Introduction to differential and Riemannian geometry

Stefan Sommer, Tom Fletcher, Xavier Pennec

Introduction to differential and Riemannian geometry

Stefan Sommer,∗, Tom Fletcher∗∗ and Xavier Pennec†

∗ University of Copenhagen, Department of Computer Science, Copenhagen, Denmark.
∗∗ University of Virginia, Dept. of Electrical & Computer Engineering and Computer Science, Charlottesville, VA USA
† Université Côte d'Azur and Inria, Epione team, 2004 Route des Lucioles, BP 93, F-06902 Sophia Antipolis Cedex, France

Abstract

This chapter introduces the basic concepts of differential geometry: Manifolds, charts, curves, their derivatives, and tangent spaces. The addition of a Riemannian metric enables length and angle measurements on tangent spaces giving rise to the notions of curve length, geodesics, and thereby the basic constructs for statistical analysis of manifold valued data. Lie groups appear when the manifold in addition has smooth group structure, and homogeneous spaces arise as quotients of Lie groups. We discuss invariant metrics on Lie groups and their geodesics.

The goal is to establish the mathematical bases that will allow in the sequel to build a simple but consistent statistical computing framework on manifolds. In the later part of the chapter, we describe computational tools, the Exp and Log maps, derived from the Riemannian metric. The implementation of these atomic tools will then constitute the basis to build more complex generic algorithms in the following chapters.

1.1. Introduction

When data exhibit non-linearity, the mathematical description of the data space must often depart from the convenient linear structure of Euclidean vector spaces. Non-linearity prevents global vector space structure but we can nevertheless ask which mathematical properties from the Euclidean case can be kept while still preserving the accurate modelling of the data. It turns out that in many cases, local resemblance to a Euclidean vector space is one such property. In other words, up to some second order approximation, the data space can be linearized in limited regions while forcing a linear model on the entire space would introduce too much distortion.

The concept of local similarity to Euclidean spaces brings us exactly to the setting of manifolds. Topological, differential, and Riemannian manifolds are characterized by the existence of local maps, charts, between the manifold and a Euclidean space. These charts are structure preserving: They are homeomorphisms in the case of topological manifolds, diffeomorphisms in the case of differential manifolds, and, in the case of Riemannian manifolds, they carry local inner products that encode the non-Euclidean geometry.

The following sections describe these foundational concepts and how they lead to notions commonly associated with geometry: curves, length, distances, geodesics, curvature, parallel transport, volume form. In addition to the differential and Riemannian structure, we describe one extra layer of structure: Lie groups that are manifolds equipped with smooth group structure. Lie groups and their quotients are examples of homogeneous spaces. The group structure provides relations between distant points on the group and thereby additional ways of constructing Riemannian metrics and deriving geodesic equations.

Topological, differential, and Riemannian manifolds are often covered by separate graduate courses in mathematics. In this much briefer overview, we describe the general concepts, often sacrificing mathematical rigor to instead provide intuitive reasons for the mathematical definitions. For a more in-depth introduction to geometry, the interested reader may for example refer to the sequence of books by John M. Lee on topological, differentiable, and Riemannian manifolds [Lee00, Lee03, Lee97] or to the book on Riemannian geometry by do Carmo [dC92]. More advanced references include [KSM93], [Jos11] and [Pos10].

1.2. Manifolds

A manifold is a collection of points that locally, but not globally, resembles Euclidean space. When the Euclidean space is of finite dimension, we can without loss of generality relate it to $\mathbb{R}^d$ for some $d > 0$. The abstract mathematical definition of a manifold specifies the topological, differential, and geometric structure by using charts, maps between parts of the manifold and $\mathbb{R}^d$, and collections of charts denoted atlases. We will discuss this construction shortly, however, we first focus on the case when the manifold is a subset of a larger Euclidean space. This viewpoint is often less abstract and closer to our natural intuition of a surface embedded in our surrounding 3D Euclidean space.

Let us exemplify this by the surface of the earth embedded in $\mathbb{R}^3$. We are constrained by gravity to live on the surface of the earth. This surface seems locally flat with two dimensions only, and we use two dimensional maps to navigate the surface. When traveling far, we sometimes need to change from one map to another. We then find charts that overlap in small parts, and we assume that the charts provide roughly the same view of the surface in those overlapping parts. For a long time, the earth was even considered to be flat because its curvature was not noticeable at the scale at which it was observed. When considering the earth surface as a two dimensional restriction of the 3D ambient space, the surface is an embedded submanifold of $\mathbb{R}^3$. On the other hand, when using maps and piecing the global surface together using the compatibility of the overlapping parts, we take the abstract view using charts and atlases.

1.2.1. Embedded Submanifolds

Arguably the simplest example of a 2-dimensional manifold is the sphere $\mathbb{S}^2$. Relating to the previous example, when embedded in $\mathbb{R}^3$, we can view it as an idealized model for the surface of the earth. The sphere with radius 1 can be described as the set of unit vectors in $\mathbb{R}^3$, i.e. the set

$$\mathbb{S}^2 = \{(x^1, x^2, x^3) \in \mathbb{R}^3 \mid (x^1)^2 + (x^2)^2 + (x^3)^2 = 1\}. \tag{1.1}$$

Notice from the definition of the set that all points of $\mathbb{S}^2$ satisfy the equation $(x^1)^2 + (x^2)^2 + (x^3)^2 - 1 = 0$. We can generalize this way of constructing a manifold to the following definition:

Definition 1 (Embedded manifold)**.** Let $F : \mathbb{R}^k \to \mathbb{R}^m$ be a differentiable map such that the Jacobian matrix $dF(x) = (\frac{\partial}{\partial x^j} F^i(x))^i_j$ has constant rank $k - d$ for all $x \in F^{-1}(0)$. Then the zero-level set $\mathcal{M} = F^{-1}(0)$ is an embedded manifold of dimension d.
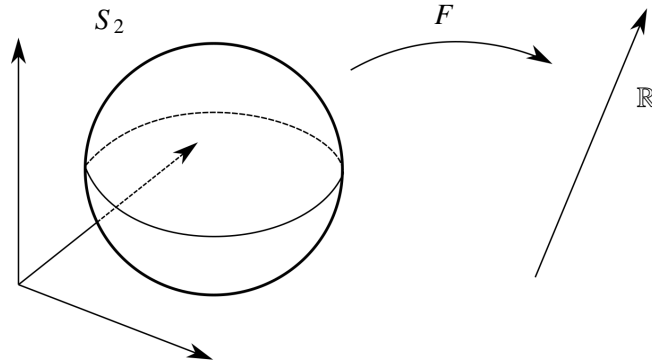

Figure 1.1 An embedded manifold arises as the zero level subset $\mathcal{M} = F^{-1}(0)$ of the map $F : \mathbb{R}^k \to \mathbb{R}^m$. Here, $F : \mathbb{R}^3 \to \mathbb{R}$ is given by the sphere equation $x \mapsto (x^1)^2 + (x^2)^2 + (x^3)^2 - 1$, and the manifold $\mathcal{M} = \mathbb{S}^2$ is of dimension $3 - 1 = 2$.

The map F is said to give an implicit representation of the manifold. In the above example, we used the definition with $F(x) = (x^1)^2 + (x^2)^2 + (x^3)^2 - 1$.

The fact that $\mathcal{M} = F^{-1}(0)$ is a manifold is often taken as the consequence of the submersion level set theorem instead of a definition. The theorem states that with the above assumptions, $\mathcal{M}$ has a manifold structure as constructed with charts and atlases. In addition, the topological and differentiable structure of M is in a certain way compatible with that of $\mathbb{R}^k$ letting us denote M as *embedded* in $\mathbb{R}^k$. For now, we will be somewhat relaxed about the details and use the construction as a working

definition of what we think of as a manifold.

The map F can be seen as a set of m constraints that points in $\mathcal{M}$ must satisfy. The Jacobian matrix $dF(x)$ at a point in $x \in \mathcal{M}$ linearizes the constraints around x and its rank $k - d$ indicates how many of them are linearly independent. In addition to the unit length constraints of vectors in $\mathbb{R}^3$ defining $\mathbb{S}^2$, additional examples of commonly occurring manifolds that we will see in this book arise directly from embedded manifolds or as quotients of embedded manifolds.

Example 1. *d-dimensional spheres* $\mathbb{S}^d$ embedded in $\mathbb{R}^{d+1}$. Here we express the unit length equation generalizing (1.1) by

$$\mathbb{S}^d = \{x \in \mathbb{R}^{n+1} \mid \|x\|^2 - 1 = 0\} . \tag{1.2}$$

The squared norm $\|x\|^2$ is the standard squared Euclidean norm on $\mathbb{R}^{d+1}$.

Example 2. *Orthogonal matrices* $O(k)$ on $\mathbb{R}^k$ have the property that the inner products $\langle U_i, U_j \rangle$ of columns U_i, U_j of the matrix $U \in M_{(k,k)}$ vanish for $i \neq j$ and equal 1 for $i = j$. This gives k^2 constraints, and $O(k)$ is thus an embedded manifold in $M_{(k,k)}$ by the equation

$$O(k) = \left\{ U \in M_{(k,k)} \mid UU^\top - \mathrm{Id}_k = 0 \right\} \tag{1.3}$$

with Id_k being the identity matrix on $\mathbb{R}^k$ We will see in Section 1.7.3 that the rank of the map $F(U) = UU^\top - \mathrm{Id}_k$ is $\frac{k(k+1)}{2}$ on $O(k)$, and it follows that $O(k)$ has dimension $\frac{k(k-1)}{2}$.

1.2.2. Charts and Local Euclideaness

We now describe how charts, local parametrizations of the manifold, can be constructed from the implicit representation above. We will use this to give the more abstract definition of a differentiable manifold.

When navigating the surface of the earth, we seldom use curved representations of the surface but instead rely on charts that give a flat, 2D representation of regions limited in extent. It turns out that this analogy can be extended to embedded manifolds with a rigorous mathematical formulation.

Definition 2. A chart on the d-dimensional manifold $\mathcal{M}$ is a diffeomorphic mapping $\phi : U \to \tilde{U}$ from an open set $U \subset \mathcal{M}$ to an open set $\tilde{U} \subseteq \mathbb{R}^d$.

The definition exactly captures the informal idea of representing a local part of the surface, the open set U, with a mapping to a Euclidean space, in the surface case $\mathbb{R}^2$.

When using charts, we often say that we work *in coordinates*. Instead of accessing
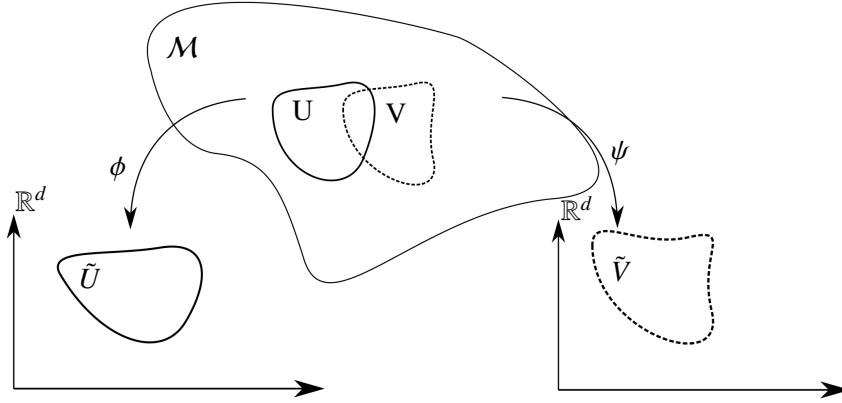
Figure 1.2 Charts $\phi : U \to \tilde{U}$ and $\psi : V \to \tilde{V}$, members of the atlas covering the manifold $\mathcal{M}$, from the open sets $U, V \subset \mathcal{M}$ to open sets $\tilde{U}, \tilde{V}$ of $\mathbb{R}^d$, respectively. The compatibility condition ensures that ϕ and ψ agree on the overlap $U \cap V$ between U and V in the sense that the composition $\psi \circ \phi^{-1}$ is a differentiable map.

points on $\mathcal{M}$ directly, we take a chart $\phi : U \to \tilde{U}$ and use points in $\phi(U) \subset \mathbb{R}^d$ instead. This gives us the convenience of having a coordinate system present. However, we need to be aware that the choice of the coordinate system affects the analysis, both theoretically and computationally. When we say that we work in coordinates $x = (x^1, \ldots, x^d)$, we implicitly imply that there is a chart ϕ such that $\phi^{-1}(x)$ is a point on $\mathcal{M}$.

It is a consequence of the implicit function theorem that embedded manifolds have charts. Proving it takes some work but we can sketch the idea in the case of the implicit representation map $F : \mathbb{R}^k \to \mathbb{R}^m$ having Jacobian with full rank m. Recall the setting of the implicit function theorem (see e.g. [Lee03]): Let $F : \mathbb{R}^{d+m} \to \mathbb{R}^m$ be continuously differentiable and write $(x, y) \in \mathbb{R}^{d+m}$ such that x denotes the first d coordinates and y the last m coordinates. Let $d_y F$ denote the last m columns of the Jacobian matrix dF, i.e. derivatives of F taken with respect to variations in y. If $d_y F$ has full rank m at a point (x, y) where $F(x, y) = 0$, then there exists an open neighborhood $\tilde{U} \subseteq \mathbb{R}^d$ of x and a differentiable map $g : \tilde{U} \to \mathbb{R}^m$ such that $F(x, g(x)) = 0$ for all $x \in \tilde{U}$.

The only obstruction to using the implicit function theorem directly to find charts is that we may need to rotate the coordinates on $\mathbb{R}^{d+m}$ in order to find coordinates (x, y) and a submatrix $d_y F$ with full rank. With this in mind, the map g ensures that $F(x, g(x)) = 0$ for all $x \in \tilde{U}$, i.e. the points $(x, g(x)), x \in \tilde{U}$ are included in $\mathcal{M}$. Setting $U = g(\tilde{U})$, we get a chart $\phi : U \to \tilde{U}$ directly by the mapping $(x, g(x)) \mapsto x$.

1.2.3. Abstract Manifolds and Atlases

We now use the concept of charts to define atlases as collections of charts and from this the abstract notion of a manifold.

Definition 3 (Atlas). An atlas of a differentiable set $\mathcal{M}$ is a family of charts $(\phi_i)_{i=1,\ldots,N}$ with $\phi_i : U_i \rightarrow \tilde{U}_i$ such that
- ϕ_i cover $\mathcal{M}$: For each $x \in \mathcal{M}$, there exists $i \in \{1, \ldots, N\}$ such that $x \in U_i$,
- ϕ_i are compatible: For each pair $i, j \in \{1, \ldots, N\}$ where $U_i \cap U_j$ is non-empty, the composition $\phi_i \circ \phi_j^{-1} : \phi_j(U_i \cap U_j) \rightarrow \mathbb{R}^d$ is a differentiable map.

An atlas thus ensures the existence of at least one chart covering a neighborhood of each point of $\mathcal{M}$. This allows the topological and differential structure of $\mathcal{M}$ to be given by definition from the topology and differential structure of the image of the charts, i.e. $\mathbb{R}^d$. Intuitively, the structure coming from the Euclidean spaces $\mathbb{R}^d$ is pulled back using ϕ_i to the manifold. In order for this construction to work, we must ensure that there is no ambiguity in the structure we get if the domain of multiple charts cover a given point. The compatibility condition ensures exactly that.

Definition 4 (Manifold). Let $\mathcal{M}$ be a set with an atlas $(\phi_i)_{i=1,\ldots,N}$ with $\phi_i : U_i \rightarrow \tilde{U}_i$, $\tilde{U}_i \subseteq \mathbb{R}^d$. Then $\mathcal{M}$ is a manifold of dimension d.

Remark 1. Until now, we have been somewhat loose in describing maps as being "differentiable". The differentiability of maps on a manifold comes from the differential structure which in turn is defined from the atlas and the charts mapping to $\mathbb{R}^d$. The differential structure on $\mathbb{R}^d$ allows derivatives up to any order but the charts may not support this when transferring the structure to $\mathcal{M}$. To be more precise, in the compatibility condition, we require the compositions $\phi_i \circ \phi_j^{-1}$ to be C^r as maps from $\mathbb{R}^d$ to $\mathbb{R}^d$ for some integer r. This gives a differentiable structure on $\mathcal{M}$ of the same order. In particular, when $r \geq 1$, we say that $\mathcal{M}$ is a differentiable manifold, and $\mathcal{M}$ is *smooth* if $r = \infty$. One may also require only $r = 0$ in which case $\mathcal{M}$ is a *topological* manifold with no differentiable structure.

Because of the implicit function theorem, embedded submanifolds in the sense of Definition 1 have charts and atlases. Embedded submanifolds are therefore particular examples of abstract manifolds. In fact, this goes both ways: The Whitney embedding theorem states that any d-dimensional manifold can be embedded in $\mathbb{R}^k$ with $k \leq 2d$ so that the topology is induced by the one of the embedding space. For Riemannian manifolds defined below, this theorem only provides a local C^1 embedding and not a global smooth embedding.

Example 3. *The projective space* P_d is the set of lines through the origin in $\mathbb{R}^{d+1}$. Each such line intersects the sphere $\mathbb{S}^d$ in two points that are antipodal. By identifying such points, expressed by taking the quotient using the equivalence relation $x \sim -x$, we get the representation $P_d \simeq \mathbb{S}^d / \sim$. Depending on the properties of the equivalence relation, the quotient space of a manifold may not be a manifold in general (more details will be given in Chapter 9 of this book). In the case of the projective space, we can verify the above abstract manifold definition. Therefore, the projective space cannot be seen as an embedded manifold directly, but it can be seen as the quotient space of an embedded manifold.

1.2.4. Tangent Vectors and Tangent Space

As the name implies, derivatives lies at the core of differential geometry. The differentiable structure allows taking derivatives of curves in much the same way as the usual derivatives in Euclidean space. However, spaces of tangent vectors to curves behave somewhat differently on manifolds due to the lack of the global reference frame that the Euclidean space coordinate system gives. We here discuss derivatives of curves, tangent vectors, and tangent spaces.
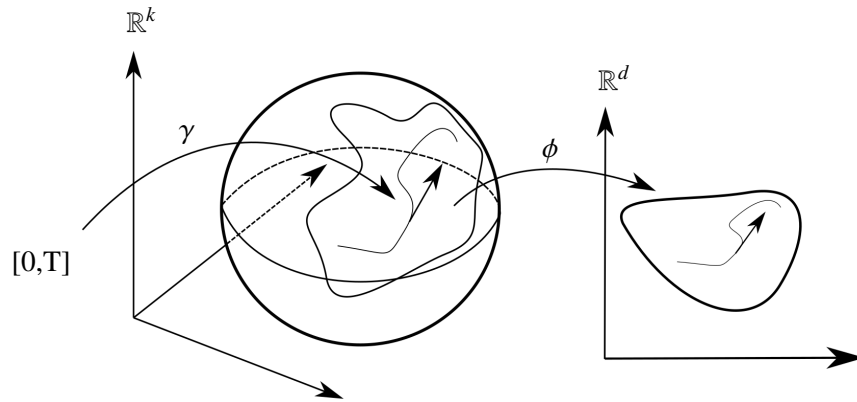
Figure 1.3 The curve γ maps the interval $[0, T]$ to the manifold. Using a chart ϕ, we can work in coordinates with the curve $\phi \circ \gamma$ in $\mathbb{R}^d$. If $\mathcal{M}$ is embedded, γ is in addition a curve in $\mathbb{R}^k$. The derivative $\dot{\gamma}(t)$ is a tangent vector in the linear tangent space $T_{\gamma(t)}\mathcal{M}$. It can be written in coordinates using ϕ as $\dot{\gamma} = \dot{\gamma}^i \partial_{x^i}$. In the embedding space, the tangent space $T_{\gamma(t)}\mathcal{M}$ is the affine d-dimensional subspace $\gamma(t) + \ker dF(\gamma(t))$ of $\mathbb{R}^k$.

Let $\gamma : [0, T] \rightarrow \mathbb{R}^k$ be a differentiable curve in $\mathbb{R}^k$ parametrized on the interval

$[0, T]$. For each t, the curve derivative is

$$\frac{d}{dt}\gamma(t) = \dot{\gamma} = \begin{pmatrix} \frac{d}{dt}\gamma^1(t) \\ \vdots \\ \frac{d}{dt}\gamma^k(t) \end{pmatrix} . \tag{1.4}$$

This tangent or velocity vector can be regarded as a vector in $\mathbb{R}^k$, denoted the tangent vector to γ at t. If $\mathcal{M}$ is an embedded manifold in $\mathbb{R}^k$ and $\gamma(t) \in \mathcal{M}$ for all $t \in [0, T]$, we can regard γ as a curve in $\mathcal{M}$. As illustrated on Figure 1.3, the tangent vectors of γ are also tangential to $\mathcal{M}$ itself. The set of tangent vectors to all curves at $p = \gamma(t)$ span a d-dimensional affine subspace of $\mathbb{R}^k$ that approximates $\mathcal{M}$ to the first order at x. This affine space has an explicit realization as $x + \ker dF(x)$ where $x = \gamma(t)$ is the foot-point and $\ker dF$ denotes the kernel (null-space) of the Jacobian matrix of F. The space is called the tangent space $T_x\mathcal{M}$ of $\mathcal{M}$ at the point x. In the embedded manifold case, tangent vectors thus arise from the standard curve derivative, and tangent spaces are affine subspaces of $\mathbb{R}^k$.

On abstract manifolds, the definition of tangent vectors becomes somewhat more intricate. Let γ be a curve in the abstract manifold $\mathcal{M}$, and consider $t \in [0, T]$. By the covering assumption on the atlas, there exists a chart $\phi : U \to \tilde{U}$ with $\gamma(t) \in U$. By the continuity of γ and openness of U, $\gamma(s) \in U$ for s sufficiently close to t. Now the curve $\tilde{\gamma} = \phi \circ \gamma$ in $\mathbb{R}^d$ is defined for such s. Thus we can take the standard Euclidean derivative $\dot{\tilde{\gamma}}$ of $\tilde{\gamma}$. This gives a vector in $\mathbb{R}^d$. In the same way as we define the differentiable structure on $\mathcal{M}$ by definition to be that inherited from the charts, it would be natural to let a tangent vector of $\mathcal{M}$ be $\dot{\tilde{\gamma}}$ by definition. However, we would like to be able to define tangent vectors independently of the underlying curve. In addition, we need to ensure that the construction does not depend on the chart ϕ.

One approach is to define tangent vectors from their actions on real valued functions on $\mathcal{M}$. Let $f : \mathcal{M} \to \mathbb{R}$ be a differentiable function. $f \circ \gamma$ is then a function from $\mathbb{R}$ to $\mathbb{R}$ whose derivative is

$$\frac{d}{dt}f \circ \gamma(t) . \tag{1.5}$$

This operation is clearly linear in f in the sense that $\frac{d}{dt}((\alpha f + \beta g) \circ \gamma) = \alpha \frac{d}{dt}(f \circ \gamma) + \beta \frac{d}{dt}(g \circ \gamma)$ when g is another differentiable function and $\alpha, \beta \in \mathbb{R}$. In addition, this derivative satisfies the usual product rule for the derivative of the pointwise product $f \cdot g$ of f and g. Operators on differentiable functions satisfying these properties are called *derivations*, and we can define tangent vectors and tangent spaces as the set of derivations. I.e. $v \in T_x\mathcal{M}$ is a tangent vector if it defines a derivation $v(f)$ on functions $f \in C^1(\mathcal{M}, \mathbb{R})$. It can now be checked that the curve derivative using a chart above defines derivations. By the chain rule, is can be seen that these derivations are independent of the chosen chart.

The construction of $T_x\mathcal{M}$ as derivations is rather abstract. In practice, it is often most convenient to just remember that there is an abstract definition and otherwise think of tangent vectors as derivatives of curves. In fact, tangent vectors and tangent spaces can also be defined without derivations using only the derivatives of curves. However, in this case one must define a tangent vector as an equivalence class of curves because multiple curves can result in the same derivative. This construction, although in some sense more intuitive, therefore has its own complexities.

The set $\{T_x\mathcal{M} \,|\, x \in M\}$ has a structure of a differentiable manifold in itself. It is called the tangent bundle $T\mathcal{M}$. It follows that tangent vectors $v \in T_x\mathcal{M}$ for some $x \in \mathcal{M}$ are also elements of $T\mathcal{M}$. $T\mathcal{M}$ is a special case of a fiber bundle (a local product of spaces whose global topology may be more complex). We will later see other examples of fiber bundles, e.g. the cotangent bundle $T^*\mathcal{M}$ and the frame bundle $F\mathcal{M}$.

A local coordinate system $x = (x^1, \ldots x^d)$ coming from a chart induces a basis $\partial_x = (\partial_{x^1}, \ldots \partial_{x^d})$ of the tangent space $T_x\mathcal{M}$. Therefore, any $v \in T_x\mathcal{M}$ can be expressed as a linear combination of $\partial_{x^1}, \ldots \partial_{x^d}$. Writing v^i for the ith entry of such linear combinations, we have $v = \sum_{i=1}^d v^i \partial_{x^i}$.

Remark 2 (Einstein summation convention)**.** We will often use the Einstein summation convention that dictates an implicit sum over indices appearing twice in lower and upper position in expressions, in particular in coordinate expressions and tensor calculations. For example, in the coordinate basis mentioned above, we have $v = v^i \partial_{x^i}$ where the sum $\sum_{i=1}^d$ is implicit because the index i appears in upper position on v^i and lower position on ∂_{x^i}.

Just as a Euclidean vector space V has a dual vector space V^* consisting of linear functionals $\xi : V \to \mathbb{R}$, the tangent spaces $T_x\mathcal{M}$ and tangent bundle $T\mathcal{M}$ have dual spaces, the cotangent spaces $T_x^*\mathcal{M}$ and cotangent bundle $T^*\mathcal{M}$. For each x, elements of the cotangent space $T_x^*\mathcal{M}$ are linear maps from $T_x\mathcal{M}$ to $\mathbb{R}$. The coordinate basis $(\partial_{x^1}, \ldots \partial_{x^d})$ induces a similar coordinate basis $(dx^1, \ldots dx^d)$ for the cotangent space. This basis is defined from evaluation on ∂_{x^i} by $dx^j(\partial_{x^i}) = \delta_j^i$ where the delta-function δ_j^i is 1 if if $i = j$ and 0 otherwise. The coordinates v^i for tangent vectors in the coordinate basis had upper indices above. Similarly, coordinates for cotangent vectors conventionally have lower indices such that $\xi = \xi_i dx^i$ for $\xi \in T_x^*\mathcal{M}$ again using the Einstein summation convention. Elements of $T^*\mathcal{M}$ are called covectors. The evaluation $\xi(v)$ of a covector ξ on a vector v is sometimes written $(\xi|v)$ or $\langle \xi, v \rangle$. Note that the latter notation with brackets is similar to the notation for inner products used below.

1.2.5. Differentials and Pushforward

The interpretation of tangent vectors as derivations above allows taking derivatives of functions. If X is a vector field on $\mathcal{M}$, we can use this pointwise to define a new function on $\mathcal{M}$ by taking derivatives at each point, i.e. $X(f)(x) = X(x)(f)$ using that $X(x)$ is a tangent vector in $T_x\mathcal{M}$ and hence a derivation that acts on functions. If instead f is a map between two manifolds $f : \mathcal{M} \to \mathcal{N}$, we get the differential $df : T\mathcal{M} \to T\mathcal{N}$ as a map between the tangent bundle of $\mathcal{M}$ and $\mathcal{N}$. In coordinates, this is $df(\partial_{x^i})^j = \partial_{x^i} f^j$ with f^j being the jth component of f. The differential df is often denoted the pushforward of f because it uses f to map, i.e. *push*, tangent vectors in $T\mathcal{M}$ to tangent vectors in $T\mathcal{N}$. For this reason, the pushforward notation $f_* = df$ is often used. When f is invertible, there exists a corresponding pullback operation $f^* = df^{-1}$.

As a special case, consider a map f between $\mathcal{M}$ and the manifold $\mathbb{R}$. Then $f_* = df$ is a map from $T\mathcal{M}$ to $T\mathbb{R}$. Because $\mathbb{R}$ is Euclidean, we can identify the tangent bundle with $\mathbb{R}$ itself, and we can consider df a map $T\mathcal{M} \to \mathbb{R}$. Being a derivative, $df|_{T_x\mathcal{M}}$ is linear for each $x \in \mathcal{M}$ and $df(x)$ is therefore a covector in $T_x^*\mathcal{M}$. Though the differential df is also a pushforward, the notation df is most often used because of its interpretation as a covector field.

1.3. Riemannian Manifolds

So far, we defined manifolds as having topological and differential structure, either inherited from $\mathbb{R}^k$ when considering embedded manifolds, or via charts and atlases with the abstract definition of manifolds. We now start including geometric and metric structures.

The topology determines the local structure of a manifold by specifying the open sets and thereby continuity of curves and functions. The differentiable structure allowed us to define tangent vectors and differentiate functions on the manifold. However, we have not yet defined a notion of how "straight" manifold valued curves are. To obtain such a notion, we need to add a geometric structure, called a connection, which allows to compare neighboring tangent spaces and characterizes the parallelism of vectors at different points. Indeed, differentiating a curve on a manifold gives tangent vectors belonging at each point to a different tangent vector space. In order to compute the second order derivative, the acceleration of the curves, we need a way to map the tangent space at a point to the tangent space at any neighboring point. This is the role of a connection $\nabla_X Y$, which specifies how the vector field $Y(x)$ is derived in the direction of the vector field $X(x)$ (Figure 1.4). In the embedding case, tangent spaces are affine spaces of the embedding vector space, and the simplest way to specify this mapping is through an affine transformation, hence the name affine connection introduced by Cartan [Car23]. A connection operator also describes how a vector is
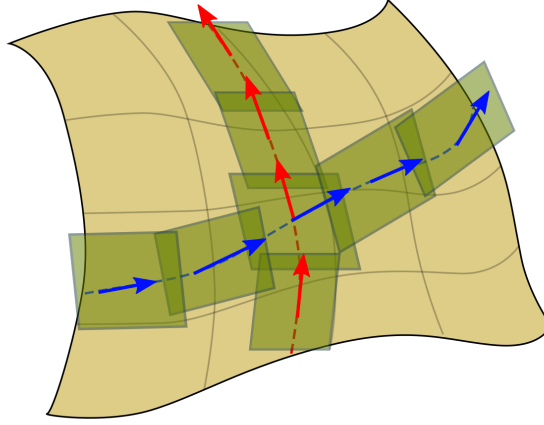
Figure 1.4 Tangent vectors along the red and blue curves drawn on the manifold belong to different tangent spaces. To define the acceleration as the difference of neighboring tangent vectors, we need to specify a mapping to connect a tangent space at one point to the tangent spaces at infinitesimally close points. In the embedding case, tangent spaces are affine spaces of the embedding vector space, and the simplest way to specify this mapping is through an affine transformation.

transported from a tangent space to a neighboring one along a given curve. Integrating this transport along the curve specifies the parallel transport along this curve. However, there is usually no global parallelism as in Euclidean space. As a matter of fact, transporting the same vector along two different curves arriving at the same point in general leads to different vectors at the endpoint. This is easily seen on the sphere where traveling from north pole to the equator, then along the equator for 90 degrees and back to north pole turns any tangent vector by 90 degrees. This defect of global parallelism is the sign of curvature.

By looking for curves that remain locally parallel to themselves, i.e. such that $\nabla_{\dot\gamma}\dot\gamma = 0$, one defines the equivalent of "straight lines" in the manifold: geodesics. One should notice that there exists many different choices of connections on a given manifold which lead to different geodesics. However, geodesics by themselves do not quantify how far away from each other two points are. For that purpose, we need an additional structure: a distance. By restricting to distances which are compatible with the differential structure, we enter into the realm of Riemannian geometry.

1.3.1. Riemannian Metric

A *Riemannian metric* is defined by a smoothly varying collection of scalar products $\langle .,. \rangle_x$ on each tangent space $T_x\mathcal{M}$ at points x of the manifold. For each x, each such
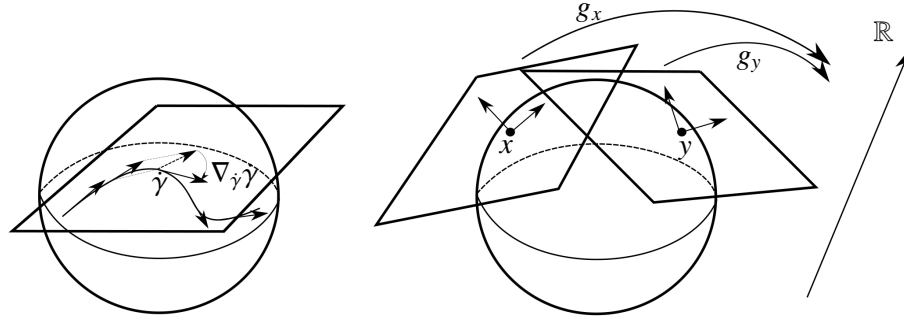
Figure 1.5 (left) Vectors along a curve, here velocity vectors $\dot{\gamma}$ along the curve γ, live in different tangent spaces and therefore cannot be compared directly. A connection ∇ defines a notion of transport of vectors along curves. This allows transport of a vector $\dot{\gamma}(t - \Delta t) \in T_{\gamma(t-\Delta t)}\mathcal{M}$ to $T_{\gamma(t)}\mathcal{M}$, and the acceleration $\nabla_{\dot{\gamma}}\dot{\gamma}$ arises by taking derivatives in $T_{\gamma(t)}\mathcal{M}$. (right) For each point $x \in \mathcal{M}$, the metric g defines a positive, bilinear map $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$. Contrary to the Euclidean case, g depends on the base point, and vectors in the tangent space $T_y\mathcal{M}$ can only be compared by g evaluated at y, i.e. the map $g_y : T_y\mathcal{M} \times T_y\mathcal{M} \to \mathbb{R}$.

scalar product is a positive definite bilinear map $\langle ., . \rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$, see Figure 1.5. The inner product gives a norm $\|\cdot\|_x : T_x\mathcal{M} \to \mathbb{R}$ by $\|v\|^2 = \langle v, v \rangle_x$. In a given chart, we can express the metric by a symmetric positive definite matrix $g(x)$. The ij^{th} entry of the matrix is denoted $g_{ij}(x)$ and given by the dot product of the coordinate basis for the tangent space: $g_{ij}(x) = \langle \partial_{x^i}, \partial_{x^j} \rangle_x$. This matrix is called the *local representation of the Riemannian metric* in the chart x and the dot products of two vectors v and w in $T_x\mathcal{M}$ is now in coordinates $\langle v, w \rangle_x = v^\top g(x)w = v^i g_{ij}(x)w^j$. The components g^{ij} of the inverse $g(x)^{-1}$ of the metric defines a metric on covectors by $\langle \xi, \eta \rangle_x = \xi_i g^{ij} \eta_j$. Notice how the upper indices of g^{ij} fit the lower indices of the covector in the Einstein summation convention. This inner product on $T_x^*\mathcal{M}$ is called a cometric.

1.3.2. Curve Length and Riemannian Distance

If we consider a curve $\gamma(t)$ on the manifold, we can compute at each abscissa t its velocity vector $\dot{\gamma}(t)$ and its norm $\|\dot{\gamma}(t)\|$, the instantaneous speed. For the velocity vector, we only need the differential structure, but for the norm, we need the Riemannian metric at the point $\gamma(t)$. To compute the length of the curve, the norm is integrated along the curve:

$$\mathcal{L}(\gamma) = \int \|\dot{\gamma}(t)\|_{\gamma(t)} \, dt = \int \left(\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} \right)^{\frac{1}{2}} dt \,.$$

The integrals here are over the domain of the curve, e.g. $[0, T]$. We write $\mathcal{L}_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\|_{\gamma(t)}\, dt$ to be explicit about the integration domain. This gives the length of the curve segment $\gamma(a)$ to $\gamma(b)$.

The distance between two points of a connected Riemannian manifold is the minimum length among the curves γ joining these points:

$$\mathrm{dist}(x, y) = \min_{\gamma(0)=x, \gamma(1)=y} \mathcal{L}(\gamma). \tag{1.6}$$

The topology induced by this Riemannian distance is the original topology of the manifold: open geodesic balls constitute a basis of open sets.

The Riemannian metric is the intrinsic way of measuring length on a manifold. The extrinsic way is to consider the manifold as embedded in $\mathbb{R}^k$ and compute the length of a curve in $\mathcal{M}$ as for any curve in $\mathbb{R}^k$. In section 1.2.4, we identified the tangent spaces of an embedded manifold with affine subspaces of $\mathbb{R}^k$. In this case, the Riemannian metric is the restriction of the dot product on $\mathbb{R}^k$ to the tangent space at each point of the manifold. Embedded manifolds thus inherit also their geometric structure in the form of the Riemannian metric from the embedding space.

1.3.3. Geodesics

In Riemannian manifolds, locally length-minimizing curves are called metric geodesics. The next subsection will show that these curves are also auto-parallel for a specific connection, so that they are simply called geodesics in general. A curve is locally length minimizing if for all t and sufficiently small s, $\mathcal{L}_t^{t+s}(\gamma) = \mathrm{dist}(\gamma(t), \gamma(t + s))$. This implies that small segments of the curve realize the Riemannian distance. Finding such curves is complicated by the fact that any time-reparameterization of the curve is authorized. Thus one often defines geodesics as critical points of the energy functional $\mathcal{E}(\gamma) = \frac{1}{2} \int \|\dot{\gamma}\|^2\, dt$. It turns out that critical points for the energy also optimize the length functional. Moreover, they are parameterized proportionally to their arc length removing the ambiguity of the parametrization.

We now define the Christoffel symbols from the metric g by

$$\Gamma^k_{\ ij} = \frac{1}{2} g^{km} \left(\partial_{x^i} g_{jm} + \partial_{x^j} g_{mi} - \partial_{x^m} g_{ij} \right). \tag{1.7}$$

Using the calculus of variations, it can be shown that the geodesics satisfy the second order differential system

$$\ddot{\gamma}^k + \Gamma^k_{\ ij} \dot{\gamma}^i \dot{\gamma}^j = 0 \tag{1.8}$$

We will see the Christoffel symbols again below for the connection.

1.3.4. Levi-Civita Connection

The fundamental theorem of Riemannian geometry states that on any Riemannian manifold, there is a unique connection which is compatible with the metric and which has the property of being torsion-free. This connection is called the Levi-Civita connection. For that choice of connection, shortest curves have zero acceleration and are thus geodesics in the sense of being "straight lines". In the following, we only consider the Levi-Civita connection unless explicitly stated.

The connection allows us to take derivatives of a vector field Y in the direction of another vector field X expressed as $\nabla_X Y$. This is also denoted the covariant derivative of Y along X. The connection is linear in X and obeys the product rule in Y so that $\nabla_X(fY) = X(f)Y + f\nabla_X Y$ for a function $f : \mathcal{M} \to \mathbb{R}$ with $X(f)$ being the derivative of f in the direction of X using the interpretation of tangent vectors as derivations. In a local coordinate system, we can write the connection explicitly using the Christoffel symbols by $\nabla_{\partial_{x^i}}\partial_{x^j} = \Gamma^k_{ij}\partial_{x^k}$. With vector fields X and Y having coordinates $X(x) = v^i(x)\partial_{x^i}$ and $Y(x) = w^i(x)\partial_{x^i}$, we can use this to compute the coordinate expression for derivatives of Y along X:

$$\nabla_X Y = \nabla_{v^i\partial_{x^i}}(w^j\partial_{x^j}) = v^i\nabla_{\partial_{x^i}}(w^j\partial_{x^j}) = v^i(\partial_{x^i}w^j)\partial_{x^j} + v^iw^j\nabla_{\partial_{x^i}}\partial_{x^j}$$
$$= v^i(\partial_{x^i}w^j)\partial_{x^j} + v^iw^j\Gamma^k_{ij}\partial_{x^k} = \left(v^i(\partial_{x^i}w^k) + v^iw^j\Gamma^k_{ij}\right)\partial_{x^k} = \left(X(w^k) + v^iw^j\Gamma^k_{ij}\right)\partial_{x^k}.$$

Using this, the connection allows us to write the geodesic equation (1.8) as the zero acceleration constraint:

$$0 = \nabla_{\dot\gamma}\dot\gamma = \left(\dot\gamma(\dot\gamma^k) + \dot\gamma^i\dot\gamma^j\Gamma^k_{ij}\right)\partial_{x^k} = \left(\ddot\gamma^k + \dot\gamma^i\dot\gamma^j\Gamma^k_{ij}\right)\partial_{x^k}.$$

The connection also defines the notion of parallel transport along curves. A vector $v \in T_{\gamma(t_0)}\mathcal{M}$ is parallel transported if it is extended to a t-dependent family of vectors with $v_t \in T_{\gamma(t)}\mathcal{M}$ and $\nabla_{\dot\gamma(t)}v_t = 0$ for each t. Parallel transport can thereby be seen as a map $P_{\gamma,t} : T_{\gamma(t_0)}\mathcal{M} \to T_{\gamma(t)}\mathcal{M}$ linking tangent spaces. The parallel transport inherits its linearity from the connection. It follows from the definition that γ is a geodesic precisely if $\dot\gamma(t) = P_{\gamma,t}(\dot\gamma(t_0))$.

It is a fundamental consequence of curvature that parallel transport is dependent on the curve along which the vector is transported: With curvature, the parallel transports $P_{\gamma,T}$ and $P_{\phi,T}$ along two curves γ and ϕ with the same end-points $\gamma(t_0) = \phi(t_0)$ and $\gamma(T) = \phi(T)$ will differ. The difference is denoted holonomy, and the holonomy of a Riemannian manifold vanishes only if $\mathcal{M}$ is flat, i.e. has zero curvature.

1.3.5. Completeness

The Riemannian manifold is said to be *geodesically complete* if the definition domain of all geodesics can be extended to $\mathbb{R}$. This means that the manifold has no boundary nor any singular point that we can reach in a finite time. For instance, $\mathbb{R}^d - \{0\}$ with

the usual metric is not geodesically complete because some geodesics will hit 0 and thus stop being defined in finite time. On the other hand, $\mathbb{R}^d$ is geodesically complete. Other examples of complete Riemannian manifolds include compact manifolds implying that $\mathbb{S}^d$ is geodesically complete. This is a consequence of the Hopf-Rinow-De Rham theorem that also states that geodesically complete manifolds are complete metric spaces with the induced distance, and that there always exists at least one minimizing geodesic between any two points of the manifold, i.e. a curve whose length is the distance between the two points.

From now on, we will assume that the manifold is geodesically complete. This assumption is one of the fundamental properties ensuring well-posedness of algorithms for computing on manifolds.

1.3.6. Exponential and Logarithm Maps

Let x be a point of the manifold that we consider as a local reference point and v a vector of the tangent space $T_x\mathcal{M}$ at that point. From the theory of second order differential equations, it can be shown that there exists one and only one geodesic $\gamma_{(x,v)}(t)$ starting from that point $x = \gamma_{(x,v)}(0)$ with tangent vector $v = \dot{\gamma}_{(x,v)}(0)$. This geodesic is at first defined in a sufficiently small interval around zero but since the manifold is assumed geodesically complete, its definition domain can be extended to $\mathbb{R}$. Thus, the points $\gamma_{(x,v)}(t)$ are defined for each t and each $v \in T_x\mathcal{M}$. This allows to map vectors in the tangent space to the manifold using geodesics: the vector $v \in T_x\mathcal{M}$ can be mapped to the point of the manifold that is reached after a unit time $t = 1$ by the geodesic $\gamma_{(x,v)}(t)$ starting at x with tangent vector v. This mapping

$$\mathrm{Exp}_x \quad : \quad \begin{array}{rcl} T_x\mathcal{M} & \longrightarrow & \mathcal{M} \\ v & \longmapsto & \mathrm{Exp}_x(v) = \gamma_{(x,v)}(1) \end{array}$$

is called the *exponential map* at point x. Straight lines passing 0 in the tangent space are transformed into geodesics passing the point x on the manifold and distances along these lines are conserved (Fig. 1.6).

When the manifold is geodesically complete, the exponential map is defined on the entire tangent space $T_x\mathcal{M}$ but it is generally one-to-one only locally around 0 in the tangent space corresponding to a local neighborhood of x on $\mathcal{M}$. In the sequel, we denote by $\overrightarrow{xy}$ or $\mathrm{Log}_x(y)$ the inverse of the exponential map where the inverse is defined: this is the smallest vector as measured by the Riemannian metric such that $y = \mathrm{Exp}_x(\overrightarrow{xy})$. In this chart, the geodesics going through x are represented by the lines going through the origin: $\mathrm{Log}_x\gamma(x, \overrightarrow{xy})(t) = t\,\overrightarrow{xy}$. Moreover, the distance with respect to the base point x is preserved:

$$\mathrm{dist}(x, y) = \|\overrightarrow{xy}\| = \sqrt{\left\langle \overrightarrow{xy}, \overrightarrow{xy} \right\rangle_x} .$$
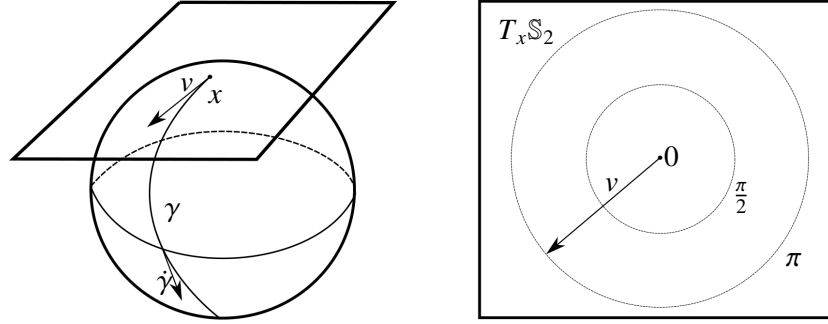
Figure 1.6 (left) Geodesics starting at x with initial velocity $v \in T_x\mathcal{M}$ are images of the exponential map $\gamma(t) = \text{Exp}_x(tv)$. They have zero acceleration $\nabla_{\dot\gamma}\dot\gamma$ and their velocity vectors are parallel transported $\dot\gamma(t) = P_{\gamma,t}(\dot\gamma(t_0))$. Geodesics locally realize the Riemannian distance so that $\text{dist}(x, \gamma(t)) = t\,\|v\|$ for sufficiently small t. (right) The tangent space $T_x\mathbb{S}^2$ and Exp_x give an exponential chart mapping vectors $v \in T_x\mathbb{S}^2$ to points in $\mathbb{S}^2$ by $\text{Exp}_x(v)$. The cut locus of x is its antipodal point, and the injectivity radius is π. Note that the equator is the set $\left\{\text{Exp}_x(v) \,\middle|\, \|v\| = \frac{\pi}{2}\right\}$.

Thus, the *exponential chart at x* gives a local representation of the manifold in the tangent space at a given point. This is also called a *normal coordinate system* or *normal chart* if it is provided with an orthonormal basis. At the origin of such a chart, the metric reduces to the identity matrix and the Christoffel symbols vanish. Note again that the exponential map is generally only invertible locally around $0 \in T_x\mathcal{M}$ and $\text{Log}_x y$ is therefore only locally defined, i.e. for points y near x.

The exponential and logarithm maps are commonly referred to as the Exp and Logs maps.

1.3.7. Cut Locus

It is natural to search for the maximal domain where the exponential map is a diffeomorphism. If we follow a geodesic $\gamma_{(x,v)}(t) = \text{Exp}_x(t\,v)$ from $t = 0$ to infinity, it is either always minimizing for all t, or it is minimizing up to a time $t_0 < \infty$. In this last case, the point $z = \gamma_{(x,v)}(t_0)$ is called a *cut point* and the corresponding tangent vector $t_0 v$ a *tangential cut point*. The set of all cut points of all geodesics starting from x is the *cut locus* $C(x) \in \mathcal{M}$, and the set of corresponding vectors the *tangential cut locus* $C(x) \in T_x\mathcal{M}$. Thus, we have $C(x) = \text{Exp}_x(C(x))$, and the maximal definition domain for the exponential chart is the domain $\mathcal{D}(x)$ containing 0 and delimited by the tangential cut locus.

It is easy to see that this domain is connected and star-shaped with respect to the

origin of $T_x\mathcal{M}$. Its image by the exponential map covers the manifold except the cut locus and the segment $[0, \overrightarrow{xy}]$ is transformed into the unique minimizing geodesic from x to y. Hence, the exponential chart has a connected and star-shaped definition domain that covers all the manifold except the cut locus $C(x)$:

$$\begin{aligned} \mathcal{D}(x) \subseteq \mathbb{R}^d &\longleftrightarrow \mathcal{M} - C(x) \\ \overrightarrow{xy} = \mathrm{Log}_x(y) &\longleftrightarrow y = \mathrm{Exp}_x(\overrightarrow{xy}) \end{aligned} \, .$$

From a computational point of view, it is often interesting to extend this representation to include the tangential cut locus. However, we have to take care of the multiple representations: Points in the cut locus where several minimizing geodesics meet are represented by several points on the tangential cut locus as the geodesics are starting with different tangent vectors (e.g. antipodal points on the sphere and rotation of π around a given axis for 3D rotations). This multiplicity problem cannot be avoided as the set of such points is dense in the cut locus.

The size of $\mathcal{D}(x)$ is quantified by the *injectivity radius* $\mathrm{i}(\mathcal{M}, x) = \mathrm{dist}(x, C(x))$, which is the maximal radius of centered balls in $T_x\mathcal{M}$ on which the exponential map is one-to-one. The injectivity radius of the manifold $\mathrm{i}(\mathcal{M})$ is the infimum of the injectivity over the manifold. It may be zero, in which case the manifold somehow tends towards a singularity (e.g. think to the surface $z = 1/\sqrt{x^2 + y^2}$ as a sub-manifold of $\mathbb{R}^3$).

Example 4. On the sphere $\mathbb{S}^d$ (center 0 and radius 1) with the canonical Riemannian metric (induced by the ambient Euclidean space $\mathbb{R}^{d+1}$), the geodesics are the great circles and the cut locus of a point x is its antipodal point $\underline{x} = -x$. The exponential chart is obtained by rolling the sphere onto its tangent space so that the great circles going through p become lines. The maximal definition domain is thus the open ball $\mathcal{D} = \mathcal{B}_d(\pi)$. On its boundary $\partial\mathcal{D} = C = \mathbb{S}^{d-1}(\pi)$, all the points represent $\underline{x}$, see Figure 1.6.

For the real projective space P_d (obtained by identification of antipodal points of the sphere $\mathbb{S}^d$), the geodesics are still the great circles, but the cut locus of the point $\{x, -x\}$ is now the equator of the two points, with antipodal points identified (thus the cut locus is P_{d-1}). The definition domain of the exponential chart is the open ball $\mathcal{D} = \mathcal{B}_d(\frac{\pi}{2})$, and the tangential cut locus is the sphere $\partial\mathcal{D} = \mathbb{S}^{d-1}(\frac{\pi}{2})$ where antipodal points are identified.

1.4. Elements of Analysis in Riemannian Manifolds

We here outline further constructions on manifolds relating to taking derivatives of functions, the intrinsic Riemannian measure, and defining curvature. These notions will be used in following chapters of this book for instance for optimization algo-

rithms.

1.4.1. Gradient and Musical Isomorphisms

Let f be a smooth function from $\mathcal{M}$ to $\mathbb{R}$. Recall that the differential $df(x)$ evaluated at the point $x \in \mathcal{M}$ is a covector in $T_x^*\mathcal{M}$. Therefore, contrary to the Euclidean situation where derivatives are often regarded as vectors, we cannot directly interpret $df(x)$ as a vector. However, thanks to the Riemannian metric, there is a canonical way to identify the linear form $df \in T_x^*\mathcal{M}$ with a unique vector $v \in T_x\mathcal{M}$. This is done by defining $v \in T_x\mathcal{M}$ to be a vector satisfying $df(w) = \langle v, w \rangle_x$ for all vectors $w \in T_x\mathcal{M}$. This mapping corresponds to the transpose operator that is implicitly used in Euclidean spaces to transform derivatives of functions (row vectors) to column vectors. On manifolds, the Riemannian metric must be specified explicitly since the coordinate system used may not be orthonormal everywhere.

The mapping works for any covector, and is often denoted the *sharp* map $^\sharp$: $T^*\mathcal{M} \to T\mathcal{M}$. It has an inverse in the *flat* map $^\flat : T\mathcal{M} \to T^*\mathcal{M}$. In coordinates, $(\xi^\sharp)^i = g^{ij}\xi_j$ for a covector $\xi = \xi_j dx^j$ and $(v^\flat)_i = g_{ij}v^j$ for a vector $v = \partial_{x^j} v^j$. The maps $^\sharp$ and $^\flat$ are denoted musical isomorphisms because they raise or lower the indices of the coordinates.

We can use the sharp map to define the Riemannian gradient as a vector:

$$\operatorname{grad} f = (df)^\sharp .$$

This definition corresponds to the classical gradient in $\mathbb{R}^k$ using the standard Euclidean inner product as Riemannian metric. Using the coordinate representation of the sharp map, we get the coordinate form $(\operatorname{grad} f)^i = g^{ij}\partial_{x^j} f$ of the gradient.

1.4.2. Hessian and Taylor Expansion

The covariant derivative of the gradient, the *Hessian*, arises from the connection ∇:

$$\operatorname{Hess} f(X, Y) = \nabla_X \nabla_Y f = (\nabla_X(df))Y = \langle \nabla_X \operatorname{grad} f, Y \rangle .$$

Here, the two expressions on the right are given using the action of the connection on the differential form df (a covector), or the vector field $\operatorname{grad} f = (df)^\sharp$. Its expression in a local coordinate system is

$$\operatorname{Hess} f = \nabla df = (\partial_{x^i x^j} f - \Gamma^k_{ij}\partial_k f)dx^i dx^j .$$

Let now f_x be the expression of f in a normal coordinate system at x. Its Taylor expansion around the origin in coordinates is:

$$f_x(v) = f_x(0) + df_x v + \frac{1}{2}v^\top H_{f_x} v + O(\|v\|^3)$$

where $df_x = (\partial_{x^i} f)$ is the Jacobian matrix of first order derivatives and $H_{f_x} = (\partial_{x^i x^j} f)$ is the Euclidean Hessian matrix. Because the coordinate system is normal, we have $f_x(v) = f(\mathrm{Exp}_x(v))$. Moreover, the metric at the origin reduces to the identity: $df_x = (\mathrm{grad}\, f)^T$, and the Christoffel symbols vanish so that the matrix of second derivatives H_{f_x} corresponds to the Hessian $\mathrm{Hess}\, f$. Thus, The Taylor expansion can be written in any coordinate system:

$$f\left(\mathrm{Exp}_x(v)\right) = f(x) + \mathrm{grad}\, f(v) + \frac{1}{2}\mathrm{Hess}\, f(v, v) + O(\|v\|^3) \qquad (1.9)$$

1.4.3. Riemannian Measure or Volume Form

In a vector space with basis $\mathcal{A} = (a_1, \ldots a_n)$, the local representation of the metric is given by $g = A^\top A$ where $A = [a_1, \ldots a_n]$ is the matrix of coordinates change from $\mathcal{A}$ to an orthonormal basis. Similarly, the measure or the infinitesimal volume element is given by the volume of the parallelepiped spanned by the basis vectors: $d\mathcal{V} = |A|\, dx = \sqrt{|g|}\, dx$ with $|\cdot|$ denoting the matrix determinant. In a Riemannian manifold $\mathcal{M}$, the Riemannian metric $g(x)$ induces an infinitesimal volume element on each tangent space, and thus a measure on the manifold that in coordinates has the expression

$$d\mathcal{M}(x) = \sqrt{|g(x)|}\, dx\,.$$

The cut locus has null measure, and we can therefore integrate indifferently in $\mathcal{M}$ or in any exponential chart. If f is an integrable function of the manifold and $f_x(v) = f(\mathrm{Exp}_x(v))$ is its image in the exponential chart at x, we have

$$\int_{x \in \mathcal{M}} f(x) d\mathcal{M}(x) = \int_{v \in \mathcal{D}(x)} f_x(v) \sqrt{|g(\mathrm{Exp}_x(v))|}\, dv\,,$$

1.4.4. Curvature

The curvature of a Riemannian manifold measures its deviance from local flatness. We often have a intuitive notion of when a surface embedded in $\mathbb{R}^3$ is flat or curved: For example, a linear subspace of $\mathbb{R}^3$ is flat while the sphere $\mathbb{S}^2$ is curved. This idea of curvature is expressed in the Gauss curvature. However, for high dimensional spaces, the mathematical description becomes somewhat more intricate. We will see below several notions of curvature capturing aspects of the nonlinearity of the manifold with varying details. Important to note is that while vanishing curvature implies local flatness of the manifold, this is not the same as the manifold being globally Euclidean. An example is the torus $\mathbb{T}_2$ which can both be embedded in $\mathbb{R}^3$ inheriting nonzero curvature and be embedded in $\mathbb{R}^4$ in a way in which it inherits a flat geometry. In both cases, the periodicity of the torus remains, which prevents it from being a vector space.

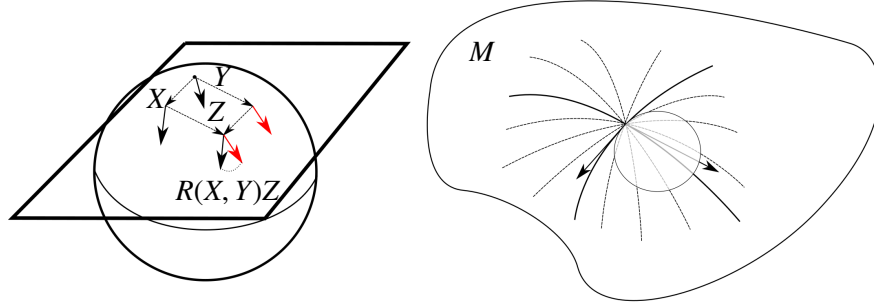The curvature of a Riemannian manifold is described by the curvature tensor R:

Figure 1.7 (Left) The curvature tensor describes the difference in parallel transport of a vector Z around an infinitesimal parallelogram spanned by the vector fields X and Y (dashed vectors). (Right) The sectional curvature measures the product of principal curvatures in a 2D submanifold given as the geodesic spray of a subspace V of $T_x\mathcal{M}$. The principal curvatures arise from comparing these geodesics to circles as for the Euclidean notion of curvature of a curve.

$T\mathcal{M} \times T\mathcal{M} \times T\mathcal{M} \to T\mathcal{M}$. It is defined from the covariant derivative by evaluation on vector fields X, Y, Z:

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]}Z. \tag{1.10}$$

The bracket $[X, Y]$ denotes the anticommutativity of the fields X and Y. If f is a differentiable function on $\mathcal{M}$, the new vector field produced by the bracket is given by its application to f: $[X, Y]f = X(Y(f)) - Y(X(f))$. The curvature tensor R can intuitively be interpreted at $x \in \mathcal{M}$ as the difference between parallel transporting the vector $Z(x)$ along an infinitesimal parallelogram with sides gives by $X(x)$ and $Y(x)$, see Figure 1.7 (left). As noted earlier, parallel transport is curve dependent, and the difference between transporting infinitesimally along $X(x)$ and then $Y(x)$ as opposed to along $Y(x)$ and then $X(x)$ is a vector in $T_x\mathcal{M}$. This difference can be calculated for any vector $z \in T_x\mathcal{M}$. The curvature tensor when evaluated at X, Y, i.e. $R(X, Y)$, is the linear map $T_x\mathcal{M} \to T_x\mathcal{M}$ given by this difference.

The reader should note that two different sign conventions exist for the curvature tensor: the definition (1.10) is used in a number of reference books in physics and mathematics [MTW73, Lee97, Kli82, Pos10, Jos11]. Other authors use a minus sign to simplify some of the tensor notations [Spi79, O'N83, dC92, GHL93, Ber03] and different order conventions for the tensors subscripts and/or a minus sign in the sectional curvature defined below (see e.g. the discussion in [GQ16, p. 399]).

The curvature can be realized in coordinates from the Christoffel symbols:

$$R(\partial_{x^i}, \partial_{x^j})\,\partial_{x^k} = R^m{}_{kij}\,\partial_{x^m} = (\Gamma^l{}_{jk}\Gamma^m{}_{il} - \Gamma^l{}_{ik}\Gamma^m{}_{jl} + \partial_i\Gamma^m{}_{jk} - \partial_j\Gamma^m{}_{ik})\,\partial_{x^m}. \tag{1.11}$$

The sectional curvature κ measures the Gaussian curvature of 2D submanifolds of $\mathcal{M}$, the Gaussian curvature of each point being the product of the principal curvatures of curves passing the point. The 2D manifolds arise as the Gaussian curvature of the geodesic spray of a 2D linear subspace of $T_x\mathcal{M}$, see Figure 1.7 (right). Such a 2-plane can be represented by basis vectors $u, v \in T_x\mathcal{M}$ in which case the sectional curvature can be expressed using the curvature tensor by

$$\kappa(u, v) = \frac{\langle R(u, v)v, u \rangle}{\|u\|^2 \|v\|^2 - \langle u, v \rangle^2} .$$

(1.12)

The curvature tensor gives the notion of Ricci and scalar curvature that both provides summary information of the full tensor R. The Ricci curvature Ric is the trace over the first and last indices of R with coordinate expression

$$R_{ij} = R_{kij}{}^k = R^k{}_{ikj} = g^{kl} R_{ikjl} .$$

(1.13)

Taking another trace, we get the scalar valued quantity the scalar curvature S:

$$S = g^{ij} R_{ij} .$$

(1.14)

Note that the cometric appears to raise one index before taking the trace.

1.5. Lie groups and Homogeneous Manifolds

A Lie group is a manifold equipped with additional group structure such that the group multiplication and group inverse are smooth mappings. Many of the interesting transformations used in image analysis, translations, rotations, affine transforms, etc., form Lie groups. We will in addition see examples of infinite dimensional Lie groups when doing shape analysis with diffeomorphisms as described in chapter 4. We begin by reviewing the definition of an algebraic group.

Definition 5 (Group)**.** A *group* is a set G with a binary operation, denoted here by concatenation or group product, such that
1. $(xy)z = x(yz)$, for all $x, y, z \in G$,
2. there is an *identity* element $e \in G$ satisfying $xe = ex = x$, for all $x \in G$,
3. each $x \in G$ has an *inverse*, $x^{-1} \in G$, satisfying $xx^{-1} = x^{-1}x = e$.

A Lie group is simultaneously a group and a manifold, with compatibility between these two mathematical concepts.

Definition 6 (Lie Group)**.** A *Lie group G* is a smooth manifold that also forms a group,

where the two group operations,

$$(x, y) \mapsto xy \qquad\qquad : \quad G \times G \to G \qquad\qquad \textit{product}$$

$$x \mapsto x^{-1} \qquad\qquad\quad : \quad G \to G \qquad\qquad\qquad \textit{inverse}$$

are smooth mappings of manifolds.

Example 5. The space of all $k \times k$ non-singular matrices forms a Lie group called the *general linear group*, denoted GL(k). The group operation is matrix multiplication, and GL(k) can be given a smooth manifold structure as an open subset of $\mathbb{R}^{k^2}$. The equations for matrix multiplication and inverse are smooth operations in the entries of the matrices. Thus, GL(k) satisfies the requirements of a Lie group in Definition 6. A *matrix group* is any closed subgroup of GL(k). Matrix groups inherit the smooth structure of GL(k) as a subset of $\mathbb{R}^{k^2}$ and are thus also Lie groups.

Example 6. The $k \times k$ rotation matrices form a closed matrix subgroup of GL(k) and thus a Lie group. This group is called the *special orthogonal group*. It is defined as SO(k) = $\{R \in$ GL(k) $: R^T R = \mathrm{Id}_k$ and $\det(R) = 1\}$. This space is a closed and bounded subset of $\mathbb{R}^{k^2}$ and thus compact.

Example 7. Classical geometric transformation groups used in image registration such as rigid-body transformations, similarities and affine transformations can also be looked upon as matrix groups via their faithful representation based on homogeneous coordinates.

For each y in a Lie group G, the following two diffeomorphisms of G are denoted left- and right-translations by y:

$$L_y : x \mapsto yx \qquad\qquad (\textit{left multiplication})$$

$$R_y : x \mapsto xy \qquad\qquad (\textit{right multiplication})$$

The differential or pushforward $(L_y)_*$ of the left translation maps the tangent space $T_x G$ to the tangent space $T_{yx}G$. In particular, $(L_y)_*$ maps any vector $u \in T_e G$ to the vector $(L_y)_* u \in T_y G$ thereby giving rise to the vector field $\tilde{u}(y) = (L_y)_* u$. Such a vector field is said to be *left-invariant* since it is invariant under left multiplication: $\tilde{u} \circ L_y = (L_y)_* \tilde{u} = \tilde{u}$ for every $y \in G$. *Right-invariant* vector fields are defined similarly. A left- or right-invariant vector field is uniquely defined by its value on the tangent space at the identity, $T_e G$.

Recall that vector fields on G can be seen as derivations on the space of smooth functions, $C^\infty(G)$. Thus, two vector fields u and v can be composed to form another operator uv on $C^\infty(G)$ but the operator uv is not necessarily a derivation as it includes

second order differential terms. However, the operator $uv - vu$ is a vector field on G. Indeed, we can check by writting this expression in a local coordinate system that the second order terms vanish. This leads to a definition of the *Lie bracket* of vector fields u, v on G, defined as

$$[u, v] = uv - vu. \tag{1.15}$$

This is also sometimes called the Lie derivative $\mathcal{L}_u v = [u, v]$ because it is conceptually the derivative of the vector field v in the direction $u(x)$ generated by u at each point $x \in G$.

Definition 7 (Lie algebra). A *Lie algebra* is a vector space V equipped with a bilinear product $[\cdot, \cdot] : V \times V \to V$, called a *Lie bracket*, that satisfies
1. $[u, v] = -[v, u]$ (skew symmetry),
2. $[[u, v], w] + [[v, w], u] + [[w, u], v] = 0$ (Jacobi identity),
for all $u, v, w \in V$.

The tangent space of a Lie group G at the identity element, $T_e G$, typically denoted $\mathfrak{g}$, forms a Lie algebra. The Lie bracket on $\mathfrak{g}$ is induced by the Lie bracket on the corresponding left-invariant vector fields. If u, v are two vectors in $\mathfrak{g}$, let $\tilde{u}, \tilde{v}$ be the corresponding unique left-invariant vector fields on G. Then the Lie bracket on $\mathfrak{g}$ is given by

$$[u, v] = [\tilde{u}, \tilde{v}](e).$$

The Lie bracket provides a test for whether the Lie group G is commutative. A Lie group G is commutative if and only if the Lie bracket on the corresponding Lie algebra $\mathfrak{g}$ is zero, i.e., $[u, v] = 0$ for all $u, v \in \mathfrak{g}$.

Example 8. The Lie algebra for Euclidean space $\mathbb{R}^k$ is again $\mathbb{R}^k$. The Lie bracket is zero, i.e., $[X, Y] = 0$ for all $X, Y \in \mathbb{R}^k$.

Example 9. The Lie algebra for GL(k) is $\mathfrak{gl}(k)$, the space of all real $k \times k$ matrices. The Lie bracket operation for $X, Y \in \mathfrak{gl}(k)$ is given by

$$[X, Y] = XY - YX.$$

Here the product XY denotes actual matrix multiplication, which turns out to be the same as composition of the vector field operators (compare to (1.15)). All Lie algebras corresponding to matrix groups are subalgebras of $\mathfrak{gl}(k)$.

Example 10. The Lie algebra for the rotation group SO(k) is $\mathfrak{so}(k)$, the space of skew-symmetric matrices. A matrix A is skew-symmetric if $A = -A^T$.

1.5.1. One-parameter Subgroups

Let $\tilde{u}(y) = (L_y)_* u$ be a left-invariant vector field. The solution $x(t)$ to the initial value problem

$$\dot{x}(t) = \tilde{u}(x(t)), \ x(0) = e$$

is called a one-parameter subgroup because it is a morphing of Lie groups: $x(s + t) = x(s)x(t)$. The Lie exponential map $\exp : \mathfrak{g} \to G$ is then given by the value of $x(t)$ at $t = 1$, i.e. $\exp(u) = x(1)$. For matrix groups where the Lie group algebra consists of ordinary matrices, exp corresponds to the matrix exponential. The group exponential should not be confused with the Riemannian exponential as they usually differ, unless the group is provided with a bi-invariant metric.

1.5.2. Actions

Let $\mathcal{M}$ be a manifold and G a Lie group. The elements of the group can often be used to produce variations of elements of the manifold, e.g. elements of $GL(k)$ linearly transforms elements of the manifold $\mathbb{R}^k$. Similarly, affine transformations applies to change images in image registration. These are examples of actions of G on $\mathcal{M}$. Such actions are usually denoted $g.x$ where $g \in G$, $x \in \mathcal{M}$. Because the action involves two manifolds, G and $\mathcal{M}$, we will below use x, y to denote elements of $\mathcal{M}$, and g, h to denote elements of G.

Definition 8 (Action). A left action of a Lie group G on a manifold $\mathcal{M}$ is a smooth mapping $. : G \times \mathcal{M} \to \mathcal{M}$ satisfying
1. $e.x = x, \forall x \in \mathcal{M}$,
2. $h.(g.x) = (hg).x, \forall x \in M$,
3. the map $x \mapsto g.x$ is a diffeomorphism of M for each $g \in G$.

We will see examples of Lie group actions throughout the book. For example, Chapter 4 on shape analysis relies fundamentally on actions of the group $\mathrm{Diff}(\Omega)$ of diffeomorphisms of a domain Ω on shape spaces $\mathcal{S}$.

Through the action, a curve $g(t)$ on the group G acts on a point $x \in \mathcal{M}$ to give a curve $g(t).x$ in $\mathcal{M}$. In particular, one-parameter subgroups define the curves $x_v(t) = \exp(tv).x$ in $\mathcal{M}$ for Lie algebra elements $v \in \mathfrak{g}$. The derivative

$$v_\mathcal{M}(x) := \frac{d}{dt} \exp(tv).x$$

in $T_{x_v(t)}\mathcal{M}$ is denoted the infinitesimal generator associated to v.

Some particularly important actions are the actions of G on itself and on the Lie algebra $\mathfrak{g}$. These include the actions by left translation $g.h := L_g(h) = gh$, and the action by conjugation $g.h := L_g(R_{g^{-1}})h = ghg^{-1}$. The pushforward of the conjugation

gives the adjoint action $g.v = (L_g \circ R_{g^{-1}})_* v$ of G on $\mathfrak{g}$. The adjoint action is also denoted $\mathrm{Ad}_g v$.

From the adjoint action, we get the adjoint operator $\mathrm{ad}_u v = \frac{d}{dt}\mathrm{Ad}_{\exp(tu)} v$ for $v, u \in \mathfrak{g}$. This operator is sometimes informally denoted "little ad", and it is related to the Lie bracket by $\mathrm{ad}_u v = [u, v]$.

The actions on the Lie algebra have dual actions as well, denoted co-actions: the co-adjoint action $G \times \mathfrak{g}^* \to \mathfrak{g}^*$, $g.\xi := \mathrm{Ad}^*_{g^{-1}}\xi$ for $\xi \in \mathfrak{g}^*$ where the dual of the adjoint is given by $\mathrm{Ad}^*_{g^{-1}}\xi(v) = \xi(\mathrm{Ad}_{g^{-1}}v)$ for all $v \in \mathfrak{g}$. Using the notation $(\xi|v)$ for evaluation $\xi(v)$ of ξ on v, the definition of the dual of the adjoint is $(\mathrm{Ad}^*_{g^{-1}}\xi|v) = (\xi|\mathrm{Ad}_{g^{-1}}v)$. The coadjoint operator $\mathrm{ad}^* : \mathfrak{g} \times \mathfrak{g}^* \to \mathfrak{g}^*$ is similarly specified by $(\mathrm{ad}^*_v\xi|u) = (\xi|\mathrm{ad}_v u)$ for $v, u \in \mathfrak{g}$ and $\xi \in \mathfrak{g}^*$.

1.5.3. Homogeneous Spaces

Let the group G act on $\mathcal{M}$. If, for any $x, y \in \mathcal{M}$, there exists $g \in G$ such that $g.x = y$, the action is said to be transitive. In this case, the manifold $\mathcal{M}$ is homogeneous. For a fixed $x \in \mathcal{M}$, the closed subgroup $H = \{g \in g \mid g.x = x\}$ is denoted the isotropy subgroup of G, and $\mathcal{M}$ is isomorphic to the quotient G/H. Similarly, a closed subgroup H of G leads to a homogeneous space G/H by quotienting out H. Examples of homogeneous spaces are the spheres $\mathbb{S}^n \equiv \mathrm{SO}(n + 1)/\mathrm{SO}(n)$ and the orbit shape spaces described in Chapter 4, e.g. the manifold of landmark configurations.

1.5.4. Invariant Metrics and Geodesics

The left- and right translation maps gives a particularly useful way of defining Riemannian metrics on Lie groups. Given an inner product $\langle \cdot, \cdot \rangle_\mathfrak{g}$ on the Lie algebra, we can extend it to an inner product on tangent spaces at all elements of the group by setting

$$\langle u, v \rangle_g := \left\langle (L_{g^{-1}})_* u, (L_{g^{-1}})_* v \right\rangle_\mathfrak{g} .$$

This defines a left-invariant Riemannian metric on G because, $\langle (L_h)_* u, (L_h)_* v \rangle_{hg} = \langle u, v \rangle_g$ for any $u, v \in T_g G$. Similarly, we can set

$$\langle u, v \rangle_g := \left\langle (R_{g^{-1}})_* u, (R_{g^{-1}})_* v \right\rangle_\mathfrak{g}$$

to get a right-invariant metric. In the particular case where the metric is invariant to both left- and right-translation, it called bi-invariant.

Geodesics for bi-invariant metrics are precisely one-parameter subgroups, and the Lie group exponential map exp therefore equals the Riemannian exponential map Exp_e. For metrics that are left- or right-invariant, but not bi-invariant, the ordinary geodesic equation (1.8) can be simplified using for example Euler-Poincaré reduction. The resulting Euler-Poincaré equations are discussed further in chapter 4 in the case

of right-invariant metrics on the group $\mathrm{Diff}(\Omega)$.

1.6. Elements of Computing on Riemannian Manifolds

The Riemannian Exp and Log maps constitute very powerful atomic functions to express most geometric operations for performing statistical computing on manifolds. The implementation of Log_x and Exp_x is therefore the algorithmic basis of programming on Riemannian manifolds, as we will see in the following.

In a Euclidean space, exponential charts are nothing but orthonormal coordinates systems translated to each point: In this case, $\overrightarrow{xy} = \mathrm{Log}_x(y) = y - x$ and $\mathrm{Exp}_x(v) = x + v$. This example is more than a simple coincidence. In fact, most of the usual operations using additions and subtractions may be reinterpreted in a Riemannian framework using the notion of *bipoint*, an antecedent of vector introduced during the 19th century. Indeed, one defines vectors as equivalent classes of bipoints, oriented couples of points, in a Euclidean space. This is possible using the canonical way to compare what happens at two different points by translating. In a Riemannian manifold, we can compare vectors using the parallel transport along curves but the curve dependence of the parallel transport prevents global comparison of vectors as in Euclidean space. This implies that each vector has to remember at which point of the manifold it is attached, as is the case for tangent vectors, which relates back to the Euclidean notion of a bipoint.

Conversely, the logarithm map may be used to map almost any bipoint (x, y) into a vector $\overrightarrow{xy} = \mathrm{Log}_x(y)$ of $T_x\mathcal{M}$. This reinterpretation of addition and subtraction using logarithm and exponential maps is very powerful when generalizing algorithms working on vector spaces to algorithms on Riemannian manifolds. This is illustrated in Table 1.1 and in the following sections.

	Euclidean space	Riemannian manifold
Subtraction	$\overrightarrow{xy} = y - x$	$\overrightarrow{xy} = \mathrm{Log}_x(y)$
Addition	$y = x + v$	$y = \mathrm{Exp}_x(v)$
Distance	$\mathrm{dist}(x, y) = \|y - x\|$	$\mathrm{dist}(x, y) = \|\overrightarrow{xy}\|_x$
Mean value (implicit)	$\sum_i(x_i - \bar{x}) = 0$	$\sum_i \overrightarrow{\bar{x}x_i} = 0$
Gradient descent	$x_{t+\varepsilon} = x_t - \varepsilon\nabla f(x_t)$	$x_{t+\varepsilon} = \mathrm{Exp}_{x_t}(-\varepsilon\,\mathrm{grad}\,f(x_t))$
Geodesic interpolation	$x(t) = x_0 + t\,\overrightarrow{x_0x_1}$	$x(t) = \mathrm{Exp}_{x_0}(t\,\overrightarrow{x_0x_1})$

Table 1.1 Re-interpretation of standard operations in a Riemannian manifold.

The Exp and Log maps are different for each manifold and for each metric. They

must therefore be determined and implemented on a case by case basis. In some cases, closed form expressions are known, examples being the spheres $\mathbb{S}^d$, rotations and rigid body transformations with left invariant metric [PT97], and covariance matrices (positive definite symmetric matrices, so called tensors in medical image analysis) [PFA06] and Chapter 3. In cases where closed form solutions are not known, geodesics with given initial velocity can be obtained by numerically solving the geodesic ODE (1.8), or by solving the variational problem of finding a minimum energy curve between two points. Thus computing $\mathrm{Exp}_x(v)$ may be posed as a numerical integration problem, see e.g. [HM94, HWL02], and computing $\overrightarrow{xy} = \mathrm{Log}_x(y)$ an optimal control problem. This opens the way to statistical computing in more complex spaces than the spaces we have considered up to now, such as spaces of curves, surfaces, and diffeomorphic transformations, as we will see in the following chapters. Geometric computation frameworks such as Theano Geometry[1] [KAS17] and Geomstats[2] provide numerical implementations of geometric operations on some commonly used manifolds. Theano Geometry uses automatic differentiation to express and compute the derivatives that are essential for differential geometric computations. This results in convenient code for computing Christoffel symbols, curvature tensors, and fiber bundle operations using the parallel transport.

1.7. Examples

Below, we survey ways to express the Exp and Log maps on selected manifolds at the same time exemplifying how particular structure of the spaces can be used for computations.

1.7.1. The Sphere

Let x be a point in $\mathbb{S}^d$. From the embedding of $\mathbb{S}^d$ in $\mathbb{R}^{d+1}$, the tangent space $T_x\mathbb{S}^d$ can be identified with the d-dimensional vector space of all vectors in $\mathbb{R}^{d+1}$ orthogonal to x. The inner product between two tangent vectors is then equivalent to the usual Euclidean inner product. The exponential map is given by a 2D rotation of x by an angle given by the norm of the tangent, i.e.,

$$\mathrm{Exp}_x(v) = \cos\theta\, x + \frac{\sin\theta}{\theta}\, v, \quad \theta = \|v\|. \tag{1.16}$$

The log map between two points x, y on the sphere can be computed by finding the initial velocity of the rotation between the two points. Let $\pi_x(y) = x \langle x, y \rangle$ denote the

---

projection of the vector y onto x. Then,

$$\text{Log}_x(y) = \frac{\theta \; (y - \pi_x(x))}{\|y - \pi_x(y)\|}, \quad \theta = \arccos(\langle x, y \rangle). \tag{1.17}$$

1.7.2. 2D Kendall Shape Space

Kendall shape space [Ken84] represents a shape as an equivalence class of all translations, rotations, and scalings of a set of k points, landmarks, in the plane. A configuration of k points in the 2D plane is considered a complex k-vector, $z \in \mathbb{C}^k$. Removing translation by requiring the centroid to be zero projects this point to the linear complex subspace $V = \{z \in \mathbb{C}^k : \sum z_i = 0\}$, which is isomorphic to the space $\mathbb{C}^{k-1}$. Next, points in this subspace are deemed equivalent if they are a rotation and scaling of each other, which can be represented as multiplication by a complex number, $\rho e^{i\theta}$, where ρ is the scaling factor and θ is the rotation angle. The set of such equivalence classes forms the complex projective space, $\mathbb{C}P^{k-2}$.

We think of a centered shape $p \in V$ as representing the complex line $L_p = \{z \, p : z \in \mathbb{C}\backslash\{0\}\}$, i.e., L_p consists of all point configurations with the same shape as p. A tangent vector at $L_p \in V$ is a complex vector, $v \in V$, such that $\langle p, v \rangle = 0$. The exponential map is given by rotating within V the complex line L_p by the initial velocity v:

$$\text{Exp}_p(v) = \cos \theta \, p + \frac{\|p\| \sin \theta}{\theta} \, v, \quad \theta = \|v\|. \tag{1.18}$$

Likewise, the log map between two shapes $p, q \in V$ is given by finding the initial velocity of the rotation between the two complex lines L_p and L_q. Let $\pi_p(q) = p \langle p, q \rangle / \|p\|^2$ denote the projection of the vector q onto p. Then the log map is given by

$$\text{Log}_p(q) = \frac{\theta \left(q - \pi_p(q)\right)}{\|q - \pi_p(q)\|}, \quad \theta = \arccos \frac{|\langle p, q \rangle|}{\|p\| \, \|q\|}. \tag{1.19}$$

We will see an example of a different landmark space equipped with a geometric structure coming from the action of the diffeomorphism group in Chapter 4.

1.7.3. Rotations

The set of orthogonal transformations $\text{O}(k)$ on $\mathbb{R}^k$ discussed in Section 1.2.1 is the subset of linear maps of $\mathbb{R}^k$, square matrices $U \in \text{M}_{(k,k)}$, that preserve the dot product: $\langle Ux, Uy \rangle = \langle x, y \rangle$. In particular, they conserve the norm of a vector: $\|Ux\|^2 = \|x\|^2$. This means that $x^\top (U^\top U - \text{Id})x = 0$ for all vectors $x \in \mathbb{R}^k$, which is possible if and only if the matrix U satisfies the quadratic constraint $U^\top U = \text{Id}$. Thus, the inverse transformation is $U^{-1} = U^\top$. The composition of two such maps obviously also preserves the dot product so this forms the group of orthogonal transformations with the

identity matrix Id_k as neutral element e:

$$\mathrm{O}(k) = \left\{ U \in \mathrm{M}_{(k,k)} \mid U^\top U = \mathrm{Id}_k \right\}.$$

Because the quadratic constraint is smooth and differentiable, $O(k)$ constitute a Lie group, submanifold of the linear space of square matrices. However, it is not connected: Taking the determinant of the constraint gives $\det(U)^2 = \det(\mathrm{Id}_k)$, so that $\det(U) = \pm 1$. We see that there are two disconnected component of determinant $+1$ and -1 that cannot be joined by any continuous curve on the space of matrices. Such a curve would have to go through matrices with determinants between -1 and 1 since the determinant is a continuous function. The component of negative determinant includes symmetries that reverse the orientation of the space. It is not a subgroup because the composition of two such transformations of negative determinant has a positive determinant.

The component of positive determinant preserves the orientation of the space and is a subgroup, the group of rotations, or special orthogonal transformations:

$$\mathrm{SO}(k) = \left\{ R \in \mathrm{M}_{(k,k)} \mid R^\top R = \mathrm{Id}_k \,,\; \det(R) = 1 \right\}.$$

Let $R(t) = R + t\dot{R} + O(t^2)$ be a curve drawn on $\mathrm{SO}(k)$, considered as an embedded manifold in the vector space of matrices $\mathrm{M}_{(k,k)}$. The constraint $R^\top R = \mathrm{Id}_k$ is differentiated into

$$\dot{R}R^\top + \left(\dot{R}R^\top\right)^\top = 0 \qquad \text{or} \qquad R^\top\dot{R} + \left(R^\top\dot{R}\right)^\top = 0$$

which means that $\dot{R}R^\top$ and $R^\top\dot{R}$ are skew-symmetric matrices. Thus, the tangent space $T_e\mathrm{SO}(k)$ at identity is the vector space of skew matrices and the tangent space at rotation $R \in \mathrm{SO}(k)$ is its left or right translation:

$$T_R\mathrm{SO}(k) = \{X \in \mathrm{M}(k,k) \,/\, R^\top X = -(R^\top X)^\top\} = \{X \in \mathrm{M}(k,k) \,/\, XR^\top = -(XR^\top)^\top\}$$

Since $k \times k$ skew-symmetric matrices have $k(k-1)/2$ free components, we also obtain that the dimension of the special orthogonal group is $k(k-1)/2$.

In order to put a metric on this Lie group, we may take a metric on the tangent space at the identity and left translate it to any other point resulting in a left-invariant metric. We may similarly right translate it to obtain a right-invariant metric. Since $\mathrm{SO}(k)$ is a submanifold of the Euclidean space of matrices $\mathrm{M}_{(k,k)}$, we may also consider the restriction of the embedding Frobenius dot product $\mathrm{Tr}(X^\top Y)$ to the tangent spaces at all points. It is common to rescale the Frobenius metric by $1/2$ to compensate the fact that we are counting twice each off diagonal coefficient of the skew-symmetric matrices. This induces the metric

$$\langle X, Y \rangle_R = \frac{1}{2}\mathrm{Tr}(XY^\top)$$

on the tangent space $T_R\mathrm{SO}(k)$.

This metric is invariant by left and right translation. The existence of this bi-invariant metric is a particular case due to the compactness of the $\mathrm{SO}(k)$ group. Bi-invariant metrics on Lie groups have very special properties that will be described in Chapter 5. In particular, as mentioned earlier, geodesics passing the identity are one-parameter subgroups whose equations are given by the matrix exponential: $\mathrm{Exp}_e(X) = \exp(X) = \sum_{k=0}^{+\infty} \frac{X^k}{k!}$. This series is absolutely convergent so that the matrix exponential always exists. Its inverse, the logarithm, may however fail to exist and is also generally not unique when it exists.

For rotations, the exponential of skew symmetric matrices covers the whole rotation group so that the log always exists, but it is not unique: For $k = 2$, rotating of an angle θ is the same as rotating of an angle $\theta + 2l\pi$ where l is an integer. In order to understand the structure of rotations in higher dimensions, we may look at the spectral decomposition of a rotation matrix R: The characteristic polynomial $P(\lambda) = \det(R - \lambda \, \mathrm{Id}_k)$ is a real polynomial of degree k. Thus the k complex eigenvalues are real or conjugate by pairs, and the polynomial can be factored into at most $\lfloor k/2 \rfloor$ quadratic terms, potentially with multiplicity, and real linear terms. The conservation of the norm by the rotation $\|Rx\| = \|x\|$ shows that the modulus of all the eigenvalues is 1. Thus, eigenvalues are $e^{\pm i\theta_j}$ or 1. Since a rotation is a normal matrix, it can be diagonalized, and we conclude that every rotation matrix, when expressed in a suitable coordinate system, partitions into $\lfloor k/2 \rfloor$ independent 2D rotations, called Givens rotations [HJ90]:

$$R(\theta_j) = \begin{pmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{pmatrix} = \exp\left(\theta_j \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right).$$

Conversely, each skew symmetric matrix $\Omega = -\Omega^\top$ decomposes the space $\mathbb{R}^k$ in a direct sum of mutually orthogonal subspaces, which are all invariant under Ω [HJ90]. The decomposition has l (possibly equal to zero) 2-dimensional vector subspaces E_j on which Ω acts non trivially, and one single subspace F of dimension $k - 2l$, the orthogonal complement of the span of other subspaces, which is the kernel of Ω. For any E_j, there exists an orthonormal basis of E_j such that Ω restricted to E_j is in this basis of the following matrix form: $\theta_j \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ where θ_j ($\neq 0$) is the jth angle of rotation of the n−dimensional rotation $\exp\Omega$.

We can now come back to the uniqueness of the Log: When the angles of the above $\lfloor k/2 \rfloor$ 2D rotations decomposing the rotation R are within $]-\pi, \pi[$, the logarithm of R is well-defined. Otherwise, one cannot define a unique logarithm. This is only the case for 2D rotations of 180 degrees, whose two 'smallest' real logarithms are the following: $\begin{pmatrix} 0 & -\pi \\ \pi & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & \pi \\ -\pi & 0 \end{pmatrix}$.

Geodesics starting at any point in the group are left, or right, translation of geodesics starting at identity. For instance, $\gamma_{(R,Y)}(t) = R \exp(t\, R^{\top} Y)$ is the unique geodesic starting at R with tangent vector Y. The following reasoning underlies this: To find the geodesic starting at R with tangent vector Y, we first left translate Y by $R^{\top}$ to the tangent space at identity, compute the geodesic starting at e with tangent vector $R^{\top} Y$, and left translate back the result by R. Since the metric is bi-invariant, the same mechanism can be implemented with right translation. The formula for the exponential map $\mathrm{Exp}_R : T_R \mathrm{SO}(k) \mapsto \mathrm{SO}(k)$ at point R is thus:

$$\mathrm{Exp}_R(X) = R\,\mathrm{Exp}_e(R^{\top} X) = R \exp(R^{\top} X) = \exp(X R^{\top})\, R. \qquad (1.20)$$

Likewise, to compute the log map of rotation U at rotation R, we first left translate both rotations by $R^{\top}$, take the log map of $R^{\top} U$ at e, and left translate back to result:

$$\mathrm{Log}_R(U) = R\,\mathrm{Log}_e(R^{\top} U) = R \log(R^{\top} U) = \log(U R^{\top})\, R. \qquad (1.21)$$

1.8. Additional References

This very compact introduction to differential geometry, Riemannian manifolds, and Lie groups provides only a brief overview of the underlying deep theory. We here provide some references to further reading. There are many excellent texts on the subjects. The lists below are therefore naturally non-exhaustive.

Introductory texts on differential and Riemannian geometry

- J. M. Lee: Introduction to topological manifolds [Lee00]; Introduction to smooth manifolds [Lee03]; Riemannian manifolds [Lee97].
- M. do Carmo: Riemannian geometry [dC92].
- J. Gallier: Notes on differential geometry manifolds, Lie groups and bundles, Chapter 3, http://www.cis.upenn.edu/ jean/gbooks/manif.html, [GQ16].
- S. Gallot, D. Hulin, J. Lafontaine; Riemannian geometry [GHL93].
- W. M. Boothby: An Introduction to differentiable manifolds and Riemannian geometry [Boo02].
- C. Small: Statistical theory of shapes [Sma96].

Advanced differential and Riemannian geometry

- J. Jost: Riemannian geometry and geometric analysis [Jos11].
- M. Berger: A panoramic view of Riemannian geometry [Ber03].
- I. Kolář, J. Slovák, P. W. M.: Natural operations in differential geometry [KSM93].
- P. W. Michor: Topics in differential geometry [Mic08].
- M. M. Postnikov: Geometry VI: Riemannian geometry [Pos10].

Lie groups

- S. Helgason: Differential geometry, Lie groups, and symmetric spaces [Hel01].
- J. Gallier: Notes on differential geometry manifolds, Lie groups and bundles, Chapter 2 and 4, `http://www.cis.upenn.edu/~jean/gbooks/manif.html`, [GQ16].

Bibliography

[Ber03] Marcel Berger. *A Panoramic View of Riemannian Geometry*. Springer, Berlin, Heidelberg, 2003.

[Boo02] William M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry, Revised, Volume 120*. Academic Press, Amsterdam ; New York, 2 edition edition, August 2002.

[Car23] Elie Cartan. Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie). *Annales scientifiques de l'École normale supérieure*, 40:325–412, 1923.

[dC92] Manfredo do Carmo. *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin, 1992.

[GHL93] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian Geometry*. Springer Verlag, 2nd edition edition, 1993.

[GQ16] Jean Gallier and Jocelyn Quaintance. *Notes on Differential Geometry and Lie Groups*. 2016.

[Hel01] Sigurdur Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*, volume 34 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. Corrected reprint of the 1978 original.

[HJ90] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, 1990.

[HM94] Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. Communication and Control Engineering Series. Springer, 1994.

[HWL02] Ernst Hairer, Gerhard Wanner, and Christian Lubich. *Geometric numerical integration : structure preserving algorithm for ordinary differential equations*, volume 31 of *Springer series in computational mathematics*. Springer, 2002.

[Jos11] Jürgen Jost. *Riemannian Geometry and Geometric Analysis*. Springer, Heidelberg ; New York, 6th ed. 2011 edition edition, August 2011.

[KAS17] Line Kühnel, Alexis Arnaudon, and Stefan Sommer. Differential geometry and stochastic dynamics with deep learning numerics. *arXiv:1712.08364 [cs, stat]*, December 2017. arXiv: 1712.08364.

[Ken84] David G. Kendall. Shape Manifolds, Procrustean Metrics, and Complex

Projective Spaces. *Bull. London Math. Soc.*, 16(2):81–121, March 1984.

[Kli82] Wilhelm Klingenberg. *Riemannian Geometry*. Walter de Gruyter, Berlin, New York, 1982.

[KSM93] Ivan Kolář, Jan Slovák, and Peter W. Michor. *Natural Operations in Differential Geometry*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.

[Lee97] John M Lee. *Riemannian manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[Lee00] John M Lee. *Introduction to topological manifolds*, volume 202 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.

[Lee03] John M Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003.

[Mic08] Peter W. Michor. *Topics in Differential Geometry*. American Mathematical Soc., 2008.

[MTW73] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973.

[O'N83] Barrett O'Neill. *Semi-Riemannian geometry: with applications to relativity*. Number 103 in Pure and applied mathematics. Academic Press, New York, 1983.

[PFA06] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.

[Pos10] Mikhail Mikhailovich Postnikov. *Geometry VI: Riemannian geometry*. Springer, Berlin, 2010.

[PT97] Xavier Pennec and Jean-Philippe Thirion. A Framework for Uncertainty and Validation of 3D Registration Methods based on Points and Frames. *International Journal of Computer Vision*, 25(3):203–229, 1997.

[Sma96] Christopher Small. *The Statistical Theory of Shape*. Springer, 1996.

[Spi79] Michael David Spivak. *Differential Geometry*, volume 1. Publish or Perish, Inc., 2nd edition, 1979.