathbf{u}) = \mathcal{R}^tB(\mathbf{v})\mathcal{R}.$$

□

We end this section by showing the explicit formula for the angle and the axis of rotation corresponding to the matrix R as a function of the entries of Λ.

Thanks to formula (1.24), we know that the expression of a matrix $R \in SO(3)$ w.r.t. the angle $\vartheta \in [0, 2\pi)$ of rotation and the unit vector $\mathbf{n} = (n_1, n_2, n_3)^t \in \mathbb{R}^3$ that defines the rotation axis is the following

$$R = \begin{pmatrix} \cos\vartheta + (1-\cos\vartheta)n_1^2 & (1-\cos\vartheta)n_1n_2 - \sin\vartheta n_3 & (1-\cos\vartheta)n_1n_3 + \sin\vartheta n_2 \\ (1-\cos\vartheta)n_1n_2 + \sin\vartheta n_3 & \cos\vartheta + (1-\cos\vartheta)n_2^2 & (1-\cos\vartheta)n_2n_3 - \sin\vartheta n_1 \\ (1-\cos\vartheta)n_1n_3 - \sin\vartheta n_2 & (1-\cos\vartheta)n_2n_3 + \sin\vartheta n_1 & \cos\vartheta + (1-\cos\vartheta)n_3^2 \end{pmatrix}.$$

The trace of R is

$$\mathrm{Tr}(R) = R_{11} + R_{22} + R_{33} = 3\cos\vartheta + (1-\cos\vartheta)(n_1^2 + n_2^2 + n_3^2),$$

but $\mathbf{n}$ is a unit vector, so $\mathrm{Tr}(R) = 2\cos\vartheta + 1$, i.e.

$$\cos\vartheta = \frac{\mathrm{Tr}(R) - 1}{2}.$$

Inserting the expressions of the matrix elements of R in terms of the entries of Λ provided by eq. (14.24) we find

$$\cos\vartheta = \frac{\Lambda_{11} + \Lambda_{22} + \Lambda_{33} - 1}{2} - \frac{\Lambda_{10}\Lambda_{01} + \Lambda_{20}\Lambda_{02} + \Lambda_{30}\Lambda_{03}}{2(\Lambda_{00}+1)}.$$

Then we observe that

$$R_{32} - R_{23} = 2\sin\vartheta n_1,\ R_{13} - R_{31} = 2\sin\vartheta n_2,\ R_{21} - R_{12} = 2\sin\vartheta n_3,$$

so

$$\sin\vartheta\mathbf{n} = \frac{1}{2}\begin{pmatrix} R_{32} - R_{23} \\ R_{13} - R_{31} \\ R_{21} - R_{12} \end{pmatrix},$$

or, using again eq. (14.24),

$$\sin\vartheta\mathbf{n} = \begin{pmatrix} \frac{\Lambda_{32}-\Lambda_{23}}{2} - \frac{\Lambda_{30}\Lambda_{02}-\Lambda_{20}\Lambda_{03}}{2(\Lambda_{00}+1)} \\ \\ \frac{\Lambda_{13}-\Lambda_{31}}{2} - \frac{\Lambda_{10}\Lambda_{03}-\Lambda_{30}\Lambda_{01}}{2(\Lambda_{00}+1)} \\ \\ \frac{\Lambda_{21}-\Lambda_{12}}{2} - \frac{\Lambda_{20}\Lambda_{01}-\Lambda_{10}\Lambda_{02}}{2(\Lambda_{00}+1)} \end{pmatrix}.$$

14.5 Relativistic sum of velocities

Let us consider again the standard configuration and a particle that moves with a uniform velocity. If $(ct, x, y, z)^t$ and $(ct', x', y', z')^t$ are the spacetime coordinates of the particle in the reference frames $\mathscr{R}$ and $\mathscr{R}'$, respectively, then the uniform velocity of the particle is

$$\mathbf{u} = (u_x, u_y, u_z),\ u_x = \frac{dx}{dt},\ u_x = \frac{dy}{dt},\ u_x = \frac{dz}{dt},$$

in the frame $\mathscr{R}$ and

$$\mathbf{u}' = (u'_x, u'_y, u'_z),\ u'_x = \frac{dx'}{dt},\ u'_x = \frac{dy'}{dt},\ u'_x = \frac{dz'}{dt},$$

in the frame $\mathscr{R}'$.

By differentiating eqs. (14.14) we obtain:

$$\begin{cases} dt' = \gamma\left(dt - \frac{v}{c^2}dx\right) \\ dx' = \gamma(dx - vdt) \\ dy' = dy,\ dz' = dz \end{cases}, \tag{14.27}$$

so, by dividing side by side the equations of the second and third line by the equation of the first line, we find:

$$\frac{dx'}{dt'} = \frac{dx - vdt}{dt - \frac{v}{c^2}dx} = \frac{\frac{dx}{dt} - v}{1 - \frac{v}{c^2}\frac{dx}{dt}},$$
$$\frac{dy'}{dt'} = \frac{dy}{\gamma(dt - \frac{v}{c^2}dy)} = \frac{1}{\gamma}\frac{\frac{dy}{dt}}{1 - \frac{v}{c^2}\frac{dy}{dt}},$$
$$\frac{dz'}{dt'} = \frac{dz}{\gamma(dt - \frac{v}{c^2}dz)} = \frac{1}{\gamma}\frac{\frac{dy}{dt}}{1 - \frac{v}{c^2}\frac{dz}{dt}}.$$

Introducing the components of the velocity vectors $\mathbf{u}$ and $\mathbf{u}'$ of the moving particle we obtain the formulae called **Einstein-Poincaré relativistic addition of velocities**:

$$\begin{cases} v \oplus u_x := u'_x = \frac{u_x - v}{1 - \frac{v}{c^2}u_x} \\ v \oplus u_y := u'_y = \frac{1}{\gamma}\frac{u_y}{1 - \frac{v}{c^2}u_y} \\ v \oplus u_z := u'_z = \frac{1}{\gamma}\frac{u_z}{1 - \frac{v}{c^2}u_z} \end{cases} \quad \text{and, analogously,} \quad \begin{cases} v \oplus u'_x := u_x = \frac{u'_x + v}{1 + \frac{v}{c^2}u'_x} \\ v \oplus u'_y := u_y = \frac{1}{\gamma}\frac{u'_y}{1 + \frac{v}{c^2}u'_y} \\ v \oplus u'_z := u_z = \frac{1}{\gamma}\frac{u'_z}{1 + \frac{v}{c^2}u'_z} \end{cases} . \tag{14.28}$$

Notice that in the Newtonian limit, $v \ll c$, we have $\gamma \simeq 1$ and $u'_x = u_x - v$, $u'_y = u_y$, $u'_z = u_z$, i.e. we come back to the Galilean addition law for velocities.

In the special case of a *particle moving uniformly along the x-axis*, we have $u_y = u'_y = u_z = u'_z = 0$, so, writing $u_x = u$ and $u'_x = u'$,

$$v \oplus u = u' = \frac{u - v}{1 - \frac{v}{c^2}u} \quad \text{and} \quad v \oplus u' = u = \frac{u' + v}{1 + \frac{v}{c^2}u'}.$$

Let us use for example the first formula to verify that, **given $u < c$, independently of how close it is to the speed of light, its relativistic sum with v is always strictly less than c**:

$$v \oplus u = u' < \frac{c - v}{1 - \frac{v}{c^2}u} = c^2\frac{c - v}{c^2 - vu} = c\frac{c - v}{c - v\frac{u}{c}} < c, \qquad \forall u, v < c.$$

Moreover, **the relativistic sum of any velocity with the speed of light is again the speed of light**, in fact, if $u = c$,

$$v \oplus c = c' = \frac{c - v}{1 - \frac{v}{c^2}c} = c\frac{c - v}{c - v} = c, \qquad \forall v < c,$$

this, of course, is consistent with the postulate of constancy of the speed of light in every inertial frame.

We can pass from these formulae relative to the standard configuration to those of the general configuration in which the velocity vector $\mathbf{v}$ can have any spatial direction with a similar procedure to that used to extend the formula for the Lorentz boosts. However, the computations needed to extend the Einstein-Poincaré relativistic sum are much more tedious and can be found at the Wikipedia url https://en.wikipedia.org/wiki/Velocity-addition_formula.

The results are:

$$\mathbf{v} \oplus \mathbf{u} := \mathbf{u}' = \frac{1}{1 - \frac{\mathbf{u}\cdot\mathbf{v}}{c^2}}\left[\frac{\mathbf{u}}{\gamma} - \mathbf{v} + \frac{1}{c^2}\frac{\gamma}{1+\gamma}(\mathbf{u}\cdot\mathbf{v})\mathbf{v}\right],$$

and

$$\mathbf{v} \oplus \mathbf{u}' := \mathbf{u} = \frac{1}{1 + \frac{\mathbf{u}' \cdot \mathbf{v}}{c^2}} \left[\mathbf{v} + \frac{\mathbf{u}'}{\gamma} + \frac{1}{c^2} \frac{\gamma}{1+\gamma} (\mathbf{u}' \cdot \mathbf{v}) \mathbf{v} \right].$$

The presence of the dot product is explained by the fact that $\mathbf{u}$, or $\mathbf{u}'$, must be decomposed on the direction parallel and orthogonal w.r.t. $\mathbf{v}$.

An important consequence of the lack of commutativity of non-collinear Lorentz boosts is that **the Einstein-Poincaré relativistic sum is neither commutative nor associative**.

Remark: David Mermin has managed to show that the results of special relativity can be derived by postulating the relativistic addition of velocities, see [12]. This highly non-trivial result permits to characterize as 'relativistic' also theories that do not deal with light speed!

14.5.1 Relativistic aberration of light

In this subsection we use the same notations of the main section, except for the fact that we suppose that the velocity vector $\mathbf{u}$ of a moving particle lies in the plane formed by the axes X and Y.

If u, u' represent the modulus of $\mathbf{u}$ in $\mathscr{R}$ and $\mathscr{R}'$, respectively, and ϑ, ϑ' are the angles formed between $\mathbf{u}$ and the axes X and X', respectively, then

$$\begin{cases} \mathbf{u} = (u_x, u_y) = (u\cos\vartheta, u\sin\vartheta) \\ \mathbf{u}' = (u'_x, u'_y) = (u'\cos\vartheta, u'\sin\vartheta). \end{cases}$$

Replacing these expression of u_x, u_y, u'_x, u'_y in eqs. (14.28) we find:

$$\frac{\cancel{u}\sin\vartheta}{\cancel{u}\cos\vartheta} = \frac{1}{\gamma} \frac{u'\sin\vartheta'}{\cancel{1 + u\cos\vartheta v/c^2}} \frac{\cancel{1 + u\cos\vartheta v/c^2}}{u'\cos\vartheta' + v},$$

i.e.

$$\tan\vartheta = \frac{1}{\gamma} \frac{u'\sin\vartheta'}{u'\cos\vartheta' + v}.$$

The Lorentz factor and the velocity term v at the denominator of the right-hand side of this last equation prevent to get the identity $\tan\vartheta = \tan\vartheta'$, this implies that $\vartheta \neq \vartheta'$, i.e. the velocity vectors $\mathbf{u}$ and $\mathbf{u}'$ do not have the same direction.

If we are not dealing with particles with mass, but with rays of light, then $u = u' = c$ and, dividing numerator and denominator by c, the previous formula becomes:

$$\tan\vartheta = \frac{1}{\gamma} \frac{\sin\vartheta'}{\cos\vartheta' + \beta},$$

which is known as the **relativistic light aberration formula**.

14.6 Minkowski diagrams

A Minkowski diagram is a 2-dimensional *spacetime diagram*, i.e. a graphical illustration of the properties of space and time, in which the vertical axis is parameterized by ct and the horizontal one is parameterized by only one spatial coordinate, typically x.

With such a choice of the vertical axis, the world-line of a ray of light is a straight line with a slope $\pi/2$ and any world-line of a particle moving uniformly with speed $v < c$ is a straight line with slope given by the fraction of the speed of light, i.e. the coefficient $\beta = v/c$, as depicted in Figure 14.1.

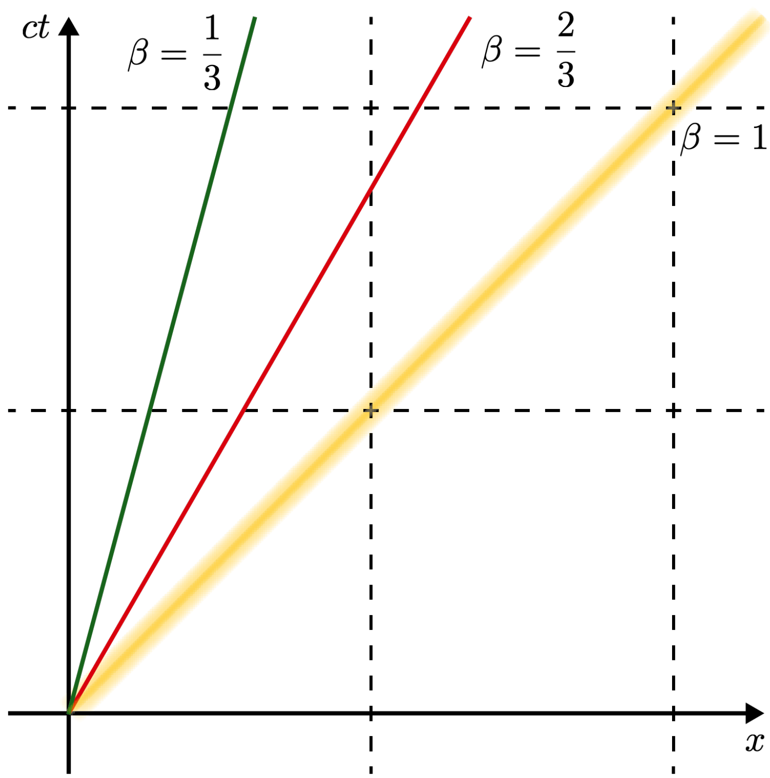


Figure 14.1: The fraction of the speed of light $\beta = v/c$ determining the slope of the world-line of a particle with uniform speed v.

As a consequence, **the world-line of a particle at rest in an inertial frame is the ct axis**, of course if we take as origin of the space coordinates the place where the particle stays still.

Let us remember that the 2-dimensional formulae for the x-Lorentz boost are given by:

$$\begin{cases} ct' = \gamma(ct - \beta x) \\ x' = \gamma(x - \beta ct) \end{cases},$$

being linear, these equations imply that the axes defining the coordinates ct' and x' of an inertial frame co-moving with the particle are **tilted towards the world-line of a ray of light by the same angle, which is determined by** β, as depicted in Figure 14.2. The larger is v, the greater is this tilt. For $v \ll c$, the tilt reduces to the Galilean one.

The fact that the tilt angle is the same should not be surprising: since we are analyzing a Lorentz transformation, the Lorentz-orthogonality between the axes x and ct must be preserved and, as we saw in section 11.2.4, this can be accomplished only if the sum of the angles that the new axes x' and ct' form with x is $\pi/2$.

The Lorentz factor $\gamma = 1/\sqrt{1-\beta^2}$ appears as a scaling coefficient equal for the axes ct and x. If γ were not the same for ct and x, the speed of light would be measured differently, which contradicts the third postulate. To better understand the geometric action of γ, let us recall that the 2-dimensional matrix associated with the Lorentz boost is:

$$B(\beta) = \gamma \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix} \implies \det(B(\beta)) = \gamma^2(1-\beta^2) = 1,$$

since $\det(B(\beta))$ represents the change in area of a coordinate grid box after the Lorentz boost, this means that, thanks to the suitable scaling operated by the Lorentz factor, the area of any grid box is preserved. Figure 14.2 gives a visual representation of this fact.

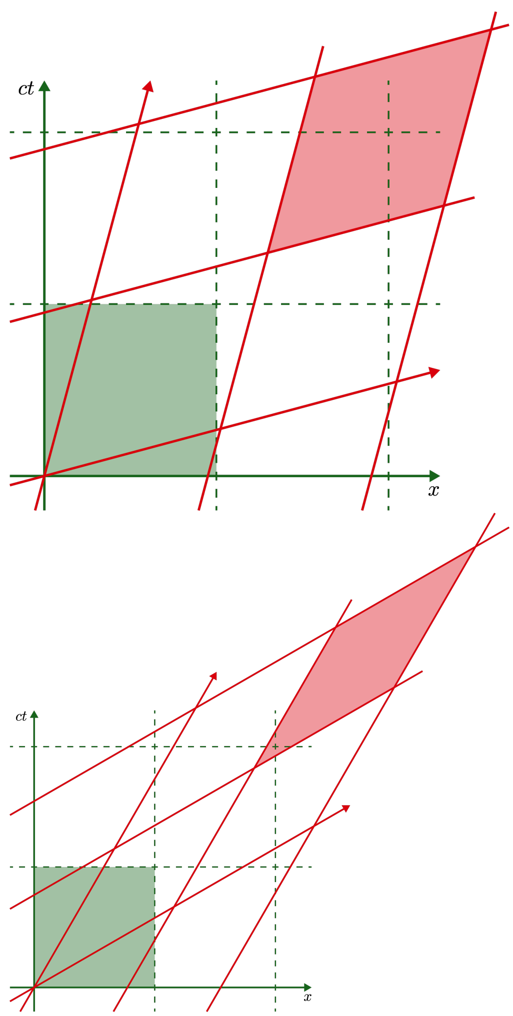


Figure 14.2: The relative position of the axes w.r.t. the original frame and frame co-moving with the particle. *First image*: $\beta = 1/3$, so $\gamma \simeq 1.06$. *Second image*: $\beta = 2/3$, so $\gamma \simeq 1.34$.

In Figure 14.3 it is shown that the tilt of the axes follows a 'hyperbolic rotation': this is of course a consequence of the fact that the Minkowski norm is preserved by a Lorentz transformation and that the level lines of this norm are hyperbola in the 2D Minkowski diagrams.

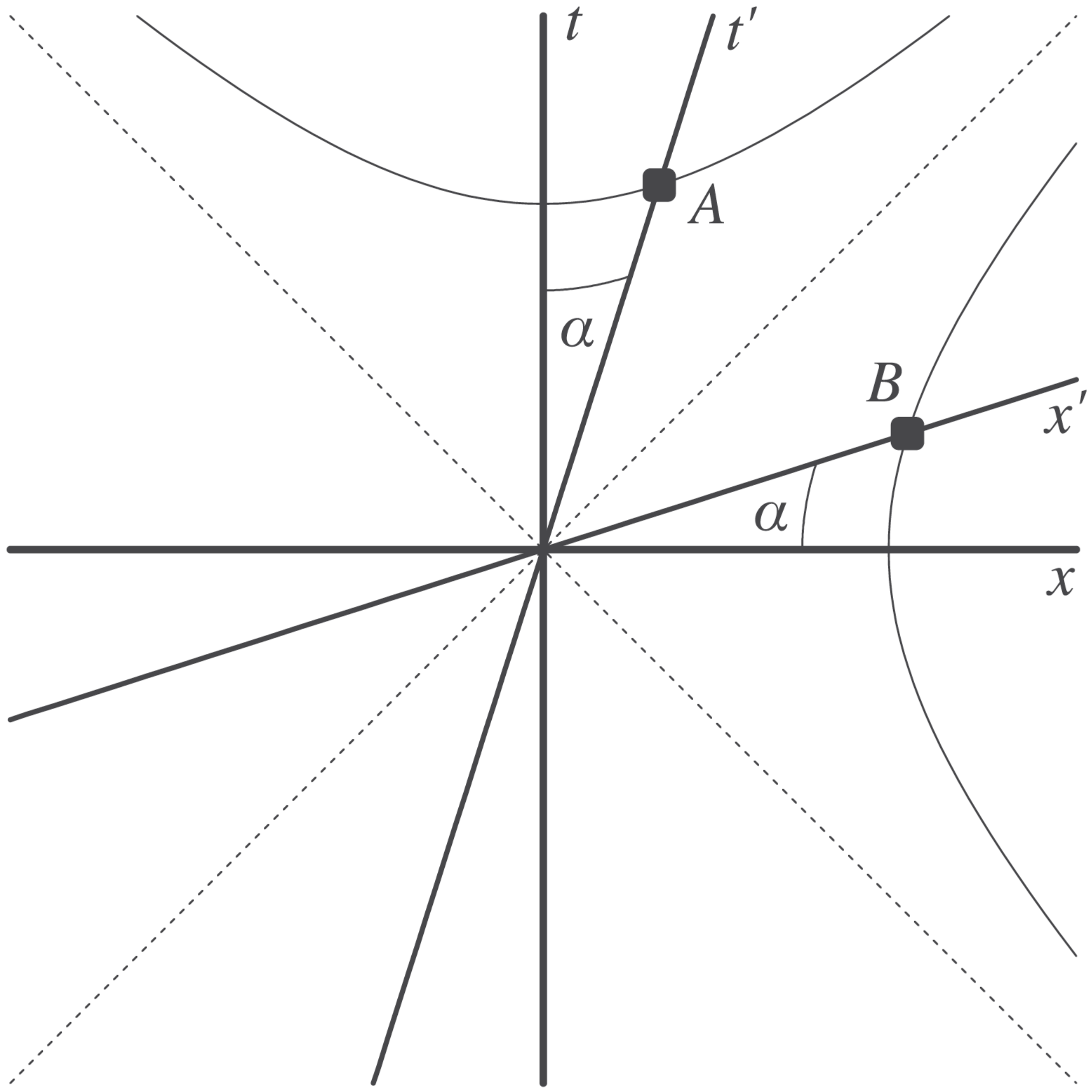


Figure 14.3: The hyperbolic rotation of the tilted axes after a x-Lorentz boost.

14.7 3-dimensional Lorentz geometry and split-quaternions

In section 1.3.3 we have recalled the basic properties of quaternions and in the subsequent sections we have related them to the geometry of the *3D-Euclidean space*, by showing, in particular, that Rodrigues' formula for the rotation of a vector in $\mathbb{R}^3$ around an axis can be performed via the conjugation with a suitable unit quaternion.

It is possible to obtain analogous results for the *3D Lorentz space*, a hyperbolic version of Rodrigues' formula. In order to do that we must replace quaternions with the so-called split quaternions, which are the topic of the present section.

The split-quaternions (or coquaternions), introduced in 1849 by James Cockle, form an algebraic structure similar to quaternions, as they define a 4-dimensional associative and non-commutative algebra, however they are not a division algebra because, as we will point out soon, there exist infinite non-trivial split-quaternions which are not invertible.

The elements of the split-quaternion algebra[5], indicated by $\mathbb{S}$, have the same form of quaternions, i.e. $q = q_0 + q_1 i + q_2 j + q_3 k$, $q_0, q_1, q_2, q_3 \in \mathbb{R}$, but the multiplication rules of i, j, k are different. They can be resumed by the formula

$$-i^2 = j^2 = k^2 = ijk = 1,$$

from which we may deduce the multiplicative properties

- $i^2 = -1$ and $j^2 = k^2 = 1$
- $k = ij$, $j = ki$ and $i = kj$
- $ij = -ji$, $ik = -ki$ and $kj = -jk$.

As it was the case for quaternions, it is useful to use the scalar/vector decomposition $q = s_q + \mathbf{v}_q$, with $s_q = q_0$ (the scalar part) and $\mathbf{v}_q = q_1 i + q_2 j + q_3 k$ (the vector part). If $s_q = 0$, i.e. if $q = \mathbf{v}_q$, then q is called a **pure split-quaternion** and the set of pure split quaternions is denoted as follows:

$$\mathbb{S}_0 = \{q = s_q + \mathbf{v}_q \in \mathbb{S},\ s_q = 0\} \cong \mathbb{R}^{1,2}.$$

The conjugate of a split-quaternion $q \in \mathbb{S}$ is defined as

$$\overline{q} = q_0 - q_1 i - q_2 j - q_3 k = s_q - \mathbf{v}_q,$$

the squared modulus is given by

$$|q|^2 = q\overline{q} = \overline{q}q = q_0^2 + q_1^2 - q_2^2 - q_3^2, \tag{14.29}$$

and reveals *the reason underlying the term 'split'*: the signs in the modulus of $q \in \mathbb{S}$ are equally partitioned (split) in two negative and two positive signs.

The most important aspect to stress about the previous formula is that **the modulus of the vector part $\mathbf{v}_q$ of a split quaternion is the Lorentz pseudo-norm in the 3-Lorentz space** $\mathbb{R}^{1,2}$ and no longer its Euclidean norm, as it was the case for quaternions.

This fact suggests to distinguish among split-quaternions using the same nomenclature introduced in chapter 11, i.e.

[5]The split-quaternion algebra can also be viewed as the Clifford algebra $C\ell(2,0)$, i.e the Clifford algebra $C\ell(\mathcal{V},\mathcal{Q})$ of the vector space $\mathcal{V} = \mathbb{R}^2$ with quadratic form $\mathcal{Q}(v) = v_1^2 + v_2^2$.

- $q \in \mathbb{S}$ is **time-like** if $|q|^2 > 0$,
- $q \in \mathbb{S}$ is **light-like** if $|q|^2 = 0$,
- $q \in \mathbb{S}$ is **space-like** if $|q|^2 < 0$.

Fixed $i \in \{0,1,2,3\}$, if we introduce $q_i = \delta_{ij}$, $j = 0,1,2,3$, in eq. (14.29), we find that 1 **and** i **are time-like split quaternion** while j **and** k **are space-like split quaternions**.

Space-like and time-like split-quaternions have a multiplicative inverse given by

$$q^{-1} = \frac{\overline{q}}{|q|^2},$$

but light-like split-quaternions do not have an inverse, this is why $\mathbb{S}$ is not a division algebra. Notice that if $|q|^2 = 1$ (so, necessarily, q is time-like), then $q^{-1} = \overline{q}$.

We recall that, given two quaternions $p, q \in \mathbb{H}$, eq. (1.15) established that their product can be written as follows

$$pq = (s_p s_q - \langle \mathbf{v}_p, \mathbf{v}_q \rangle) + (s_p \mathbf{v}_q + s_q \mathbf{v}_p + \mathbf{v}_p \times \mathbf{v}_q), \quad p, q \in \mathbb{H},$$

instead, for two split-quaternions $p, q \in \mathbb{S}$, after straightforward tedious computations, we have

$$pq = (s_p s_q - \mathbf{v}_p \circ \mathbf{v}_q) + (s_p \mathbf{v}_q + s_q \mathbf{v}_p - \mathbf{v}_p \otimes \mathbf{v}_q), \quad p, q \in \mathbb{S}, \tag{14.30}$$

where we see that the Euclidean product is replaced by the Lorentz pseudo-scalar product

$$\mathbf{v}_p \circ \mathbf{v}_q = p_1 q_1 - p_2 q_2 - p_3 q_3,$$

and the cross product is replaced by minus the Lorentzian cross product

$$\mathbf{v}_p \otimes \mathbf{v}_q = \eta(\mathbf{v}_p \times \mathbf{v}_q) = (p_2 q_3 - p_3 q_2) i + (p_1 q_3 - p_3 q_1) j - (p_1 q_2 - q_1 p_2) k,$$

where $\eta = \mathrm{diag}(1, -1, -1)$.

As a special case, if we set $p = q$, then $\mathbf{v}_p \otimes \mathbf{v}_q = \mathbf{0}$ and:

$$q^2 = (s_q^2 - \|\mathbf{v}_q\|_{\mathcal{L}}^2) + (2 s_q \mathbf{v}_q), \quad q \in \mathbb{S}, \tag{14.31}$$

where $\|\mathbf{v}_q\|_{\mathcal{L}}^2 = \mathbf{v}_q \circ \mathbf{v}_q$ is the squared Lorentz pseudo-norm of $\mathbf{v}_q$. Notice that $\mathbf{v}_{q^2} = 2 s_q \mathbf{v}_q$, so q and q^2 have a collinear vector part.

As in $\mathbb{H}$, the scalar part of the product in $\mathbb{S}$ is commutative, i.e. $s_{pq} = s_{qp}$ for all $p, q \in \mathbb{S}$ and we also have:

$$\overline{pq} = \overline{q}\,\overline{p}, \qquad \forall p, q \in \mathbb{S},$$

in fact:

$$\begin{aligned}
\overline{pq} &= \overline{(s_p s_q - \mathbf{v}_p \circ \mathbf{v}_q) + (s_p \mathbf{v}_q + s_q \mathbf{v}_p - \mathbf{v}_p \otimes \mathbf{v}_q)} = (s_p s_q - \mathbf{v}_p \circ \mathbf{v}_q) + (-s_p \mathbf{v}_q - s_q \mathbf{v}_p + \mathbf{v}_p \otimes \mathbf{v}_q) \\
&= (s_p s_q - \mathbf{v}_p \circ \mathbf{v}_q) + (-s_p \mathbf{v}_q - s_q \mathbf{v}_p - \mathbf{v}_q \otimes \mathbf{v}_p) \\
&= (s_p s_q - \mathbf{v}_p \circ \mathbf{v}_q) + (-s_p \mathbf{v}_q - s_q \mathbf{v}_p - (-\mathbf{v}_q) \otimes (-\mathbf{v}_p)) \underset{(14.30)}{=} \overline{q}\,\overline{p}.
\end{aligned}$$

If a split quaternion q is space-like, then its vector part $\mathbf{v}_q$ is a space-like vector of $\mathbb{R}^{1,2}$, in fact if in this case $q_1^2 - q_2^2 - q_3^2 = \|\mathbf{v}_q\|_{\mathcal{L}}^2 < -q_0^2 < 0$.

Instead, if q is time-like, then $\mathbf{v}_q$ can be a space-, time-, or light-like vector of $\mathbb{R}^{1,2}$ because this time we can only know that $q_1^2 - q_2^2 - q_3^2 = \|\mathbf{v}_q\|_{\mathcal{L}}^2 > -q_0^2$.

The set of space-like split quaternions is not a group w.r.t. the product of $\mathbb{S}$ because the product of two space-like split quaternions is always time-like, as it can be verified with straightforward computations. Instead, time-like split quaternions indeed form a group w.r.t. the product of $\mathbb{S}$.

As it happens with quaternions, also split quaternions have a tight link with the Pauli matrices: in fact the matrices

$$\tilde{\sigma}_2 = i\sigma_2 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

satisfy the same multiplication rules of $i, j, k \in \mathbb{S}$. If we add the matrix $\sigma_0 = I_2$, we find a basis of $M(2,\mathbb{R})$ and the map $\mathbb{S} \ni (1, i, j, k) \mapsto (\sigma_0, \tilde{\sigma}_2, \sigma_1, \sigma_3) \in M(2,\mathbb{R})$ turns out to be an algebra isomorphism which allows us to uniquely represent any fixed split quaternion with a matrix as follows:

$$q = 1\cdot q_0 + i\cdot q_1 + j\cdot q_2 + k\cdot q_3 \quad \longleftrightarrow \quad q = \sigma_0 q_0 + \tilde{\sigma}_2 q_1 + \sigma_1 q_2 + \sigma_3 q_3 = \begin{pmatrix} q_0 + q_3 & q_2 + q_1 \\ q_2 - q_1 & q_0 - q_3 \end{pmatrix} =: A_q.$$

The split quaternion multiplication can be encoded in the multiplication of the corresponding matrices, i.e. $A_{qp} = A_q A_p$, for all $q, p \in \mathbb{S}$.

As for quaternions, the determinant and the trace of A_q carry important information:

$$\det(A_q) = q_0^2 + q_1^2 - q_2^2 + q_3^2 = |q|^2, \quad \frac{1}{2}\mathrm{Tr}(A_q) = s_q, \qquad \forall q \in \mathbb{S}.$$

It follows that, when we interpret the elements of $\mathbb{S}$ as matrices, the multiplicative group of split-quaternions with nonzero modulus is isomorphic to $\mathrm{GL}(2,\mathbb{R})$ and the group of unit split quaternions is isomorphic to $\mathrm{SL}(2,\mathbb{R})$.

Moreover, given $p, q \in \mathbb{S}$, since $|p| = \sqrt{\det(A_p)}$ and $|q| = \sqrt{\det(A_q)}$, we obtain

$$|pq| = |p|\,|q|, \qquad \forall p, q \in \mathbb{S}.$$

Proposition 14.7.1 *Let $p \in \mathbb{S}_0$, then:*

1. *for all $q \in \mathbb{S}$, the operation $p \mapsto qp\overline{q}$ is an endomorphism of $\mathbb{S}_0$*
2. *for all $q \in \mathbb{S}_0$, the operation $p \mapsto qpq$ is an endomorphism of $\mathbb{S}_0$.*

Proof. The first statement can be proven as follows: thanks to the commutativity of the scalar part of the split quaternion product, $s_{qp\overline{q}} = s_{p|q|^2}$, which is null because $s_p = 0$ and, since $|q|^2 \in \mathbb{R}$, $\mathbf{v}_{|q|^2} = \mathbf{0}$, so eq. (14.30) implies

$$p|q|^2 = (\cancelto{0}{s_p} s_{|q|^2} - \mathbf{v}_p \circ \cancelto{\mathbf{0}}{\mathbf{v}_{|q|^2}}) + (\cancelto{0}{s_p}\mathbf{v}_{|q|^2} + s_{|q|^2}\mathbf{v}_p - \mathbf{v}_p \otimes \cancelto{\mathbf{0}}{\mathbf{v}_{|q|^2}}) = s_{|q|^2}\mathbf{v}_p \implies s_{p|q|^2} = 0.$$

Regarding the second statement we have: $s_{qpq} = s_{pq^2}$, but $s_p = 0$ so eq. (14.30) implies $s_{qpq} = s_{pq^2} = -\mathbf{v}_p \circ \mathbf{v}_{q^2}$, which is null because we have already seen that $\mathbf{v}_{q^2} = 2s_q\mathbf{v}_q = \mathbf{0}$ since $s_q = 0$. $\square$

The definition of norm of a split quaternion is different from the one given for quaternions: since in $\mathbb{H}$ the modulus is positive-definite, we had the possibility to define the norm of $q = q_0 + q_1 i + q_2 j + q_3 k \in \mathbb{H}$ simply as $|q| = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$, instead for split quaternions we define the norm as follows (see e.g. [16] page 326):

$$N_q := \sqrt{|q_0^2 + q_1^2 - q_2^2 - q_3^2|} = \sqrt{|\,|q|^2\,|}, \qquad q = q_0 + q_1 i + q_2 j + q_3 k \in \mathbb{S},$$

and we say that q is a **unit split quaternion** if $N_q = 1$, it follows that the set of unit split quaternions can be written explicitly as

$$\mathbb{S}_1 = \{q = s_q + \mathbf{v}_q \in \mathbb{S},\ |q|^2 = \pm 1\},$$

we also write the set of unit time-like and space-like split quaternions, respectively, as follows:

$$\mathbb{S}_1^+ = \{q = s_q + \mathbf{v}_q \in \mathbb{S},\ |q|^2 = 1\}, \quad \mathbb{S}_1^- = \{q = s_q + \mathbf{v}_q \in \mathbb{S},\ |q|^2 = -1\}.$$

It is clear that for split quaternions we have $\overline{q} = q^{-1}$ if and only if $q \in \mathbb{S}_1^+$.

Recall now that, by proposition 1.3.1, the square roots of -1 in $\mathbb{H}$ are isomorphic to S^2. Analogously, **the square roots of -1 in $\mathbb{S}$ are pure unit time-like split quaternions**, which are isomorphic to the **upper unit hyperboloid $\mathcal{H}$ in** $\mathbb{R}^{1,2}$.

Proposition 14.7.2 *Let $q \in \mathbb{S}$, then $q^2 = -1$ if and only if $q \in (\mathbb{S}_0 \cap \mathbb{S}_1^+) = \mathcal{H}$.*

Proof. Let $q = a + ib + jc + kd = s_q + \mathbf{v}_q \in \mathbb{S}_0 \cap \mathbb{S}_1^+$, then $|q|^2 = \|\mathbf{v}_q\|^2_{\mathcal{L}} = 1$, so, by eq. (14.31), $q^2 = -\|\mathbf{v}_q\|^2_{\mathcal{L}} = -1$.

Let now $q \in \mathbb{S}$ satisfy $q^2 = -1$, then, again by eq. (14.31), we have:

$$\begin{cases} s_q^2 - \|\mathbf{v}_q\|^2_{\mathcal{L}} = a^2 + b^2 - c^2 - d^2 = -1 \\ 2ab = 0 \\ 2ac = 0 \\ 2ad = 0. \end{cases}$$

If a were not 0, the last three equations would imply $b = c = d = 0$ and the first equation would give $a^2 = -1$, which is impossible because $a \in \mathbb{R}$. Hence $a = s_q = 0$, so $q \in \mathbb{S}_0$ and also $|q|^2 = \|\mathbf{v}_q\|^2_{\mathcal{L}} = 1$, so $q \in \mathbb{S}_1^+$. $\square$

As for the quaternions, also for the split quaternions we have an Euler exponential formula related to the square roots of -1.

Proposition 14.7.3 (Euler's formula for $\mathbb{S}$) *Let $q \in \mathbb{S}$ be a square root of -1. Then, for all $t \in \mathbb{R}$,*

$$e^{tq} = \cos(t) + \sin(t) q.$$

The proof is identical to the one provided in section 1.4.3.

14.7.1 Split quaternions and Lorentz transformations

In chapter 1 we have seen that quaternions allow us to elegantly represent rotations of vectors in the 3-dimensional Euclidean space via conjugation by unit quaternions. Having seen that split-quaternions are associated to the 3-dimensional Lorentz space, it is not surprising that they allow to represent Lorentz transformations via conjugation by unit split quaternions.

In particular, we are interested in proper orthochronous Lorentz transformations belonging to $\mathrm{SO}^+(1,2)$, the restricted Lorentz group $\mathscr{L}_+^\uparrow$. Thanks to theorem 14.4.1, if we prove that both rotations and boosts can be written via the conjugation by a unit split quaternion, then the result will follow.

Let us start with rotations: given $q = q_0 + q_1 i + q_2 j + q_3 k \in \mathbb{S}$, consider the angle $\theta \in [0, 2\pi)$ and the unit split quaternion

$$\boxed{r_\theta := \cos(\theta/2) + \sin(\theta/2) i},$$

clearly $r_\theta \in \mathbb{S}_1^+$ and so $r_\theta^{-1} = \overline{r_\theta}$, with a straightforward calculation it can be verified that

$$r_\theta\, q\, \overline{r_\theta} = q_0 + q_1 i + (\cos\theta q_2 - \sin\theta q_3) j + (\sin\theta q_2 + \cos\theta q_3) k.$$

If we write q as a column vector $(q_0, q_1, q_2, q_3)^t$, then

$$r_\theta\, q\, \overline{r_\theta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \equiv \mathcal{R}_\theta\, q,$$

where $\mathcal{R}_\theta$ is an extended rotation of the form

$$\mathcal{R}_\theta = \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & R_\theta \end{array}\right),$$

$R_\theta \in \mathrm{SO}(3)$ represents a rotation of $\mathbf{v}_q = (q_1, q_2, q_3)^t \in \mathbb{R}^{1,2}$ by the angle θ in the plane defined by the space-like split quaternion units j and k, about the axis defined by the time-like unit i.

Being i a square root of -1, we can write the previous result in the following exponential form:

$$\mathcal{R}_\theta\, q = e^{i\frac{\theta}{2}}\, q\, e^{-i\frac{\theta}{2}}, \qquad q \in \mathbb{S}.$$

If, instead of the conjugation, we consider the two-side product by r_θ we obtain

$$r_\theta\, q\, r_\theta = (\cos\theta q_0 - \sin\theta q_1) + (\cos\theta q_1 + \sin\theta q_0) i + q_2 j + q_3 k,$$

or, in matrix form

$$r_\theta\, q\, r_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \equiv \tilde{\mathcal{R}}_\theta\, q = e^{i\frac{\theta}{2}}\, q\, e^{i\frac{\theta}{2}}, \quad q \in \mathbb{S},$$

where $\tilde{\mathcal{R}}_\theta$ is an extended rotation of the form

$$\tilde{\mathcal{R}}_\theta = \left(\begin{array}{c|c} \tilde{R}_\theta & \mathbf{0} \\ \hline \mathbf{0}^t & 1 \end{array}\right),$$

$\tilde{R}_\theta \in \mathrm{SO}(3)$ *represents a rotation of* $\mathbf{v}_q = (q_0, q_1, q_2)^t \in \mathbb{R}^{1,2}$ *by the angle* θ *in the plane defined by the time-like split quaternion units* 1 *and* i*, about the axis defined by the time-like unit* j*.*

Let us now pass to boosts. Given the rapidity $\zeta \in \mathbb{R}$, and

$$\boxed{b_\zeta := \cosh(\zeta/2) + \sinh(\zeta/2)j},$$

since $|b_\zeta|^2 = \cosh^2(\zeta/2) - \sinh^2(\zeta/2) = 1$, we have $b_\zeta \in \mathbb{S}_1^+$ and $b_\zeta^{-1} = \overline{b_\zeta}$. With a direct computation, taken $q = q_0 + q_1 i + q_2 j + q_3 k \in \mathbb{S}$, we obtain:

$$\begin{aligned} b_\zeta\, q\, \overline{b_\zeta} = q_0 &+ [(\cosh^2(\zeta/2) + \sinh^2(\zeta/2))q_1 - 2\cosh(\zeta/2)\sinh(\zeta/2)q_3]i + q_2 j \\ &+ [-2\cosh(\zeta/2)\sinh(\zeta/2)q_1 + (\cosh^2(\zeta/2) + \sinh^2(\zeta/2))q_3]k, \end{aligned}$$

which, thanks to the identities $\cosh^2(\zeta/2) + \sinh^2(\zeta/2) = \cosh(\zeta)$ and $2\cosh(\zeta/2)\sinh(\zeta/2) = \sinh(\zeta)$, can be re-written as

$$b_\zeta\, q\, \overline{b_\zeta} = q_0 + (\cosh(\zeta)q_1 - \sinh(\zeta)q_3)i + q_2 j + (-\sinh(\zeta)q_1 + \cosh(\zeta)q_3)k.$$

By interpreting again q as a column vector $(q_0, q_1, q_2, q_3)^t$, we can write the transformation $q \mapsto b_\zeta q \overline{b_\zeta}$ as follows:

$$b_\zeta\, q\, \overline{b_\zeta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cosh(\zeta) & 0 & -\sinh(\zeta) \\ 0 & 0 & 1 & 0 \\ 0 & -\sinh(\zeta) & 0 & \cosh(\zeta) \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \equiv \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & B_k(\zeta) \end{array}\right) q,$$

or, using the identities $\cosh\zeta = \gamma$ and $\sinh\zeta = \beta\gamma$,

$$b_\zeta\, q\, \overline{b_\zeta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \gamma & 0 & -\beta\gamma \\ 0 & 0 & 1 & 0 \\ 0 & -\beta\gamma & 0 & \gamma \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \equiv \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & B_k(\beta) \end{array}\right) q.$$

Since the scalar part q_0 is kept fixed by the conjugation, it is clear that if we interpret the i-axis as the ct-axis of section 14.4, then $B_k(\beta)$ and $B_k(\zeta)$ represent **a Lorentz boost of** $\mathbf{v}_q \in \mathbb{R}^{1,2}$ **in the** k**-th direction**.

If, instead of the conjugation, we consider the two-side product by b_ζ we obtain

$$b_\zeta\, q\, b_\zeta = [\cosh(\zeta)q_0 + \sinh(\zeta)q_2] + q_1 i + [\cosh(\zeta)q_2 + \sinh(\zeta)q_0]j + q_3 k,$$

or, in matrix form,

$$b_\zeta\, q\, \overline{b_\zeta} = \begin{pmatrix} \cosh(\zeta) & 0 & \sinh(\zeta) & 0 \\ 0 & 1 & 0 & 0 \\ \sinh(\zeta) & 0 & \cosh(\zeta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \equiv \left(\begin{array}{c|c} B_j(-\zeta) & \mathbf{0} \\ \hline \mathbf{0}^t & 1 \end{array}\right) q,$$

$B_j(-\zeta)$ represents **an inverse Lorentz boost of** $(q_0, q_1, q_2)^t \in \mathbb{R}^{1,2}$ **in the** j**-th direction**.

If we consider $c_\zeta := \cosh(\zeta/2) + \sinh(\zeta/2)k$, then again $c_\zeta \in \mathbb{S}_1^+$ and

$$c_\zeta\, q\, \overline{c_\zeta} = q_0 + (\cosh(\zeta)q_1 + \sinh(\zeta)q_2)i + (\sinh(\zeta)q_1 + \cosh(\zeta)q_2)j + q_3 k,$$

which, interpreting q as a column vector $(q_0, q_1, q_2, q_3)^t$, gives

$$c_\zeta\, q\, \overline{c_\zeta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cosh(\zeta) & \sinh(\zeta) & 0 \\ 0 & \sinh(\zeta) & \cosh(\zeta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \equiv \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & B_j(-\zeta) \end{array}\right) q,$$

or

$$c_\zeta\, q\, \overline{c_\zeta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \gamma & \beta\gamma & 0 \\ 0 & \beta\gamma & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \equiv \left(\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mathbf{0} & B_j(-\beta) \end{array}\right) q,$$

where $B_j(-\zeta)$ and $B_j(-\beta)$ represent **an inverse Lorentz boost of $\mathbf{v}_q \in \mathbb{R}^{1,2}$ in the j-th direction**.

Finally, let us consider the two-side product by c_ζ, which gives

$$c_\zeta\, q\, c_\zeta = [\cosh(\zeta)q_0 + \sinh(\zeta)q_3] + q_1 i + q_2 j + [\cosh(\zeta)q_3 + \sinh(\zeta)q_0]k,$$

or, in matrix form,

$$c_\zeta\, q\, c_\zeta = \begin{pmatrix} \cosh(\zeta) & 0 & 0 & \sinh(\zeta) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh(\zeta) & 0 & 0 & \cosh(\zeta) \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix},$$

i.e. an inverse Lorentz boost in the k-th direction.

If we extend this result by showing that a Lorentz boost in any direction can be written as a conjugation by a unit split quaternion, then the proof that every proper orthochronous Lorentz transformation can be obtained through this kind of operations will be accomplished.

TO BE WRITTEN...

Def. 14.7.1 ($\pm$ split with respect to a square root of -1) *Let $p \in \mathbb{S}$ be a square root of -1. For $q \in \mathbb{S}$, we define the $\pm$ split associated to p as*

$$q_+ = \frac{1}{2}(q + pqp^*) \quad \textit{and} \quad q_- = \frac{1}{2}(q - pqp^*)$$

The $\pm$ split $q = q_+ + q_-$ produces a unique decomposition into **commuting** and **anticommuting** parts, i.e $pq_+ = q_+p$ and $pq_- = -q_-p$.
Most importantly, from Euler's formula we obtain

$$e^{tp}q_+ = q_+e^{tp},$$

and

$$e^{tp}q_- = q_-e^{-tp}.$$

Lemma 14.7.1 *A split quaternion $q \in \mathbb{S}$ commutes with p, a square root of -1, if and only if $v_q \in span(p)$.*

Proof.

$$
\begin{aligned}
pq = qp &\iff p(s_q + v_q) = (s_q + v_q)p \\
&\iff pv_q = v_q p \\
&\iff p \otimes v_q + \cancel{(p \circ v_q)} = v_q \otimes p + \cancel{(p \circ v_q)} \\
&\iff p \otimes v_q = 0 \\
&\iff v_q \in \operatorname{span}(p)
\end{aligned}
$$

□

Chapter 15

Rudiments of general relativity

The concepts of connection, curvature, geodesics, etc. introduced at the end of chapter **??**, find a natural application in the general theory of relativity of Albert Einstein. This theory, published in 1915, Einstein proposes a completely novel way to describe gravity, w.r.t. the Newtonian theory, in terms of curvature of spacetime.

In this chapter we will only see some rudiments of Einstein's theory, the reader interested in a more thorough description is referred to the several treatises about general relativity available in literature.

We start with the main assumptions of the theory. The spacetime is modeled to be a pseudo-Riemannian manifold (M, g) where M has dimension 4, 3 of which are needed to describe space and 1 for time, and g has Lorentzian signature, i.e. $(1, 3)$.

In the vacuum, i.e. in a region of spacetime without mass or energy, the metric is taken to be

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mu, \nu = 0, \dots, 3,$$

called Minkowski metric. The units of measure have been chosen in such a way that the speed of light is set to $c = 1$.

On the contrary, Einstein assumed that the presence of mass or energy in a region of spacetime the metric $g_{\mu\nu}$ is modified and it is no longer the Minkowski metric $\eta_{\mu\nu}$, while the spacetime manifold M remains the same topologically. So, the metric $g_{\mu\nu}$ of M becomes the unknown variable of some dynamical equations, proposed by Einstein and since then called after him. The **Einstein equations** are:

$$\boxed{R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}}, \qquad \mu, \nu = 0, \dots, 3, \tag{15.1}$$

where:

- $R_{\mu\nu}$ is the Ricci tensor of g and R is the corresponding scalar curvature;
- G is Newton's constant of universal gravitation;
- $T_{\mu\nu}$ is the stress energy tensor which depends on the distribution of mass and energy inside the spacetime region under investigation.

All the tensors appearing in Einstein's equations are symmetric, so we pass from the 20 independent equations corresponding to the independent components of a manifold of dimension 4 (see chapter **??**), to 10.

They are **highly non-linear partial differential equations of order 2** for the components of the metric $g_{\mu\nu}$ because, as seen in chapter **??**, $R_{\mu\nu}$ (and thus also R), can be expressed in terms of the Christoffel symbols, which, in turn, can be written in terms of $g_{\mu\nu}$ and its second-order partial differential equations w.r.t. the spacetime coordinates.

For simplicity, if we introduce the Einstein gravitational tensor whose components are

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu},$$

then the first Bianchi identity can be translated in the following four equations

$$\nabla^{\mu} G_{\mu\nu} = 0, \qquad \nu = 0, \ldots, 3$$

where, denoted with $\nabla_{\mu} \equiv \nabla_{\partial_{\mu}}$ the covariant derivative along $\frac{\partial}{\partial x^{\mu}}$ induced by the Levi-Civita connection of M, we define ∇^{μ} as usual by raising the index thanks to the inverse of the metric, i.e. $\nabla^{\mu} = g^{\mu\nu} \nabla_{\nu}$.

The 4 Bianchi identities written above reduce the independent Einstein equations to 6, which remain hugely complicated to solve. In fact, in general the Einstein equations cannot be solved exactly unless in very simple cases, as we will see shortly.

Einstein subsequently proposed another version of his equation, namely:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \qquad \mu, \nu = 0, \ldots, 3,$$

where $\Lambda \in \mathbb{R}$ is called **cosmological constant**. To understand its meaning, let us suppose to be in a region of space free from mass and energy, so that $T_{\mu\nu} = 0$ and the previous equations become

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = -\Lambda g_{\mu\nu}, \qquad \mu, \nu = 0, \ldots, 3,$$

compared with the original proposal of Einstein, we see that, even in empty spacetime, there appear to exists 'something', i.e. $-\Lambda g_{\mu\nu}$, that replaces the stress-energy tensor, for this reason $-\Lambda g_{\mu\nu}$ is interpreted as the 'energy density' intrinsically possessed by the vacuum. In cosmology, this energy is supposed to contribute to the expansion of the universe, even if its role and its existence are still very much debated.

In summary the main idea underlying Einstein's general theory of relativity can be resumed in the following steps:

1. a spacetime region of empty space is supposed to be the flat Riemannian manifold $(M, \eta_{\mu\nu})$;

2. if, instead of being empty, M is filled with a known distribution of mass and energy, then we compute the stress-energy tensor $T_{\mu\nu}$;

3. by solving, exactly or approximately, the Einstein equations we obtain $g_{\mu\nu}$, this creates a novel Riemannian manifold $(M, g_{\mu\nu})$, where M remains the same, but the metric on it changes: the original flat Minkowski metric $\eta_{\mu\nu}$ has been bent into $g_{\mu\nu}$ by the presence of a non-null $T_{\mu\nu}$;

4. another hugely important, and bold, assumption of general relativity is that *the bending of spacetime replaces the concept of gravitational force*, so that everything, including light, is supposed to *free-fall* in $(M, g_{\mu\nu})$, i.e. not to be subjected to any force;

5. in geometrical terms, as seen in chapter **??**, this is translated in the fact that everything travels along the geodesics of $(M, g_{\mu\nu})$;

6. the geodesics of every object can be computed thanks to $g_{\mu\nu}$ which leads to the computation of the coefficients of the Levi-Civita connection, i.e. the Christoffel symbols ${\Gamma_{\mu\nu}}^{\sigma}$.

15.1 The Schwarzschild solution of Einstein's equations

The first exact solution of the Einstein equations was found by Schwarzschild (literally: black shield) in the case of a spherically symmetric gravitational field for empty regions of space. Of course, the motivation that led Schwarzschild to work under this hypothesis is twofold: on one side, the spherical symmetry provides a strong simplification of the mathematical complexity underlying Einstein equations, on the other side, the gravitation field produced by a non-rotating spherical object is known to possess this kind of symmetry in Newtonian gravitation. Stars and planet have approximately spherical shape, but they rotate, so the Schwarzschild solution, even though mathematically exact, is only valid with a certain degree of precision for what concerns physics. Searching for solutions of the Einstein equations for empty regions of space means simply that we are interested in determining how the star or the planet bends spacetime at their exterior, in a portion of the universe without mass or energy.

Of course, under such an hypothesis, it is ideal to use the spherical coordinates for what concerns space, i.e.

$$\begin{cases} x^1 = r\sin\vartheta\cos\varphi \\ x^2 = r\sin\vartheta\sin\varphi \\ x^3 = r\cos\vartheta \end{cases} .$$

The expression of the spatial part of the metric in spherical coordinates is

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 = dr^2 + r^2(d\vartheta^2 + \sin^2\vartheta d\varphi^2) \equiv dr^2 + r^2 d\Omega^2.$$

By using the coordinates $(t, r, \vartheta, \varphi)$ (with $c = 1$), instead of x^0, x^1, x^2, x^3, it can be proven that the most general metric that satisfies the request of spherical symmetry, i.e. whose spatial part varies only with r and not with the angles, is

$$ds^2 = m(t,r)dt^2 + n(t,r)dr^2 + r^2 d\Omega^2,$$

where $m(t,r)$ and $n(t,r)$ are two unknown function of t and r. We can see that the request of spherical symmetry has hugely simplified the problem, because in the expression of the unknown metric now there are only two functions that must be determined.

From ds^2 we can compute the Christoffel symbols ${\Gamma_{\mu\nu}}^{\sigma}$ of the Levi-Civita connection and from them we can compute the components R_{ijhk} of the Riemann curvature tensor, the components $R_{\mu\nu}$ of the Ricci tensor and the scalar curvature R.

Since we are looking at the solutions of Einstein equations in vacuum, we introduce what we have found in the formula

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 0$$

and we solve them determining once and for all the functions $m(t, r)$ and $n(t, r)$.

The metric that one finds, called **Schwarzschild metric**, has the following expression:

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2,$$

where M is a constant that has the physical interpretation of the mass that produces the gravitational field.

If we use reintroduce the speed of light c, the Schwarzschild metric has the following expression:

$$ds^2 = -\left(1 - \frac{2GM}{rc^2}\right) c^2 dt^2 + \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 + r^2 d\Omega^2,$$

which shows that there are two radii for which two denominators become 0:

- $r = 0$, which nullifies the fraction in the first term, is a real singularity of the metric, in fact it can be checked that the scalar curvature $R \underset{r \to 0}{\to} \infty$. However, the Schwarzschild metric has no physical meaning in $r = 0$ because that value of the radio surely refers to a point inside the mass that produced the gravitational field, which goes against the fact that the metric obtained comes from the Einstein equations in vacuum!

- $r = 2GM/c^2$, which nullified the coefficient of the second term, which appears elevated to -1, so it is actually a denominator. This value of the radius is not a real singularity, but an apparent one: in fact R does not diverge when $r \to 2GM/c^2$. This apparent contradiction can be actually easily explained: by the fact that this 'singularity' only appears in spherical coordinates, by using other coordinates, i.e. another local chart, the no singularity appears for that value of r.

$r_S = 2GM/c^2$ is called **Schwarzschild radius**. In most cases, r_S is well inside the mass that created the gravitational field and so it does not correspond to a real solution of Einstein equations. For instance, for the Sun $r_S \simeq 3km$, against the sun radius of $\simeq 7 \cdot 10^5 km$ and for the Earth $r_S \simeq 9mm$, against the Earth radius of $\simeq 6400km$. However, for black holes, the mass M is so huge that overcomes c^2 at the denominator and r_S lies outside them. In this case, it represent the **event horizon** of the black hole.

Having at disposal the Schwarzschild metric, we can compute the geodesics and determine the orbits of the planets around the Sun. These ones happen to be almost ellipses as in the Newtonian theory of gravity, however the slight deviation from the elliptical orbit can explain with a great precision phenomena such as the precession of the perihelion of Mercury, which the Newtonian theory cannot explain. So far, no experimental test has proven wrong any prediction of Einstein general relativity.

15.2 Tidal forces in general relativity

As we have seen at the beginning of this chapter, one of the main assumptions of the theory of general relativity is that *points* in a spacetime region M move along the geodesics of (M, g), where g is the pseudo-Riemannian metric generated by the distribution of mass and energy around those points.

We can ask ourselves what happens if we do not consider points but *extended objects*. In this case, since geodesics passing through different points of this object are different and not, in general, 'covariantly parallel' to each other, the different points of extended object tend to move differently, thus deforming the object itself. Of course the electromagnetic interactions between the different parts of the object tend to oppose to this gravitational deformation, but if either the object is elastic or the gravitational field is strong enough, the deformation will occur. The final effect is called '**tidal force**' experienced by the object.

This is, e.g., how general relativity explains water tides of seas and oceans or the so-called 'spaghettification' when a body enters in a black hole.

15.3 Time dilation in general relativity and its consequences on the GPS system

The coordinate t that is used in the Schwarzschild solution of Einstein's equations *is not* the time measured by a clock placed in some point of the spacetime. Exactly as the other parameters, i.e. r, ϑ, φ, t is only a coordinate that allows us to parameterize explicitly the spacetime points through a local chart and, as such, does not have an intrinsic meaning. For this reason, t is sometimes called *coordinate time*.

The time τ measured by a clock, called **proper time**, can be obtained from the relationship $c^2 d\tau^2 + ds^2 = 0$, which leads to $d\tau^2 = -ds^2/c^2$, by integrating $d\tau$ along a world-line, which depends on the metric, hence, in general relativity, also time depends on gravity.

In the case of the Schwarzschild metric we have:

$$d\tau^2 = \left(1 - \frac{2GM}{rc^2}\right) dt^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} \frac{dr^2}{c^2} - \frac{r^2}{c^2} d\Omega^2.$$

Let us use this expression for a very interesting application: the general-relativistic correction needed to make the GPS system function correctly. Consider an artificial satellite which orbits the Earth with tangential speed of constant modulus v along a planar circular orbit at a distance r from the center of the Earth. We can of course suppose that the mass m of the artificial satellite is negligible w.r.t. the mass M of the Earth, so that the presence of m does not alter the gravitational field generated by the Earth.

Since the orbit of the artificial satellite is circular the radius r is constant and because of the planar assumption we can arbitrarily fix one of the two angles, e.g. $\vartheta = \pi/2$ to simplify the computations, we obtain

$$\begin{cases} r(t) = r \quad \text{(constant)} \\ \vartheta(t) = \frac{\pi}{2} \quad \text{(constant)} \\ \varphi(t) = \omega t, \end{cases}$$

where $\omega = v/r$ is the angular velocity, which is a constant too because v and r are, thence, $dr = 0$, $d\vartheta = 0$, $d\varphi = \omega dt$.

Recalling that $d\Omega^2 = d\vartheta + \sin^2\vartheta d\varphi^2$, thanks to the fact that we have chosen $\vartheta(t) = \frac{\pi}{2}$ we obtain $d\Omega^2 = d\varphi^2 = \omega^2 dt^2$.

Introducing these expressions of the differentials in the formula for $d\tau^2$ we find

$$d\tau^2 = \left(1 - \frac{2GM}{rc^2}\right) dt^2 - \frac{r^2\omega^2}{c^2} d\Omega^2 dt^2 = \left(1 - \left(\frac{2GM}{rc^2} + \frac{r^2\omega^2}{c^2}\right)\right) dt^2,$$

which shows that the proper time measured by a clock on the artificial satellite is linked to the coordinate time t by a correction factor that depends on the mass of the Earth, the distance at which the satellite orbits our planet and its angular speed.

The artificial satellites used for the GPS system are equipped with an atomic clock which measures time with great precision. Of course this clock measure the proper time of the satellite. Usual GPS satellite parameters are the following: $r \simeq 26.6 \cdot 10^6 m$ and $v \simeq 3.87 \cdot 10^3 m/s$, G, M and c are known, so we find

$$d\tau_{\text{sat}} = \sqrt{1 - \left(\frac{2GM}{rc^2} + \frac{r^2\omega^2}{c^2}\right)}\, dt \simeq \sqrt{1 - 0.498 \cdot 10^{-9}} dt \simeq (1 - 0.249 \cdot 10^{-9}) dt,$$

having used the Taylor expansion of the square root: $\sqrt{1+\varepsilon} \simeq 1 + \frac{1}{2}\varepsilon - \frac{1}{8}\varepsilon^2$.

Instead, for a clock on the surface of the Earth we have to use as r the Earth radius $r \simeq 6.37 \cdot 10^6 m$, which leads to

$$d\tau_{\text{Earth}} \simeq \sqrt{1 - 1.392 \cdot 10^{-9}} dt \simeq (1 - 0.696 \cdot 10^{-9}) dt.$$

From these computations we find

$$d\tau_{\text{sat}} = \frac{1 - 0.249 \cdot 10^{-9}}{1 - 0.696 \cdot 10^{-9}}\, d\tau_{\text{Earth}},$$

since $1/(1+\varepsilon)$ is the sum of the geometric series, if we stop it at the linear term we find

$$d\tau_{\text{sat}} = (1 - 0.249 \cdot 10^{-9})(1 + 0.696 \cdot 10^{-9})\, d\tau_{\text{Earth}} \simeq (1 + (0.696 - 0.249) \cdot 10^{-9})\, d\tau_{\text{Earth}},$$

having discarded the products of the small terms which is of order 10^{-18}. So, finally

$$d\tau_{\text{sat}} \simeq (1 + 0.447 \cdot 10^{-9})\, d\tau_{\text{Earth}},$$

which means that aboard the satellite the proper time flows faster than on the Earth[1].

Each time a clock measures $1s$ on the surface of the Earth, the satellite clock goes ahead of it by the quantity $0.447 \cdot 10^{-9}$. Since in an Earth day we have $86400s$, in a day the satellite clock accumulates $0.447 \cdot 10^{-9} \cdot 8.64 \cdot 10^4 s = 38.6 \cdot 10^{-6} s = 38.6 \mu s$. over the Earth clock. This may seem a negligible effect, however, since the GPS communications with the satellite is performed through radio waves, which travel at the speed of light $c \simeq 3 \cdot 10^8 m/s$, we see that, if the clock discrepancies were not continuously compensated, in only a day the GPS system would fail to pinpoint a place on Earth to hundreds of meters!

The fact that the compensation computed by the equations of general relativity allows us to use the GPS system with great precision is another confirmation of the soundness of Einstein's theory of general relativity.

[1] In general, this discrepancy increases with the mass M that generated the gravitational field. In extreme cases as black holes it can become so large that when an object reached the event horizon it is seen by a sufficiently large distance as completely still.

Part IV:

Appendices

Appendix A

Einstein's convention

In differential geometry, we often deal with expressions with many indices and sums. To simplify the notation, it is common to use **Einstein's convention** and implicitly assume a sum over repeated indices *above and below* in an algebraic expression, the sum being of course performed over the range of index variability, e.g. if $i = 1, \ldots, n$, then

$$a^i b_i := \sum_{i=1}^{n} a_i b_i.$$

This notation is consistent as long as we agree to write the indices below for the basis vectors of $\mathbb{R}^n$ and above for the components w.r.t. them. The convention for the dual space $(\mathbb{R}^n)^*$ is inverted. Coherently with that, the **canonical basis of** $\mathbb{R}^n$ will be denoted with $(e_i)_{i=1}^n$, while its **dual basis** will be written as $(\varepsilon^j)_{j=1}^n$, $\varepsilon^j \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}) \equiv (\mathbb{R}^n)^*$, the two bases are linked via the pairing:

$$\varepsilon^j(e_i) = \delta_i^j, \qquad i, j = 1, \ldots, n.$$

Given the vector $x = x^i e_i \in \mathbb{R}^n$, $x^i \in \mathbb{R}$, for all $i = 1, \ldots, n$, the action of the linear functional ε^j on x is:

$$\varepsilon^j(x) = \varepsilon^j(x^i e_i) = x^i \varepsilon^j(e_i) = x^i \delta_i^j = x^j,$$

i.e. ε^j **simply extracts the** j**-th component of the vector** $x \in \mathbb{R}^n$ **w.r.t. the canonical basis** $(e_i)_{i=1}^n$.

Vectors in $\mathbb{R}^n$, or any other vector space V, will always be considered as column vectors, while their duals, belonging to $(\mathbb{R}^n)^*$, or V^*, will be considered as row vectors.

It is very important to make explicit the use of the Einstein convention when we deal with matrices associated with *linear maps between vector spaces* and with *bilinear forms on a vector space*. Let $f : V \to W$ be a linear function between the vector spaces V and W of dimension n and m, respectively, then, if we denote the matrix associated to f with $A = (a_j^i)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, i.e. we write the **row index above** and the **column index below**, the Einstein convention can be coherently applied to compute the product of A with a *column vector* $v = (v^1, \ldots, v^n)^t$ of V, in fact:

$$Av = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} = \begin{pmatrix} a_1^1 v^1 + \cdots + a_n^1 v^n \\ \vdots \\ a_1^m v^1 + \cdots + a_n^m v^n \end{pmatrix} = (a_j^i v^j),$$

which is a column vector with m rows belonging to W. Notice that, if the matrix is square, then the **trace** of A is $\text{Tr}(A) = a^i_i$.

Now we examine the case of a bilinear form. Let $g : V \times V \to \mathbb{R}$ be an $\mathbb{R}$-bilinear form over the vector space V of dimension n, then, by fixing a basis $(u_1, \ldots, u_n)$ of V we can associate to g the matrix $G = (g_{ij})$, where the matrix elements are defined by the formula:

$$g(u_i, u_j) := g_{ij},$$

so that $G = (g_{ij})_{i,j=1,\ldots,n}$. Notice that now the matrix elements of G are written with two indices below, this is the only way of being coherent with Einstein's notation, in fact, if $v, w \in V$, $v = (v^i)$ and $w = (w^j)$ where v^i and w^j, $i, j = 1, \ldots, n$, are the components of v and w w.r.t. the basis (u_i) of V, then

$$g(v, w) = g_{ij} v^i w^j := \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} v_i w_j,$$

will be a real scalar, as correctly expected.

Appendix B

Recap of ordinary calculus in $\mathbb{R}^n$

We collect here some basic results and definitions of ordinary calculus in $\mathbb{R}^n$. We assume that the reader is already familiar with this topic, the aim of this appendix is just to recap the most important concepts of standard calculus.

In particular, we stress some concepts, as e.g. the spaces between which partial derivatives act or the role of the dual basis of $\mathbb{R}^n$ or that of the evaluation map, that are sometimes hidden when presenting ordinary calculus but that are essential for the development of differential calculus on manifolds.

It is convenient to fix the notation that will be used, unless otherwise specified, in this appendix:

- $x_0 \in \Omega \subseteq \mathbb{R}^n$, Ω open set
- $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$
- $L(\mathbb{R}^n, \mathbb{R}^m)$ is the vector space of linear operators from $\mathbb{R}^n$ to $\mathbb{R}^m$
- $U(0)$ is an open neighborhood of the null vector $0 \in \mathbb{R}^n$
- $U(x_0)$ is the open neighborhood of x_0 obtained by translation of $U(0)$ by the vector x_0:

$$U(x_0) = \{x_0 + h,\ h \in U(0)\}$$

- a **curve**, or **path**, in $\mathbb{R}^d$, $d \geqslant 1$, is a continuous function $\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^d$, where I is an open real interval.
- Modulo a translation and a rescaling, it is always possible to consider I to be $(-\varepsilon, \varepsilon)$ for a suitable $\varepsilon > 0$.
- We say that γ **passes through** $x_0 \in \mathbb{R}^d$ if $\gamma(0) = x_0$.

The main idea behind differential calculus in $\mathbb{R}^n$ is the concept of **local linearization**, which leads directly to the definition of derivative. For functions of only one real variable there is only one derivative, but for functions of more than one real variable two different (and not equivalent) derivatives can be considered: the total and the directional derivative along a vector. We start recalling the definition of the total derivative, which has been formalized as follows by Fréchet.

Def. B.0.1 *f is said to be Fréchet-differentiable (or simply differentiable) in $x_0 \in \Omega$ if there exist:*

- *an open neighborhood $U(0) \subseteq \Omega$*
- *a linear operator*[1] *$Df(x_0) \in L(\mathbb{R}^n, \mathbb{R}^m)$, that, in general, depends on x_0*
- *a rest function $\rho_{x_0} : U(0) \subseteq \mathbb{R}^n \to \mathbb{R}^m$,*

such that:

1. $f(x_0 + h) = f(x_0) + Df(x_0)h + \rho_{x_0}(h), \qquad \forall h \in U(0)$

2. $\rho_{x_0}(0) = 0$

3. $\frac{\|\rho_{x_0}(h)\|}{\|h\|} \underset{\|h\|\to 0}{\longrightarrow} 0$, *i.e.* $\rho_{x_0}(h) = o(\|h\|)$.

f is differentiable on Ω if it is differentiable in every point of Ω.

This definition is the precise formalization of the intuitive statement that it is possible to approximate the action of f on nearby points $x = x_0 + h$ around x_0 by a linear function and that the error in doing this tends to zero faster than the distance between x and x_0, i.e. $\|h\| = \|x - x_0\|$.

Def. B.0.2 *$Df(x_0)$ is called the **total derivative**, the **Fréchet derivative**, or simply the derivative of f in x_0.*

Condition **1.** and **3.** imply an important equation, to find its expression let us rewrite **1.** as $f(x_0 + h) - f(x_0) - Df(x_0)h = \rho_{x_0}(h)$, so that $\|f(x_0 + h) - f(x_0) - Df(x_0)h\| = \|\rho_{x_0}(h)\|$ thus, dividing by $\|h\|$ and taking the limit $\|h\| \to 0$, thanks to **3.** we obtain:

$$\lim_{\|h\|\to 0} \frac{\|f(x_0 + h) - f(x_0) - Df(x_0)h\|}{\|h\|} = 0. \tag{B.1}$$

Theorem B.0.1 (Uniqueness of the total derivative) *If $Df(x_0)$ exists, then it is unique.*

Proof. We must proof that if $D_1f(x_0)$ and $D_2f(x_0)$ are two total derivatives of f in x_0, then they must agree as linear operators belonging to $L(\mathbb{R}^n, \mathbb{R}^m)$.

To this aim, observe that (B.1) implies, in particular, that the numerator tends to 0 as $\|h\| \to 0$, i.e. $\forall \varepsilon > 0\ \exists \delta_\varepsilon > 0$ such that $\|h\| < \delta_\varepsilon$ implies both

$$\|f(x_0 + h) - f(x_0) - D_1f(x_0)h\| < \frac{\varepsilon}{2}\|h\| \ \text{ and } \ \|f(x_0 + h) - f(x_0) - D_2f(x_0)h\| < \frac{\varepsilon}{2}\|h\|, \tag{B.2}$$

having used the arbitrariness of ε. Now, thanks to the triangular inequality, we have:

$$\begin{aligned} \|D_1f(x_0)h - D_2f(x_0)h\| &= \|f(x_0 + h) - f(x_0) - D_2f(x_0)h - (f(x_0 + h) - f(x_0) - D_1f(x_0)h)\| \\ &< \|f(x_0 + h) - f(x_0) - D_2f(x_0)h\| + \|f(x_0 + h) - f(x_0) - D_1f(x_0)h\| \\ &\underset{(B.2)}{<} \varepsilon\|h\|. \end{aligned}$$

[1]Sometimes $Df(x_0)$ is written as $f'(x_0)$.

Noticing that $D_1 f(x_0)h - D_2 f(x_0)h = D_1 f(x_0) - D_2 f(x_0))h$, we can write

$$\frac{\|(D_1 f(x_0) - D_2 f(x_0))h\|}{\|h\|} < \varepsilon \qquad \forall \|h\| < \delta_\varepsilon.$$

When $h = 0$, it is clear that $D_1 f(x_0)h = D_2 f(x_0)h = 0$ because the total derivative is linear, so, let us consider $h \neq 0$, then, from the previous expression we get:

$$\|D_1 f(x_0) - D_2 f(x_0)\| := \sup_{h \neq 0} \frac{\|(D_1 f(x_0) - D_2 f(x_0))h\|}{\|h\|} < \varepsilon \qquad \forall \|h\| < \delta_\varepsilon,$$

which implies that $D_1 f(x_0) = D_2 f(x_0)$. □

Because of the uniqueness of the total derivative, many authors say that $Df(x_0)$ **provides the best linear approximation of f in a neighborhood of x_0.**

Remark: if we replace $\mathbb{R}^n$ and $\mathbb{R}^m$ by any two finite-dimensional normed spaces, then the definitions and results above remain valid.

The reason why $Df(x_0)$ is called the total derivative is that it contains, as special cases, all the derivatives of f in x_0 along any possible directions, as we are going to formalize.

Def. B.0.3 *The* ***straight line passing through*** x_0 ***and directed as the vector*** $v \in \mathbb{R}^n$ *is the curve in* $\mathbb{R}^n$ *defined as follows:*

$$\begin{array}{rccl} r_{x_0,v}: & \mathbb{R} & \longrightarrow & \mathbb{R}^n \\ & t & \longmapsto & r_{x_0,v}(t) = x_0 + tv. \end{array}$$

In order to define the concept of directional derivative, we just need to observe that the composed function $f \circ r_{x_0,v} : \mathbb{R} \to \mathbb{R}^m$ is a curve in $\mathbb{R}^m$ passing through $f(x_0) = f(r_{x_0,v}(0))$.

Def. B.0.4 *Given* $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *and* $x_0 \in \Omega$, *if the following limit (in* $\mathbb{R}^m$*) exists*[2]

$$D_v f(x_0) = \lim_{t \to 0} \frac{(f \circ r_{x_0,v})(t) - f(x_0)}{t} = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} \equiv (f \circ r_{x_0,v})^{\cdot}(0), \qquad \text{(B.3)}$$

then we call it the ***directional derivative of the function in*** x_0 ***along the vector*** v.

We say that f is ***Gateaux differentiable*** *in x_0 if the directional derivatives of f in x_0 exists for every direction v. f is Gateaux differentiable on Ω if it is Gateaux differentiable in every point of Ω.*

As a particular vector v we can choose e_i, the i-th element of the canonical basis of $\mathbb{R}^n$, in this case the directional derivative of f in x_0 is called **partial derivative** of f in x_0 and denoted with

$$\frac{\partial f}{\partial x^i}(x_0) := D_{e_i} f(x_0).$$

Each function

$$\begin{array}{rccl} f: & \Omega \subseteq \mathbb{R}^n & \longrightarrow & \mathbb{R}^m \\ & x & \longmapsto & f(x) = (y_1, \ldots, y_m), \end{array}$$

[2]Notice that $D_v f(x_0)$ *is a vector* in $\mathbb{R}^m$ and *not* a linear operator.

is uniquely associated to an ordered collection of m real-valued functions, the so-called **component functions** $f^1, \dots, f^m : \Omega \to \mathbb{R}$, defined as follows:

$$\forall x \in \Omega, \quad f(x) = (y_1, \dots, y_m) =: (f^1(x), \dots, f^m(x)).$$

The partial derivatives $\frac{\partial f^j}{\partial x^i}(x_0)$, $i = 1, \dots, n$, $j = 1, \dots, m$, of the component functions can be organized in a $m \times n$ matrix with real entries called **Jacobian matrix** of f in x_0 and denoted with $Jf(x_0) \in M(m \times n, \mathbb{R})$:

$$(Jf(x_0))^j_i = \frac{\partial f^j}{\partial x^i}(x_0) \iff Jf(x_0) = \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(x_0) & \dots & \frac{\partial f^1}{\partial x^n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x^1}(x_0) & \dots & \frac{\partial f^m}{\partial x^n}(x_0) \end{pmatrix}. \tag{B.4}$$

Let us now prove that the vector $D_v f(x_0) \in \mathbb{R}^m$ can be recovered by $Df(x_0) \in L(\mathbb{R}^n, \mathbb{R}^m)$ simply by applying this linear operator to the vector v, it is in this sense that the total derivative contains all the information on the directional derivatives.

The easiest and more profound way to prove this relationship is by first examining the special case $n = 1$, i.e. curves $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^m$.

In standard differential calculus we prove that a function of one real variable, as γ, is differentiable in $x_0 \in (-\varepsilon, \varepsilon)$ if and only if the limit

$$\dot{\gamma}(x_0) := \lim_{t \to 0} \frac{\gamma(x_0 + t) - \gamma(x_0)}{t} \equiv \frac{d\gamma}{dt}(x_0) \in \mathbb{R}^m$$

exists and it is finite. In this case, $\dot{\gamma}(x_0)$ is called *the value of the derivative* of γ in x_0.

The fact that, in this special case, the existence of the total derivative of γ in x_0, i.e. the linear operator $D\gamma(x_0) \in L(\mathbb{R}, \mathbb{R}^m)$, is equivalent to the existence of its derivative $\dot{\gamma}(x_0)$ in x_0 should not be surprising if we think about the canonical identification of the vector space $L(\mathbb{R}, \mathbb{R}^m)$ with $\mathbb{R}^m$ via the linear isomorphism given by

$$\begin{array}{ccc} L(\mathbb{R}, \mathbb{R}^m) & \xrightarrow{\sim} & \mathbb{R}^m \\ T & \longmapsto & T1, \end{array} \tag{B.5}$$

i.e. the application of any linear operator $T \in L(\mathbb{R}, \mathbb{R}^m)$ to the only element of the canonical basis of $\mathbb{R}$, i.e. 1.

Let us use again the special element 1 of $\mathbb{R}$ to define the directional derivative of γ in x_0 and examine its relationship with the total derivative. 1 identifies the only possible direction in $\mathbb{R}$, so the straight line in $\mathbb{R}$ passing through $x_0 \in \mathbb{R}$ and directed as the vector $1 \in \mathbb{R}$ is:

$$\begin{array}{cccc} r_{x_0,1} : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & t & \longmapsto & r_{x_0,1}(t) = x_0 + t. \end{array}$$

The curve $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^m$ admits a directional derivative in $x_0 \in (-\varepsilon, \varepsilon)$ towards the only possible direction defined by $1 \in \mathbb{R}$ if it exists and it is finite the vector of $\mathbb{R}^m$ defined by the limit:

$$D_1\gamma(x_0) = \lim_{t \to 0} \frac{(\gamma \circ r_{x_0,1})(t) - \gamma(x_0)}{t} = \lim_{t \to 0} \frac{\gamma(x_0 + t) - \gamma(x_0)}{t} \equiv \dot{\gamma}(x_0),$$

i.e.

$$D_1\gamma(x_0) - \dot{\gamma}(x_0) = 0 \iff \lim_{t \to 0} \frac{\gamma(x_0 + t) - \gamma(x_0) - \dot{\gamma}(x_0)t}{t} = 0, \tag{B.6}$$

but the limit above is nothing but the 1-dimensional version of eq. (B.1). In eq. (B.1) the role of $\dot{\gamma}(x_0)$ is played by the total derivative $Df(x_0)$ which is an operator belonging to $L(\mathbb{R}^n, \mathbb{R}^m)$, this apparent mismatch can be corrected thanks to the canonical isomorphism (B.5), which allows us identifying $\dot{\gamma}(x_0)$ with the linear operator $D\gamma(x_0)1$.

So, eq. (B.6) becomes

$$D_1\gamma(x_0) - \dot{\gamma}(x_0) = 0 \iff \dot{\gamma}(x_0) = D\gamma(x_0)1,$$

and we come to the conclusion that the directional derivative of γ in x_0 along the direction $1 \in \mathbb{R}$ exists if and only if the total derivative $D\gamma(x_0)$ exists and, moreover, $D_1\gamma(x_0)$ is nothing but the application of the total derivative $D\gamma(x_0)$ to 1, as represented by the suggestive equation:

$$\boxed{\dot{\gamma}(x_0) = D_1\gamma(x_0) = D\gamma(x_0)1}\ , \tag{B.7}$$

in which 1 plays two different roles: in the expression $D_1\gamma(x_0)$ it must be interpreted as a vector defining the only possible direction of derivation in $\mathbb{R}$, while in the expression $D\gamma(x_0)1$ it must be though as the only canonical basis element of the vector space $\mathbb{R}$.

The extension of this result to a function $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is almost straightforward if we build, analogously to what we have done before, the curve $f \circ r_{x_0,v}$ in $\mathbb{R}^m$ passing through $f(x_0)$ by composing f with the straight line $r_{x_0,v}(t) = x_0 + tv$, $x_0 \in \Omega$, $v \in \mathbb{R}^n$.

Supposing that the curve $f \circ r_{x_0,v}$ is differentiable in 0, we have[3]:

$$\begin{aligned}
(f \circ r_{x_0,v})^{\cdot}(0) &\underset{(B.7)}{=} D(f \circ r_{x_0,v})(0)1 \\
&\underset{\text{(chain rule)}}{=} Df(r_{x_0,v}(0))Dr_{x_0,v}(0)1 \\
&\underset{(B.7)}{=} Df(x_0)\dot{r}_{x_0,v}(0) \\
&= Df(x_0)\frac{d(x_0 + tv)}{dt}(0) \\
&= Df(x_0)v,
\end{aligned}$$

but $(f \circ r_{x_0,v})^{\cdot}(0)$ is precisely $D_v f(x_0)$ thanks to eq. (B.3), so we have proven that, if f is Fréchet differentiable in x_0, then f is also Gateaux differentiable in x_0 and the directional derivative can be simply obtained by applying the total derivative to the vector v:

$$\boxed{D_v f(x_0) = Df(x_0)v}\ . \tag{B.8}$$

Counter-examples show that the reverse is not true: even if a function has directional derivatives in every direction in a point, it can be not differentiable. Thus, the Fréchet differentiability of a function of multiple real variables is stronger than the Gateaux derivability, whereas for one variable the two concepts collapse due to the canonical isomorphism $L(\mathbb{R}, \mathbb{R}^m) \simeq \mathbb{R}^m$.

Remark: this way of proving the relationship between directional and total derivative for functions of several real variables is neither the easiest, nor the standard one. However, we chose to present it because this way of reasoning is the closest to the one used in differential geometry, as the reader can appreciate starting from chapter 2.

[3] we omit the composition sign between linear operators, as conventional.

There is a last special case to consider, that of a scalar function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. In this case, the linear operator $Df(x_0)$ is an element of $L(\mathbb{R}^n, \mathbb{R}) \equiv (\mathbb{R}^n)^*$, the dual space of $\mathbb{R}^n$, i.e. $Df(x_0)$ is a linear functional on $\mathbb{R}^n$. $\mathbb{R}^n$ and its dual are canonically isomorphic via the correspondence

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\sim} & (\mathbb{R}^n)^* \\ (e_1, \ldots, e_n) & \longmapsto & (\varepsilon^1, \ldots, \varepsilon^n), \end{array}$$

where $\varepsilon^j(e_i) = \delta_i^j$, $i, j = 1, \ldots, n$, is the **dual canonical basis of** $\mathbb{R}^n$. This isomorphism allows us to identify $Df(x_0) \in (\mathbb{R}^n)^*$ with a vector of $\mathbb{R}^n$ called the **gradient of** f **in** x_0 and denoted with $\nabla f(x_0)$.

The representation of $\nabla f(x_0)$ in components, with respect to the canonical basis of $\mathbb{R}^n$, is given by the column vector:

$$\nabla f(x_0) = \left(\frac{\partial f}{\partial x^1}(x_0), \ldots, \frac{\partial f}{\partial x^n}(x_0) \right)^t,$$

i.e. the Jacobian matrix in x_0 of a real-valued function of n real variables collapses to a vector whose components are the partial derivatives of f calculated in x_0.

The directional derivative of f along a vector $v \in \mathbb{R}^n$ can be obtained via the general formula (B.8). In this case, since the Jacobian matrix is simply a vector, its action on v reduces to the scalar product of $\nabla f(x_0)$ with v:

$$D_v f(x_0) = \langle \nabla f(x_0), v \rangle, \tag{B.9}$$

which can be also seen as a particular instance of the finite-dimensional version of the Riesz isomorphism theorem: the action of the linear functional $Df(x_0) \in (\mathbb{R}^n)^*$ of $v \in \mathbb{R}^n$ is the scalar product of the vector of $\mathbb{R}^n$ uniquely associated to $Df(x_0)$, i.e. $\nabla f(x_0)$, and v.

Thanks to the linearity of the limit, $D_v f$ is linear w.r.t f, but we can say more: if we express the vector v as the linear combination $k_1 v_1 + k_2 v_2$, $v_1, v_2 \in \mathbb{R}^n$, $k_1, k_2 \in \mathbb{R} \backslash \{0\}$, then, by the bilinearity of the real scalar product, we get:

$$D_v f(x_0) = \langle \nabla f(x_0), k_1 v_1 + k_2 v_2 \rangle = k_1 \langle \nabla f(x_0), v_1 \rangle + k_2 \langle \nabla f(x_0), v_2 \rangle = k_1 D_{v_1} f(x_0) + k_2 D_{v_2} f(x_0),$$

i.e.

$$\boxed{D_{k_1 v_1 + k_2 v_2} f(x_0) = k_1 D_{v_1} f(x_0) + k_2 D_{v_2} f(x_0)}, \tag{B.10}$$

so the directional derivative $D_v f$ is linear w.r.t both f and v. This property is crucial in chapter **??**.

B.0.1 Noticeable examples of gradients and total derivatives

We show here how to compute the gradients and total derivatives of particularly important functions.

Directional derivatives of the squared Euclidean norm and of the Euclidean scalar product

In the proofs that will follow we will often use the equality

$$\|a + b\|^2 = \|a\|^2 + \|b\|^2 + 2\langle a, b \rangle,$$

which holds for all $a, b \in \mathbb{R}^n$.

Theorem B.0.2 *Let $x, a \in \mathbb{R}^n$, $f(x) = \|x\|^2$ and $g_a(x) = \|x-a\|^2$, then $\forall x \in \mathbb{R}^n$ it holds that:*

- $\nabla f(x) = 2x$
- $\nabla g_a(x) = 2(x-a)$.

This theorem has a clear interpretation: the computation of the gradient of the square Euclidean norm and of its translations is formally identical to that of the first derivative of the square function in $\mathbb{R}$ and its translations.

Proof. By direct computation:

$$\begin{aligned} D_v f(x) &= \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon v) - f(x)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{\|x+\varepsilon v\|^2 - \|x\|^2}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\|x\|^2 + \|\varepsilon v\|^2 + 2\langle x, \varepsilon v\rangle - \|x\|^2}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\varepsilon^2 \|v\|^2 + 2\varepsilon\langle x, v\rangle}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \left(\varepsilon\|v\|^2 + 2\langle x, v\rangle\right) = 2\langle x, v\rangle. \end{aligned}$$

By (B.9), $D_v f(x) = \langle \nabla f(x), v\rangle = 2\langle x, v\rangle$, i.e. $\langle \nabla f(x), v\rangle = \langle 2x, v\rangle$, or $\langle \nabla f(x) - 2x, v\rangle = 0$ for all directions v, but this is possible if and only if $\nabla f(x) - 2x = 0$, i.e. $\nabla f(x) = 2x$.

Analogously,

$$\begin{aligned} D_v g_a(x) &= \lim_{\varepsilon \to 0} \frac{\|x+\varepsilon v - a\|^2 - \|x-a\|^2}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\|(x-a)+\varepsilon v\|^2 - \|x-a\|^2}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\|x-a\|^2 + \|\varepsilon v\|^2 + 2\langle x-a, \varepsilon v\rangle - \|x-a\|^2}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\varepsilon^2\|v\|^2 + \varepsilon\langle 2(x-a), v\rangle}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \left(\varepsilon\|v\|^2 + \langle 2(x-a), v\rangle\right) = \langle 2(x-a), v\rangle. \end{aligned}$$

The same argument used above leads to the equation $\langle \nabla g_a(x) - 2(x-a), v\rangle = 0$ for all directions v, hence $\nabla g_a(x) = 2(x-a)$. □

Theorem B.0.3 *Let $x, a \in \mathbb{R}^n$, $f_a(x) = \langle a, x\rangle$, then $\nabla f_a(x) = a$.*

Interpretation: the computation of the gradient of the Euclidean scalar product function between two vectors in $\mathbb{R}^n$ is formally identical to that of the first derivative of the function in $\mathbb{R}$ given by the product between a a scalar coefficient and a real variable.

Proof. By direct computation:

$$
\begin{aligned}
D_v f_a(x) &= \lim_{\varepsilon\to 0} \frac{\langle a, x+\varepsilon v\rangle - \langle a, x\rangle}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \frac{\langle a, x\rangle + \varepsilon\langle a, v\rangle - \langle a, x\rangle}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \frac{\varepsilon\langle a, v\rangle}{\varepsilon} \\
&= \langle a, v\rangle.
\end{aligned}
$$

So $\langle \nabla f(x) - a, v\rangle = 0$ for all directions v, i.e. $\nabla f(x) = a$. $\square$

Corollary B.0.1 *Let $x, a \in \mathbb{R}^n$, $f_a : \mathbb{R}^n \to \mathbb{R}^n$ given by $f_a(x) = \langle a, x\rangle a$, then*

$$
Jf_a(x) = \begin{pmatrix} a_1a_1 & a_2a_1 & \dots & a_na_1 \\ a_1a_2 & a_2a_2 & \dots & a_na_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_1a_n & a_2a_n & \dots & a_na_n \end{pmatrix} \equiv (a_ia_j)_{1\leqslant i,j\leqslant n}. \tag{B.11}
$$

Proof. It is enough to consider the component functions $(f_a)_j(x) = \langle a, x\rangle a_j$, $j = 1, \dots, n$ and then apply to the previous theorem, obtaining $\nabla(f_a)_j(x) = aa_j$. Since the rows of $Jf_a(x)$ are $\nabla(f_a)_j(x)$, we get the result. $\square$

Theorem B.0.4 *Let $x, a \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x) = \frac{1}{\|x\|^2}$ and $g : \mathbb{R}^n \to \mathbb{R}^n$ given by $g(x) = \frac{x}{\|x\|^2}$, then $\forall x \in \mathbb{R}^n$ it holds that:*

- $\nabla f(x) = -\frac{2x}{\|x\|^4}$
- $J_g(x) = \frac{1}{\|x\|^2}\left(I_n - 2\frac{(x_ix_j)_{1\leqslant i,j\leqslant n}}{\|x\|^2}\right)$,

where $(x_ix_j)_{1\leqslant i,j\leqslant n}$ is the matrix given by

$$
(x_ix_j)_{1\leqslant i,j\leqslant n} = \begin{pmatrix} x_1x_1 & x_2x_1 & \dots & x_nx_1 \\ x_1x_2 & x_2x_2 & \dots & x_nx_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1x_n & x_2x_n & \dots & x_nx_n \end{pmatrix}.
$$

So, even when $\|x\|^2$ appears at the denominator of a fraction we can compute the gradient or the Jacobian matrix by considering $\|x\|^2$ as a real variable and using the derivation rules.

Proof. By direct computation:

$$
\begin{aligned}
D_v f(x) &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[\frac{1}{\|x+\varepsilon v\|^2} - \frac{1}{\|x\|^2}\right] = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[\frac{1}{\|x\|^2 + 2\varepsilon\langle x, v\rangle + \varepsilon^2\|v\|^2} - \frac{1}{\|x\|^2}\right] \\
&= \frac{1}{\|x\|^2} \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[\frac{1}{1 + \varepsilon\left\langle \frac{2x}{\|x\|^2}, v\right\rangle + \varepsilon^2\frac{\|v\|^2}{\|x\|^2}} - 1\right],
\end{aligned}
$$

recalling the Taylor expansion $\frac{1}{1+\xi} \underset{\xi\to 0}{=} 1-\xi+\mathcal{O}(\xi^2)$ we have that

$$\frac{1}{1+\varepsilon\left\langle\frac{2x}{\|x\|^2},v\right\rangle+\varepsilon^2\frac{\|v\|^2}{\|x\|^2}}-1 \underset{\varepsilon\to 0}{\sim} \cancel{1}-\varepsilon\left\langle\frac{2x}{\|x\|^2},v\right\rangle-\varepsilon^2\frac{\|v\|^2}{\|x\|^2}-\cancel{1}=-\varepsilon\left\langle\frac{2x}{\|x\|^2},v\right\rangle-\varepsilon^2\frac{\|v\|^2}{\|x\|^2},$$

so that

$$D_v f(x)=\frac{1}{\|x\|^2}\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[-\varepsilon\left\langle\frac{2x}{\|x\|^2},v\right\rangle-\varepsilon^2\frac{\|v\|^2}{\|x\|^2}\right]=\left\langle-\frac{2x}{\|x\|^4},v\right\rangle$$

By the same argument used in the proof of the previous theorems we get $\nabla f(x)=-\frac{2x}{\|x\|^4}$.

The formula for $J_g(x)$ follows immediately from that of $\nabla f(x)$ and the Leibnitz property of the directional derivative applied to the component functions $g_j(x)=x_j\frac{1}{\|x\|^2}$, $j=1,\dots,n$, of $g(x)$. □

Theorem B.0.5 *Let $x\in\mathbb{R}^n$, $b\in\mathbb{R}^m$, $A\in M(m\times n,\mathbb{R})$ and $f_{A,b}(x)=\frac{1}{2}\|Ax-b\|^2$, then $\nabla f_{A,b}(x)=A^t(Ax-b)$.*

Proof. Let us compute $f_{A,b}(x+\varepsilon v)$:

$$\begin{aligned}f_{A,b}(x+\varepsilon v)&=\frac{1}{2}\|A(x+\varepsilon v)-b\|^2=\frac{1}{2}\|(Ax-b)+\varepsilon Av\|^2\\&=\frac{1}{2}\left(\|Ax-b\|^2+\varepsilon^2\|Av\|^2+2\varepsilon\langle Ax-b,Av\rangle\right).\end{aligned}$$

Then:

$$\begin{aligned}D_v f_{A,b}(x)&=\lim_{\varepsilon\to 0}\frac{\|Ax-b\|^2+\varepsilon^2\|Av\|^2+2\varepsilon\langle Ax-b,Av\rangle-\|Ax-b\|^2}{2\varepsilon}\\&=\lim_{\varepsilon\to 0}\frac{\varepsilon^2\|Av\|^2+2\varepsilon\langle Ax-b,Av\rangle}{2\varepsilon}\\&=\lim_{\varepsilon\to 0}\left(\frac{\varepsilon\|Av\|^2}{2}+\langle Ax-b,Av\rangle\right)=\langle Ax-b,Av\rangle\\&=\langle A^t(Ax-b),v\rangle.\end{aligned}$$

So $\langle\nabla f_{A,b}(x)-A^t(Ax-b),v\rangle=0$ for all directions u, i.e. $\nabla f_{A,b}(x)=A^t(Ax-b)$. □

The total derivative of the determinant

Finally, we show how to compute the total derivative of the determinant in some special cases. First of all we notice that $\det: M(n,\mathbb{R})\to\mathbb{R}$, so for all $M\in M(n,\mathbb{R})$, $D\det(M)\in L(M(n,\mathbb{R}),\mathbb{R})\cong(M(n,\mathbb{R}))^*\cong(\mathbb{R}^{n^2})^*$, i.e. $D\det(M)$ is a linear functional on $M(n,\mathbb{R})$, so, when it is applied to a matrix of $M(n,\mathbb{R})$, it gives back a real number.

First of all, let us compute $D\det(I_n)$, I_n being the identity matrix $n\times n$. For all $h\in\mathbb{R}$, $h\to 0$ and for all matrix $M\in M(n,\mathbb{R})$, I_n+hA is an infinitesimal perturbation of I_n, thus, if $D\det(I_n)$ exists, it satisfies the equation:

$$\det(I_n+hM)=\det(I_n)+D\det(I_n)hM+\rho_{I_n}(hM), \tag{B.12}$$

where $\rho_{I_n}: M(n,\mathbb{R}) \to \mathbb{R}$ is such that $\frac{\rho_{I_n}(hM)}{h} \underset{h\to 0}{\to} 0$, i.e. $\rho_{I_n}(hM) = o(h)$. In order to make eq. (B.12) explicit, we recall that the dominant coefficient of the characteristic polynomial $p(t) = \det(M - tI_n)$, $t \in \mathbb{R}$, of a generic matrix $M \in M(n,\mathbb{R})$ is $(-1)^n$. Thus, thanks to the fundamental theorem of algebra we can write:

$$\det(M - tI_n) = (-1)^n \prod_{i=1}^{n} (t - \lambda_i), \qquad \lambda_i \in \mathbb{C},$$

or

$$(-1)^n \det(M - tI_n) = \prod_{i=1}^{n} (t - \lambda_i),$$

where λ_i are the complex eigenvalues of M. Using the property $\det(cM) = c^n \det(M)$ for all $c \in \mathbb{R}$ and for all $n \times n$ matrix M, we get:

$$\det(-M + tI_n) = \prod_{i=1}^{n} (t - \lambda_i).$$

If we now operate the change of variable defined by $t = -\frac{1}{h}$, $h \neq 0$, the previous equation can be written in terms of h as follows:

$$\det(-M - \frac{1}{h} I_n) = \prod_{i=1}^{n} (-\frac{1}{h} - \lambda_i) = \prod_{i=1}^{n} -\frac{1}{h}(1 + h\lambda_i) = \left(-\frac{1}{h}\right)^n \prod_{i=1}^{n} (1 + h\lambda_i),$$

i.e., since $\left(-\frac{1}{h}\right)^n = \frac{(-1)^n}{h^n}$,

$$(-1)^n h^n \det(-M - \frac{1}{h} I_n) = \prod_{i=1}^{n} (1 + h\lambda_i),$$

using again the property $\det(cM) = c^n \det(M)$ we get:

$$\det(I_n + hM) = \prod_{i=1}^{n} (1 + h\lambda_i).$$

We can expand the right-hand side of the previous equation by performing the multiplications, which give 1 plus a linear term in h plus terms containing h^j, $j = 2, \ldots, n$ that can be rearranged into $o(h)$ as follows[4]:

$$\prod_{i=1}^{n} (1 + h\lambda_i) = 1 + h \sum_{i=1}^{n} \lambda_i + o(h) = \det(I_n) + h\mathrm{Tr}(M) + o(h),$$

thus

$$\det(I_n + hM) = \det(I_n) + h\mathrm{Tr}(M) + o(h), \tag{B.13}$$

which, by comparison with eq. (B.12), gives

$$\boxed{D\det(I_n)M = \mathrm{Tr}(M)}\,, \qquad \forall M \in M(n,\mathbb{R}),$$

[4]Recall that the trace of a square matrix is the sum of its (possible complex) eigenvalues even if the matrix is real as a consequence of Jordan's canonical form.

i.e. $D\det(I_n)$, **the total derivative of the determinant computed in the identity matrix is the trace**:

$$\boxed{D\det(I_n) = \mathrm{Tr}}\ .$$

Thanks to this result, we can compute $D\det(A)$, for a generic $A \in GL(n,\mathbb{R})$. Let $M \in M(n,\mathbb{R})$, then, since A is invertible, we can write, for all $h \in \mathbb{R}$, $h \to 0$:

$$\begin{aligned}
\det(A + hM) &= \det(A(I_n + A^{-1}hM)) \\
&= \det(A)\det(I_n + A^{-1}hM) \\
&\quad \text{by using } (B.13) \text{ we get:} \\
&= \det(A)(1 + h\mathrm{Tr}(A^{-1}M) + o(h)) \\
&= \det(A) + h\mathrm{Tr}(\det(A)A^{-1}M) + o(h),
\end{aligned}$$

thus, by comparison with (B.12), we find:

$$\boxed{D\det(A)M = \det(A)\mathrm{Tr}(A^{-1}M)}\ , \quad \forall A \in GL(n,\mathbb{R}),\ M \in M(n,\mathbb{R}),$$

and so, in particular,

$$\boxed{D\det(A)M = \mathrm{Tr}(A^{-1}M)}\ , \quad \forall A \in SL(n,\mathbb{R}),\ M \in M(n,\mathbb{R}).$$

Since $\mathrm{Tr}(A^{-1}A) = \mathrm{Tr}(I_n) = n$, if we apply the total derivative to the matrix A itself we get:

$$D\det(A)A = n\det(A), \qquad \forall A \in GL(n,\mathbb{R}),$$

and

$$D\det(A)A = n, \qquad \forall A \in SL(n,\mathbb{R}). \tag{B.14}$$

B.1 The classes of functions $\mathscr{C}^1, \ldots, \mathscr{C}^k, \ldots, \mathscr{C}^\infty$

A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ belongs to the class $\mathscr{C}^0(\Omega)$ if is it continuous in every point of Ω.

The notion of continuous differentiability is more complicated than in the case of functions of only one variable. Let us start with the continuous differentiability.

Def. B.1.1 ($\mathscr{C}^1$-differentiability) *A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is said to belong to the class $\mathscr{C}^1(\Omega)$ if it is Fréchet differentiable for any point $x_0 \in \Omega$ and if the Fréchet derivative function, i.e. the map that associates to each point of Ω the Fréchet derivative of f in it:*

$$\begin{array}{rccl} Df: & \Omega \subseteq \mathbb{R}^n & \longrightarrow & L(\mathbb{R}^n,\mathbb{R}^m) \cong \mathbb{R}^{nm} \\ & x_0 & \longmapsto & Df(x_0), \end{array}$$

is continuous.

Suppose $f \in \mathscr{C}^1(\Omega)$, then it is possible to introduce the following continuous linear functional on $\mathscr{C}^1(\Omega)$ that has a great importance in differential geometry:

$$\boxed{\left.\frac{\partial}{\partial x^i}\right|_{x_0} := ev_{x_0} \circ \frac{\partial}{\partial x^i}} \in \mathscr{C}^1(\Omega)^*, \tag{B.15}$$

where ev_{x_0} is the evaluation map in x_0, so:

$$\begin{array}{rccl} \frac{\partial}{\partial x^i}\Big|_{x_0}: & \mathscr{C}^1(\Omega) & \longrightarrow & \mathbb{R} \\ & f & \longmapsto & \frac{\partial}{\partial x^i}\Big|_{x_0}(f) = \left(ev_{x_0} \circ \frac{\partial}{\partial x^i}\right)(f) = \frac{\partial f}{\partial x^i}(x_0), \end{array}$$

so, *first* we compute the partial derivative of f w.r.t x^i and *then* we evaluate the resulting function in x_0.

Suppose that $f \in \mathscr{C}^1(\Omega)$, then we can ask ourselves if the Fréchet derivative function is Fréchet differentiable in a point $x_0 \in \Omega$. If this is the case, then we say that f is two-times Fréchet differentiable in x_0 and we denote its second Fréchet derivative in x_0 as $D^2 f(x_0)$.

Of course $D^2 f(x_0)$ will still be a linear operator, but this time it will belong to the vector space $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ because the domain of Df is still $\Omega \subseteq \mathbb{R}^n$, but its range is the vector space of linear operators from $\mathbb{R}^n$ to $\mathbb{R}^m$, i.e. $L(\mathbb{R}^n, \mathbb{R}^m)$.

A useful result of linear algebra allows us to naturally identify $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \cong \mathbb{R}^{n^2 m}$ with the vector space $\mathrm{Bil}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m)$ of bilinear maps from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^m$:

$$\begin{array}{rccl} \phi: & L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) & \xrightarrow{\sim} & \mathrm{Bil}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m) \\ & T & \longleftrightarrow & \phi_T, \quad \phi_T(x, y) := (Tx)y, \quad \forall x, y \in \mathbb{R}^n, \end{array}$$

perfectly well-defined: $T \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$, so T acts linearly on $x \in \mathbb{R}^n$ to get $Tx \in L(\mathbb{R}^n, \mathbb{R}^m)$, which acts linerly on y to get $(Tx)y \in \mathbb{R}^m$. The naturalness of the isomorphism comes from the fact that no other structure than the very nature of the elements of the spaces involved in the definition is used.

These considerations justify the following definition.

Def. B.1.2 ($\mathscr{C}^2$-differentiability) *If $D^2 f(x_0)$ exists for every $x_0 \in \Omega$, then*

$$\begin{array}{rccl} D^2 f: & \Omega \subseteq \mathbb{R}^n & \longrightarrow & \mathrm{Bil}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m) \\ & x_0 & \longmapsto & D^2 f(x_0), \end{array}$$

is called second total derivative function of f.

f is said to belong to the class $\mathscr{C}^2(\Omega)$ if the function $D^2 f$ exists and it is continuous in every point of Ω.

Theorem B.1.1 (Schwarz's theorem) *If $f \in \mathscr{C}^2(\Omega)$, then $D^2 f(x_0) \in \mathrm{Bil}_S(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m)$ for all $x_0 \in \Omega$, where Bil_S stays for symmetric bilinear functions.*

Of course, we can iterate the procedure and consider $D^k f(x_0)$, the k-th total derivative of f in x_0, which will be an element of the multilinear maps from k copies of $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$D^k f(x_0) \in \mathrm{Mul}^k(\mathbb{R}^n \times \ldots \times \mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{n^k m},$$

i.e. $D^k f$ transforms linearly each variable of the Cartesian product $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ (k times), taken separately, to an element of $\mathbb{R}^m$.

Def. B.1.3 ($\mathscr{C}^k$-differentiability) *If $D^k f(x_0)$ exists for every $x_0 \in \Omega$, then*

$$\begin{array}{rccl} D^k f: & \Omega \subseteq \mathbb{R}^n & \longrightarrow & \mathrm{Mul}(\mathbb{R}^n \times \cdots \times \mathbb{R}^n, \mathbb{R}^m) \\ & x_0 & \longmapsto & D^k f(x_0), \end{array}$$

is called the k-th total derivative function of f.

f is said to belong to the class $\mathscr{C}^k(\Omega)$ if the function $D^k f$ is continuous on Ω.

Schwarz's theorem implies that, if $f \in \mathscr{C}^k(\Omega)$, then $D^k f(x_0) \in \mathrm{Mul}_S^k(\mathbb{R}^n \times \cdots \times \mathbb{R}^n, \mathbb{R}^m)$ for all $x_0 \in \Omega$, where Mul_S^k stays for *symmetric* multilinear functions.

Def. B.1.4 ($\mathscr{C}^\infty$-differentiability or smoothness) *f is said to belong to the class $\mathscr{C}^\infty(\Omega)$, or simply to be smooth on Ω, if $D^k f$ exists and it is continuous on Ω for all $k \in \mathbb{N}$.*

The continuous linear functional $\left.\frac{\partial}{\partial x^i}\right|_{x_0} := ev_{x_0} \circ \frac{\partial}{\partial x^i} \in \mathscr{C}^\infty(\Omega)^*$ plays a crucial role in differential geometry.

Appendix C

Recap of projective geometry

Appendix D

Recap of group theory

Index

Bibliography

[1] Marco Abate and Francesca Tovena. *Geometria differenziale.* Springer Science & Business Media, 2011.

[2] Giusto Bellavitis. Teoria delle figure inverse, e loro uso nella geometria elementare. *Annali delle Scienze del Regno Lombardo-Veneto, Opera Periodica*, 6:126–141, 1836.

[3] J. Callahan. *The geometry of spacetime: an introduction to special and general relativity.* Springer Science & Business Media, 2000.

[4] Charles Ehresmann. Sur les espaces fibrés associés à une variété differentiable. *CR Acad. Sci. Paris*, 216:628–630, 1943.

[5] J. Faraut and A. Koranyi. *Analysis on Symmetric Cones.* Clarendon Press, Oxford, 1994.

[6] C.J. Isham. *Modern differential geometry for physicists*, volume 61. World Scientific Publishing Company, 1999.

[7] Klaus Jänich and Leslie Kay. *Vector analysis*, volume 5. Springer, 2001.

[8] L. D. Landau and E. M. Lifshitz. *The classical theory of fields.* Pergamon, 1971.

[9] Kurt Lechner. *Classical Electrodynamics.* Springer, 2018.

[10] J. M. Lee. *Introduction to Smooth Manifolds.* Springer, 2013.

[11] P. Lounesto. *Clifford algebras and spinors.* Cambridge University Press, 1997.

[12] N David Mermin. Relativity without light. *American Journal of Physics*, 52(2):119–124, 1984.

[13] Valter Moretti. The interplay of the polar decomposition theorem and the lorentz group. *arXiv preprint math-ph/0211047*, 2002.

[14] Gregory L Naber. *The geometry of Minkowski spacetime: An introduction to the mathematics of the special theory of relativity*, volume 92. Springer Science & Business Media, 2012.

[15] B. O'neill. *Semi-Riemannian geometry with applications to relativity.* Academic press, 1983.

[16] M Özdemir and Abdullah A Ergin. Rotations with unit timelike quaternions in minkowski 3-space. *Journal of geometry and physics*, 56(2):322–336, 2006.

[17] Henri Poincaré. *Leçons sur la théorie mathématique de la lumière*, volume 1. G. Carré, 1889.

[18] J.G. Ratcliffe. *Foundations of hyperbolic manifolds*, volume 149. Springer.

[19] Gregorio Ricci-Curbastro and Tullio Levi-Civita. Méthodes de calcul différentiel absolu et leurs applications. *Mathematische Annalen*, 54(1-2):125–201, 1900.

[20] B. Riemann. Über die hypothesen, welche der geometric zu grunde liegen. In *The Collected Works of Bernhard Riemann*. Dover Books on Mathematics, New York, 2017.

[21] Roman U Sexl and Helmuth K Urbantke. *Relativity, groups, particles: special relativity and relativistic symmetry in field and particle physics*. Springer Science & Business Media, 2012.

[22] M Spivak. Differential geometry, volumes one and two. *Brandeis university*, 1970.

[23] Roberto Torretti. *Relativity and geometry*. Courier Corporation, 1996.

[24] Helmuth K Urbantke. Elementary proof of moretti's polar decomposition theorem for lorentz transformations. *arXiv preprint math-ph/0211077*, 2002.

[25] Hassler Whitney. The imbedding of manifolds in families of analytic manifolds. *Annals of Mathematics*, pages 865–878, 1936.

[26] Ernst Witt. Theorie der quadratischen Formen in beliebigen Körpern. 1937.