Chapter Three

Matrix Manifolds: First-Order Geometry

The constraint sets associated with the examples discussed in Chapter 2 have a particularly rich geometric structure that provides the motivation for this book. The constraint sets are *matrix manifolds* in the sense that they are manifolds in the meaning of classical differential geometry, for which there is a natural representation of elements in the form of matrix arrays.

The matrix representation of the elements is a key property that allows one to provide a natural development of differential geometry in a matrix algebra formulation. The goal of this chapter is to introduce the fundamental concepts in this direction: manifold structure, tangent spaces, cost functions, differentiation, Riemannian metrics, and gradient computation.

There are two classes of matrix manifolds that we consider in detail in this book: embedded submanifolds of $\mathbb{R}^{n\times p}$ and quotient manifolds of $\mathbb{R}^{n\times p}$ (for $1 \leq p \leq n$). Embedded submanifolds are the easiest to understand, as they have the natural form of an explicit constraint set in matrix space $\mathbb{R}^{n\times p}$. The case we will be mostly interested in is the set of orthonormal $n \times p$ matrices that, as will be shown, can be viewed as an embedded submanifold of $\mathbb{R}^{n\times p}$ called the Stiefel manifold $\mathrm{St}(p, n)$. In particular, for $p = 1$, the Stiefel manifold reduces to the unit sphere S^{n-1}, and for $p = n$, it reduces to the set of orthogonal matrices $O(n)$.

Quotient spaces are more difficult to visualize, as they are not defined as sets of matrices; rather, each point of the quotient space is an equivalence class of $n \times p$ matrices. In practice, an example $n \times p$ matrix from a given equivalence class is used to represent an element of matrix quotient space in computer memory and in our numerical development. The calculations related to the geometric structure of a matrix quotient manifold can be expressed directly using the tools of matrix algebra on these representative matrices.

The focus of this first geometric chapter is on the concepts from differential geometry that are required to generalize the steepest-descent method, arguably the simplest approach to unconstrained optimization. In $\mathbb{R}^n$, the steepest-descent algorithm updates a current iterate x in the direction where the first-order decrease of the cost function f is most negative. Formally, the update direction is chosen to be the unit norm vector η that minimizes the directional derivative

$$\mathrm{D} f\,(x)\,[\eta] = \lim_{t\to 0} \frac{f(x + t\eta) - f(x)}{t}. \tag{3.1}$$

When the domain of f is a manifold $\mathcal{M}$, the argument $x + t\eta$ in (3.1) does not make sense in general since $\mathcal{M}$ is not necessarily a vector space. This leads to the important concept of a tangent vector (Section 3.5). In order to define the notion of a steepest-descent direction, it will then remain to define the *length* of a tangent vector, a task carried out in Section 3.6 where the concept of a Riemannian manifold is introduced. This leads to a definition of the gradient of a function, the generalization of steepest-descent direction on a Riemannian manifold.

3.1 MANIFOLDS

We define the notion of a manifold in its full generality; then we consider the simple but important case of linear manifolds, a linear vector space interpreted as a manifold with Euclidean geometric structure. The manifold of $n \times p$ real matrices, from which all concrete examples in this book originate, is a linear manifold.

A d-dimensional manifold can be informally defined as a set $\mathcal{M}$ covered with a "suitable" collection of coordinate patches, or charts, that identify certain subsets of $\mathcal{M}$ with open subsets of $\mathbb{R}^d$. Such a collection of coordinate charts can be thought of as the basic structure required to do differential calculus on $\mathcal{M}$.

It is often cumbersome or impractical to use coordinate charts to (locally) turn computational problems on $\mathcal{M}$ into computational problems on $\mathbb{R}^d$. The numerical algorithms developed later in this book rely on exploiting the natural matrix structure of the manifolds associated with the examples of interest, rather than imposing a local $\mathbb{R}^d$ structure. Nevertheless, coordinate charts are an essential tool for addressing fundamental notions such as the differentiability of a function on a manifold.

3.1.1 Definitions: charts, atlases, manifolds

The abstract definition of a manifold relies on the concepts of charts and atlases.

Let $\mathcal{M}$ be a set. A bijection (one-to-one correspondence) φ of a subset $\mathcal{U}$ of $\mathcal{M}$ onto an open subset of $\mathbb{R}^d$ is called a d-dimensional *chart of the set* $\mathcal{M}$, denoted by $(\mathcal{U}, \varphi)$. When there is no risk of confusion, we will simply write φ for $(\mathcal{U}, \varphi)$. Given a chart $(\mathcal{U}, \varphi)$ and $x \in \mathcal{U}$, the elements of $\varphi(x) \in \mathbb{R}^d$ are called the *coordinates* of x in the chart $(\mathcal{U}, \varphi)$.

The interest of the notion of chart $(\mathcal{U}, \varphi)$ is that it makes it possible to study objects associated with $\mathcal{U}$ by bringing them to the subset $\varphi(\mathcal{U})$ of $\mathbb{R}^d$. For example, if f is a real-valued function on $\mathcal{U}$, then $f \circ \varphi^{-1}$ is a function from $\mathbb{R}^d$ to $\mathbb{R}$, with domain $\varphi(\mathcal{U})$, to which methods of real analysis apply. To take advantage of this idea, we must require that each point of the set $\mathcal{M}$ be at least in one chart domain; moreover, if a point x belongs to the domains of two charts $(\mathcal{U}_1, \varphi_1)$ and $(\mathcal{U}_2, \varphi_2)$, then the two charts must give compatible information: for example, if a real-valued function f is defined

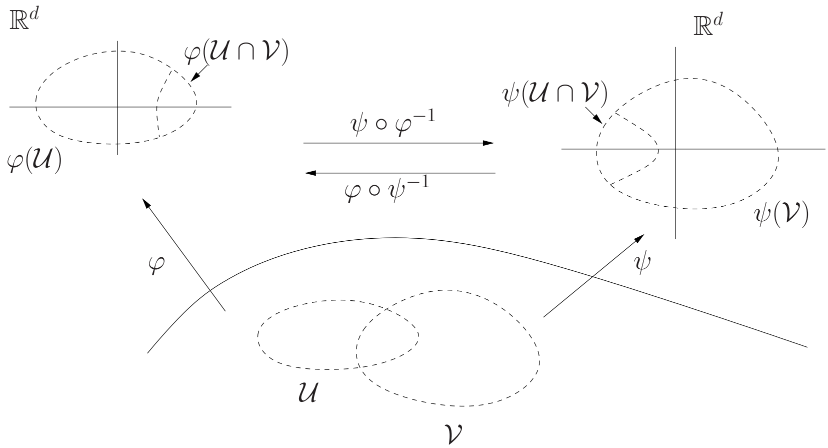


Figure 3.1 Charts.

on $\mathcal{U}_1 \cap \mathcal{U}_2$, then $f \circ \varphi_1^{-1}$ and $f \circ \varphi_2^{-1}$ should have the same differentiability properties on $\mathcal{U}_1 \cap \mathcal{U}_2$.

The following concept takes these requirements into account. A (C^∞) *atlas of $\mathcal{M}$ into $\mathbb{R}^d$* is a collection of charts $(\mathcal{U}_\alpha, \varphi_\alpha)$ of the set $\mathcal{M}$ such that

1. $\bigcup_\alpha \mathcal{U}_\alpha = \mathcal{M}$,
2. for any pair α, β with $\mathcal{U}_\alpha \cap \mathcal{U}_\beta \neq \emptyset$, the sets $\varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$ and $\varphi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$ are open sets in $\mathbb{R}^d$ and the change of coordinates
$$\varphi_\beta \circ \varphi_\alpha^{-1} : \mathbb{R}^d \to \mathbb{R}^d$$
(see Appendix A.3 for our conventions on functions) is *smooth* (class C^∞, i.e., differentiable for all degrees of differentiation) on its domain $\varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$; see illustration in Figure 3.1. We say that the elements of an atlas *overlap smoothly.*

Two atlases $\mathcal{A}_1$ and $\mathcal{A}_2$ are *equivalent* if $\mathcal{A}_1 \cup \mathcal{A}_2$ is an atlas; in other words, for every chart $(\mathcal{U}, \varphi)$ in $\mathcal{A}_2$, the set of charts $\mathcal{A}_1 \cup \{(\mathcal{U}, \varphi)\}$ is still an atlas. Given an atlas $\mathcal{A}$, let $\mathcal{A}^+$ be the set of all charts $(\mathcal{U}, \varphi)$ such that $\mathcal{A} \cup \{(\mathcal{U}, \varphi)\}$ is also an atlas. It is easy to see that $\mathcal{A}^+$ is also an atlas, called the *maximal atlas* (or *complete atlas*) generated by the atlas $\mathcal{A}$. Two atlases are equivalent if and only if they generate the same maximal atlas. A maximal atlas of a set $\mathcal{M}$ is also called a *differentiable structure* on $\mathcal{M}$.

In the literature, a manifold is sometimes simply defined as a set endowed with a differentiable structure. However, this definition does not exclude certain unconventional topologies. For example, it does not guarantee that convergent sequences have a single limit point (an example is given in Section 4.3.2). To avoid such counterintuitive situations, we adopt the following classical definition. A *(d-dimensional) manifold* is a couple $(\mathcal{M}, \mathcal{A}^+)$, where $\mathcal{M}$ is a set and $\mathcal{A}^+$ is a maximal atlas of $\mathcal{M}$ into $\mathbb{R}^d$, such that the topology induced by $\mathcal{A}^+$ is Hausdorff and second-countable. (These topological issues are discussed in Section 3.1.2.)

A maximal atlas of a set $\mathcal{M}$ that induces a second-countable Hausdorff topology is called a *manifold structure* on $\mathcal{M}$. Often, when $(\mathcal{M}, \mathcal{A}^+)$ is a manifold, we simply say "the manifold $\mathcal{M}$" when the differentiable structure is clear from the context, and we say "the set $\mathcal{M}$" to refer to $\mathcal{M}$ as a plain set without a particular differentiable structure. Note that it is not necessary to specify the whole maximal atlas to define a manifold structure: it is enough to provide an atlas that generates the manifold structure.

Given a manifold $(\mathcal{M}, \mathcal{A}^+)$, an atlas of the set $\mathcal{M}$ whose maximal atlas is $\mathcal{A}^+$ is called an *atlas of the manifold* $(\mathcal{M}, \mathcal{A}^+)$; a chart of the set $\mathcal{M}$ that belongs to $\mathcal{A}^+$ is called a *chart of the manifold* $(\mathcal{M}, \mathcal{A}^+)$, and its domain is a *coordinate domain* of the manifold. By a chart around a point $x \in \mathcal{M}$, we mean a chart of $(\mathcal{M}, \mathcal{A}^+)$ whose domain $\mathcal{U}$ contains x. The set $\mathcal{U}$ is then a *coordinate neighborhood* of x.

Given a chart φ on $\mathcal{M}$, the inverse mapping φ^{-1} is called a *local parameterization* of $\mathcal{M}$. A family of local parameterizations is equivalent to a family of charts, and the definition of a manifold may be given in terms of either.

3.1.2 The topology of a manifold*

Recall that the star in the section title indicates material that can be readily skipped at a first reading.

It can be shown that the collection of coordinate domains specified by a maximal atlas $\mathcal{A}^+$ of a set $\mathcal{M}$ forms a basis for a topology of the set $\mathcal{M}$. (We refer the reader to Section A.2 for a short introduction to topology.) We call this topology the *atlas topology* of $\mathcal{M}$ induced by $\mathcal{A}$. In the atlas topology, a subset $\mathcal{V}$ of $\mathcal{M}$ is open if and only if, for any chart $(\mathcal{U}, \varphi)$ in $\mathcal{A}^+$, $\varphi(\mathcal{V} \cap \mathcal{U})$ is an open subset of $\mathbb{R}^d$. Equivalently, a subset $\mathcal{V}$ of $\mathcal{M}$ is open if and only if, for each $x \in \mathcal{V}$, there is a chart $(\mathcal{U}, \varphi)$ in $\mathcal{A}^+$ such that $x \in \mathcal{U} \subset \mathcal{V}$. An atlas $\mathcal{A}$ of a set $\mathcal{M}$ is said to be *compatible* with a topology $\mathcal{T}$ on the set $\mathcal{M}$ if the atlas topology is equal to $\mathcal{T}$.

An atlas topology always satisfies separation axiom T_1, i.e., given any two distinct points x and y, there is an open set $\mathcal{U}$ that contains x and not y. (Equivalently, every singleton is a closed set.) But not all atlas topologies are *Hausdorff* (i.e., T_2): two distinct points do not necessarily have disjoint neighborhoods. Non-Hausdorff spaces can display unusual and counterintuitive behavior. From the perspective of numerical iterative algorithms the most worrying possibility is that a convergent sequence on a non-Hausdorff topological space may have several distinct limit points. Our definition of manifold rules out non-Hausdorff topologies.

A topological space is *second-countable* if there is a countable collection $\mathcal{B}$ of open sets such that every open set is the union of some subcollection of $\mathcal{B}$. Second-countability is related to *partitions of unity*, a crucial tool in resolving certain fundamental questions such as the existence of a Riemannian metric (Section 3.6) and the existence of an affine connection (Section 5.2).

The existence of partitions of unity subordinate to arbitrary open coverings is equivalent to the property of *paracompactness*. A set endowed with a Hausdorff atlas topology is paracompact (and has countably many components) if (and only if) it is second-countable. Since manifolds are assumed to be Hausdorff and second-countable, they admit partitions of unity.

For a manifold $(\mathcal{M}, \mathcal{A}^+)$, we refer to the atlas topology of $\mathcal{M}$ induced by $\mathcal{A}$ as the *manifold topology of* $\mathcal{M}$. Note that several statements in this book also hold without the Hausdorff and second-countable assumptions. These cases, however, are of marginal importance and will not be discussed.

Given a manifold $(\mathcal{M}, \mathcal{A}^+)$ and an open subset $\mathcal{X}$ of $\mathcal{M}$ (open is to be understood in terms of the manifold topology of $\mathcal{M}$), the collection of the charts of $(\mathcal{M}, \mathcal{A}^+)$ whose domain lies in $\mathcal{X}$ forms an atlas of $\mathcal{X}$. This defines a differentiable structure on $\mathcal{X}$ of the same dimension as $\mathcal{M}$. With this structure, $\mathcal{X}$ is called an *open submanifold* of $\mathcal{M}$.

A manifold is *connected* if it cannot be expressed as the disjoint union of two nonempty open sets. Equivalently (for a manifold), any two points can be joined by a piecewise smooth curve segment. The connected components of a manifold are open, thus they admit a natural differentiable structure as open submanifolds. The optimization algorithms considered in this book are iterative and oblivious to the existence of connected components other than the one to which the current iterate belongs. Therefore we have no interest in considering manifolds that are not connected.

3.1.3 How to recognize a manifold

Assume that a computational problem involves a search space $\mathcal{X}$. How can we check that $\mathcal{X}$ is a manifold? It should be clear from Section 3.1.1 that this question is not well posed: by definition, a manifold is not simply a set $\mathcal{X}$ but rather a couple $(\mathcal{X}, \mathcal{A}^+)$ where $\mathcal{X}$ is a set and $\mathcal{A}^+$ is a maximal atlas of $\mathcal{X}$ inducing a second-countable Hausdorff topology.

A well-posed question is to ask whether a given set $\mathcal{X}$ admits an atlas. There are sets that do not admit an atlas and thus cannot be turned into a manifold. A simple example is the set of rational numbers: this set does not even admit charts; otherwise, it would not be denumerable. Nevertheless, sets abound that admit an atlas. Even sets that do not "look" differentiable may admit an atlas. For example, consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2 : \gamma(t) = (t, |t|)$ and let $\mathcal{X}$ be the range of γ; see Figure 3.2. Consider the chart $\varphi : \mathcal{X} \to \mathbb{R} : (t, |t|) \mapsto t$. It turns out that $\mathcal{A} := \{(\mathcal{X}, \varphi)\}$ is an atlas of the set $\mathcal{X}$; therefore, $(\mathcal{X}, \mathcal{A}^+)$ is a manifold. The incorrect intuition that $\mathcal{X}$ cannot be a manifold because of its "corner" corresponds to the fact that $\mathcal{X}$ is not a *submanifold* of $\mathbb{R}^2$; see Section 3.3.

A set $\mathcal{X}$ may admit more than one maximal atlas. As an example, take the set $\mathbb{R}$ and consider the charts $\varphi_1 : x \mapsto x$ and $\varphi_2 : x \mapsto x^3$. Note that φ_1 and φ_2 are not compatible since the mapping $\varphi_1 \circ \varphi_2^{-1}$ is not differentiable at the origin. However, each chart individually forms an atlas of the set $\mathbb{R}$. These two atlases are not equivalent; they do not generate the same maximal

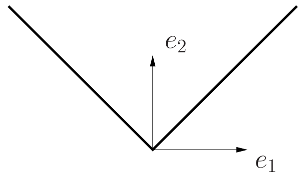


Figure 3.2 Image of the curve $\gamma : t \mapsto (t, |t|)$.

atlas. Nevertheless, the chart $x \mapsto x$ is clearly more natural than the chart $x \mapsto x^3$. Most manifolds of interest admit a differentiable structure that is the most "natural"; see in particular the notions of embedded and quotient matrix manifold in Sections 3.3 and 3.4.

3.1.4 Vector spaces as manifolds

Let $\mathcal{E}$ be a d-dimensional vector space. Then, given a basis $(e_i)_{i=1,\ldots,d}$ of $\mathcal{E}$, the function

$$\psi : \mathcal{E} \to \mathbb{R}^d : x \mapsto \begin{bmatrix} x^1 \\ \vdots \\ x^d \end{bmatrix}$$

such that $x = \sum_{i=1}^d x^i e_i$ is a chart of the set $\mathcal{E}$. All charts built in this way are compatible; thus they form an atlas of the set $\mathcal{E}$, which endows $\mathcal{E}$ with a manifold structure. Hence, every vector space is a *linear manifold* in a natural way.

Needless to say, the challenging case is the one where the manifold structure is *nonlinear*, i.e., manifolds that are not endowed with a vector space structure. The numerical algorithms considered in this book apply equally to linear and nonlinear manifolds and reduce to classical optimization algorithms when the manifold is linear.

3.1.5 The manifolds $\mathbb{R}^{n\times p}$ and $\mathbb{R}_*^{n\times p}$

Algorithms formulated on abstract manifolds are not strictly speaking numerical algorithms in the sense that they involve manipulation of differential-geometric objects instead of numerical calculations. Turning these abstract algorithms into numerical algorithms for specific optimization problems relies crucially on producing adequate numerical representations of the geometric objects that arise in the abstract algorithms. A significant part of this book is dedicated to building a toolbox of results that make it possible to perform this "geometric-to-numerical" conversion on matrix manifolds (i.e., manifolds obtained by taking embedded submanifolds and quotient manifolds of $\mathbb{R}^{n\times p}$). The process derives from the manifold structure of the set $\mathbb{R}^{n\times p}$ of $n \times p$ real matrices, discussed next.

The set $\mathbb{R}^{n\times p}$ is a vector space with the usual sum and multiplication by a scalar. Consequently, it has a natural linear manifold structure. A chart of this manifold is given by $\varphi : \mathbb{R}^{n\times p} \to \mathbb{R}^{np} : X \mapsto \mathrm{vec}(X)$, where $\mathrm{vec}(X)$ denotes the vector obtained by stacking the columns of X below one another. We will refer to the set $\mathbb{R}^{n\times p}$ with its linear manifold structure as the *manifold* $\mathbb{R}^{n\times p}$. Its dimension is np.

The manifold $\mathbb{R}^{n\times p}$ can be further turned into a Euclidean space with the inner product

$$\langle Z_1, Z_2\rangle := \mathrm{vec}(Z_1)^T \mathrm{vec}(Z_2) = \mathrm{tr}(Z_1^T Z_2). \tag{3.2}$$

The norm induced by the inner product is the *Frobenius norm* defined by

$$\|Z\|_F^2 = \mathrm{tr}(Z^T Z),$$

i.e., $\|Z\|_F^2$ is the sum of the squares of the elements of Z. Observe that the manifold topology of $\mathbb{R}^{n\times p}$ is equivalent to its canonical topology as a Euclidean space (see Appendix A.2).

Let $\mathbb{R}_*^{n\times p}$ $(p \leq n)$ denote the set of all $n \times p$ matrices whose columns are linearly independent. This set is an open subset of $\mathbb{R}^{n\times p}$ since its complement $\{X \in \mathbb{R}^{n\times p} : \det(X^T X) = 0\}$ is closed. Consequently, it admits a structure of an open submanifold of $\mathbb{R}^{n\times p}$. Its differentiable structure is generated by the chart $\varphi : \mathbb{R}_*^{n\times p} \to \mathbb{R}^{np} : X \mapsto \mathrm{vec}(X)$. This manifold will be referred to as the *manifold* $\mathbb{R}_*^{n\times p}$, or the *noncompact Stiefel manifold* of full-rank $n \times p$ matrices.

In the particular case $p = 1$, the noncompact Stiefel manifold reduces to the Euclidean space $\mathbb{R}^n$ with the origin removed. When $p = n$, the noncompact Stiefel manifold becomes the general linear group GL_n, i.e., the set of all invertible $n \times n$ matrices.

Notice that the chart $\mathrm{vec} : \mathbb{R}^{n\times p} \to \mathbb{R}^{np}$ is unwieldy, as it destroys the matrix structure of its argument; in particular, $\mathrm{vec}(AB)$ cannot be written as a simple expression of $\mathrm{vec}(A)$ and $\mathrm{vec}(B)$. In this book, the emphasis is on preserving the matrix structure.

3.1.6 Product manifolds

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be manifolds of dimension d_1 and d_2, respectively. The set $\mathcal{M}_1 \times \mathcal{M}_2$ is defined as the set of pairs (x_1, x_2), where x_1 is in $\mathcal{M}_1$ and x_2 is in $\mathcal{M}_2$. If $(\mathcal{U}_1, \varphi_1)$ and $(\mathcal{U}_2, \varphi_2)$ are charts of the manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively, then the mapping $\varphi_1 \times \varphi_2 : \mathcal{U}_1 \times \mathcal{U}_2 \to \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} : (x_1, x_2) \mapsto (\varphi_1(x_1), \varphi_2(x_2))$ is a chart for the set $\mathcal{M}_1 \times \mathcal{M}_2$. All the charts thus obtained form an atlas for the set $\mathcal{M}_1 \times \mathcal{M}_2$. With the differentiable structure defined by this atlas, $\mathcal{M}_1 \times \mathcal{M}_2$ is called the *product* of the manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$. Its manifold topology is equivalent to the product topology. Product manifolds will be useful in some later developments.

3.2 DIFFERENTIABLE FUNCTIONS

Mappings between manifolds appear in many places in optimization algorithms on manifolds. First of all, any optimization problem on a manifold $\mathcal{M}$ involves a cost function, which can be viewed as a mapping from the manifold $\mathcal{M}$ into the manifold $\mathbb{R}$. Other instances of mappings between manifolds are inclusions (in the theory of submanifolds; see Section 3.3), natural projections onto quotients (in the theory of quotient manifolds, see Section 3.4), and retractions (a fundamental tool in numerical algorithms on manifolds; see Section 4.1). This section introduces the notion of differentiability for functions between manifolds. The coordinate-free definition of a differential will come later, as it requires the concept of a tangent vector.

Let F be a function from a manifold $\mathcal{M}_1$ of dimension d_1 into another manifold $\mathcal{M}_2$ of dimension d_2. Let x be a point of $\mathcal{M}_1$. Choosing charts φ_1 and φ_2 around x and $F(x)$, respectively, the function F around x can be "read through the charts", yielding the function

$$\hat{F} = \varphi_2 \circ F \circ \varphi_1^{-1} : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}, \tag{3.3}$$

called a *coordinate representation* of F. (Note that the domain of $\hat{F}$ is in general a subset of $\mathbb{R}^{d_1}$; see Appendix A.3 for the conventions.)

We say that F is *differentiable* or *smooth* at x if $\hat{F}$ is of class C^∞ at $\varphi_1(x)$. It is easily verified that this definition does not depend on the choice of the charts chosen at x and $F(x)$. A function $F : \mathcal{M}_1 \to \mathcal{M}_2$ is said to be *smooth* if it is smooth at every point of its domain.

A (smooth) *diffeomorphism* $F : \mathcal{M}_1 \to \mathcal{M}_2$ is a bijection such that F and its inverse F^{-1} are both smooth. Two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ are said to be *diffeomorphic* if there exists a diffeomorphism on $\mathcal{M}_1$ onto $\mathcal{M}_2$.

In this book, all functions are assumed to be smooth unless otherwise stated.

3.2.1 Immersions and submersions

The concepts of immersion and submersion will make it possible to define submanifolds and quotient manifolds in a concise way. Let $F : \mathcal{M}_1 \to \mathcal{M}_2$ be a differentiable function from a manifold $\mathcal{M}_1$ of dimension d_1 into a manifold $\mathcal{M}_2$ of dimension d_2. Given a point x of $\mathcal{M}_1$, the *rank* of F at x is the dimension of the range of $\mathrm{D}\hat{F}\,(\varphi_1(x))\,[\cdot] : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, where $\hat{F}$ is a coordinate representation (3.3) of F around x, and $\mathrm{D}\hat{F}(\varphi_1(x))$ denotes the *differential* of $\hat{F}$ at $\varphi_1(x)$ (see Section A.5). (Notice that this definition does not depend on the charts used to obtain the coordinate representation $\hat{F}$ of F.) The function F is called an *immersion* if its rank is equal to d_1 at each point of its domain (hence $d_1 \leq d_2$). If its rank is equal to d_2 at each point of its domain (hence $d_1 \geq d_2$), then it is called a *submersion*.

The function F is an immersion if and only if, around each point of its domain, it admits a coordinate representation that is the *canonical immersion* $(u^1, \ldots, u^{d_1}) \mapsto (u^1, \ldots, u^{d_1}, 0, \ldots, 0)$. The function F is a submersion if and only if, around each point of its domain, it admits the *canonical submersion*

$(u^1, \ldots, u^{d_1}) \mapsto (u^1, \ldots, u^{d_2})$ as a coordinate representation. A point $y \in \mathcal{M}_2$ is called a *regular value* of F if the rank of F is d_2 at every $x \in F^{-1}(y)$.

3.3 EMBEDDED SUBMANIFOLDS

A set $\mathcal{X}$ may admit several manifold structures. However, if the set $\mathcal{X}$ is a subset of a manifold $(\mathcal{M}, \mathcal{A}^+)$, then it admits at most one submanifold structure. This is the topic of this section.

3.3.1 General theory

Let $(\mathcal{M}, \mathcal{A}^+)$ and $(\mathcal{N}, \mathcal{B}^+)$ be manifolds such that $\mathcal{N} \subset \mathcal{M}$. The manifold $(\mathcal{N}, \mathcal{B}^+)$ is called an *immersed submanifold* of $(\mathcal{M}, \mathcal{A}^+)$ if the inclusion map $i : \mathcal{N} \to \mathcal{M} : x \mapsto x$ is an immersion.

Let $(\mathcal{N}, \mathcal{B}^+)$ be a submanifold of $(\mathcal{M}, \mathcal{A}^+)$. Since $\mathcal{M}$ and $\mathcal{N}$ are manifolds, they are also topological spaces with their manifold topology. If the manifold topology of $\mathcal{N}$ coincides with its subspace topology induced from the topological space $\mathcal{M}$, then $\mathcal{N}$ is called an *embedded submanifold*, a *regular submanifold*, or simply a *submanifold* of the manifold $\mathcal{M}$. Asking that a subset $\mathcal{N}$ of a manifold $\mathcal{M}$ be an embedded submanifold of $\mathcal{M}$ removes all freedom for the choice of a differentiable structure on $\mathcal{N}$:

Proposition 3.3.1 *Let $\mathcal{N}$ be a subset of a manifold $\mathcal{M}$. Then $\mathcal{N}$ admits at most one differentiable structure that makes it an embedded submanifold of $\mathcal{M}$.*

As a consequence of Proposition 3.3.1, when we say in this book that a subset of a manifold "is" a submanifold, we mean that it admits one (unique) differentiable structure that makes it an embedded submanifold. The manifold $\mathcal{M}$ in Proposition 3.3.1 is called the *embedding space*. When the embedding space is $\mathbb{R}^{n \times p}$ or an open subset of $\mathbb{R}^{n \times p}$, we say that $\mathcal{N}$ is a *matrix submanifold*.

To check whether a subset $\mathcal{N}$ of a manifold $\mathcal{M}$ is an embedded submanifold of $\mathcal{M}$ and to construct an atlas of that differentiable structure, one can use the next proposition, which states that every embedded submanifold is locally a coordinate slice. Given a chart $(\mathcal{U}, \varphi)$ of a manifold $\mathcal{M}$, a *φ-coordinate slice* of $\mathcal{U}$ is a set of the form $\varphi^{-1}(\mathbb{R}^m \times \{0\})$ that corresponds to all the points of $\mathcal{U}$ whose last $n - m$ coordinates in the chart φ are equal to zero.

Proposition 3.3.2 (submanifold property) *A subset $\mathcal{N}$ of a manifold $\mathcal{M}$ is a d-dimensional embedded submanifold of $\mathcal{M}$ if and only if, around each point $x \in \mathcal{N}$, there exists a chart $(\mathcal{U}, \varphi)$ of $\mathcal{M}$ such that $\mathcal{N} \cap \mathcal{U}$ is a φ-coordinate slice of $\mathcal{U}$, i.e.,*

$$\mathcal{N} \cap \mathcal{U} = \{x \in \mathcal{U} : \varphi(x) \in \mathbb{R}^d \times \{0\}\}.$$

In this case, the chart $(\mathcal{N} \cap \mathcal{U}, \varphi)$, where φ is seen as a mapping into $\mathbb{R}^d$, is a chart of the embedded submanifold $\mathcal{N}$.

The next propositions provide sufficient conditions for subsets of manifolds to be embedded submanifolds.

Proposition 3.3.3 (submersion theorem) *Let $F : \mathcal{M}_1 \to \mathcal{M}_2$ be a smooth mapping between two manifolds of dimension d_1 and d_2, $d_1 > d_2$, and let y be a point of $\mathcal{M}_2$. If y is a regular value of F (i.e., the rank of F is equal to d_2 at every point of $F^{-1}(y)$), then $F^{-1}(y)$ is a closed embedded submanifold of $\mathcal{M}_1$, and $\dim(F^{-1}(y)) = d_1 - d_2$.*

Proposition 3.3.4 (subimmersion theorem) *Let $F : \mathcal{M}_1 \to \mathcal{M}_2$ be a smooth mapping between two manifolds of dimension d_1 and d_2 and let y be a point of $F(\mathcal{M}_1)$. If F has constant rank $k < d_1$ in a neighborhood of $F^{-1}(y)$, then $F^{-1}(y)$ is a closed embedded submanifold of $\mathcal{M}_1$ of dimension $d_1 - k$.*

Functions on embedded submanifolds pose no particular difficulty. Let $\mathcal{N}$ be an embedded submanifold of a manifold $\mathcal{M}$. If f is a smooth function on $\mathcal{M}$, then $f|_{\mathcal{N}}$, the *restriction* of f to $\mathcal{N}$, is a smooth function on $\mathcal{N}$. Conversely, any smooth function on $\mathcal{N}$ can be written locally as a restriction of a smooth function defined on an open subset $\mathcal{U} \subset \mathcal{M}$.

3.3.2 The Stiefel manifold

The (orthogonal) Stiefel manifold is an embedded submanifold of $\mathbb{R}^{n\times p}$ that will appear frequently in our practical examples.

Let $\mathrm{St}(p, n)$ ($p \le n$) denote the set of all $n \times p$ orthonormal matrices; i.e.,

$$\mathrm{St}(p, n) := \{X \in \mathbb{R}^{n\times p} : X^T X = I_p\}, \tag{3.4}$$

where I_p denotes the $p \times p$ identity matrix. The set $\mathrm{St}(p, n)$ (endowed with its submanifold structure as discussed below) is called an *(orthogonal or compact) Stiefel manifold*. Note that the Stiefel manifold $\mathrm{St}(p, n)$ is distinct from the noncompact Stiefel manifold $\mathbb{R}^{n\times p}_*$ defined in Section 3.1.5.

Clearly, $\mathrm{St}(p, n)$ is a subset of the set $\mathbb{R}^{n\times p}$. Recall that the set $\mathbb{R}^{n\times p}$ admits a linear manifold structure as described in Section 3.1.5. To show that $\mathrm{St}(p, n)$ is an embedded submanifold of the manifold $\mathbb{R}^{n\times p}$, consider the function $F : \mathbb{R}^{n\times p} \to \mathcal{S}_{\mathrm{sym}}(p) : X \mapsto X^T X - I_p$, where $\mathcal{S}_{\mathrm{sym}}(p)$ denotes the set of all symmetric $p \times p$ matrices. Note that $\mathcal{S}_{\mathrm{sym}}(p)$ is a vector space. Clearly, $\mathrm{St}(p, n) = F^{-1}(0_p)$. It remains to show that F is a submersion at each point X of $\mathrm{St}(p, n)$. The fact that the domain of F is a vector space exempts us from having to read F through a chart: we simply need to show that for all $\widehat{Z}$ in $\mathcal{S}_{\mathrm{sym}}(p)$, there exists Z in $\mathbb{R}^{n\times p}$ such that $\mathrm{D}F(X)[Z] = \widehat{Z}$. We have (see Appendix A.5 for details on matrix differentiation)

$$\mathrm{D}F(X)[Z] = X^T Z + Z^T X.$$

It is easy to see that $\mathrm{D}F(X)\left[\frac{1}{2}X\widehat{Z}\right] = \widehat{Z}$ since $X^T X = I_p$ and $\widehat{Z}^T = \widehat{Z}$. This shows that F is full rank. It follows from Proposition 3.3.3 that the set $\mathrm{St}(p, n)$ defined in (3.4) is an embedded submanifold of $\mathbb{R}^{n\times p}$.

To obtain the dimension of $\mathrm{St}(p,n)$, observe that the vector space $\mathcal{S}_{\mathrm{sym}}(p)$ has dimension $\frac{1}{2}p(p+1)$ since a symmetric matrix is completely determined by its upper triangular part (including the diagonal). From Proposition 3.3.3, we obtain

$$\dim(\mathrm{St}(p,n)) = np - \tfrac{1}{2}p(p+1).$$

Since $\mathrm{St}(p,n)$ is an embedded submanifold of $\mathbb{R}^{n\times p}$, its topology is the subset topology induced by $\mathbb{R}^{n\times p}$. The manifold $\mathrm{St}(p,n)$ is closed: it is the inverse image of the closed set $\{0_p\}$ under the continuous function $F : \mathbb{R}^{n\times p} \mapsto \mathcal{S}_{\mathrm{sym}}(p)$. It is bounded: each column of $X \in \mathrm{St}(p,n)$ has norm 1, so the Frobenius norm of X is equal to $\sqrt{p}$. It then follows from the Heine-Borel theorem (see Section A.2) that the manifold $\mathrm{St}(p,n)$ is *compact*.

For $p = 1$, the Stiefel manifold $\mathrm{St}(p,n)$ reduces to the *unit sphere* S^{n-1} in $\mathbb{R}^n$. Notice that the superscript $n-1$ indicates the dimension of the manifold.

For $p = n$, the Stiefel manifold $\mathrm{St}(p,n)$ becomes the *orthogonal group* O_n. Its dimension is $\frac{1}{2}n(n-1)$.

3.4 QUOTIENT MANIFOLDS

Whereas the topic of submanifolds is covered in any introductory textbook on manifolds, the subject of quotient manifolds is less classical. We develop the theory in some detail because it has several applications in matrix computations, most notably in algorithms that involve subspaces of $\mathbb{R}^n$. Computations involving subspaces are usually carried out using matrices to represent the corresponding subspace generated by the span of its columns. The difficulty is that for one given subspace, there are infinitely many matrices that represent the subspace. It is then desirable to partition the set of matrices into classes of "equivalent" elements that represent the same object. This leads to the concept of quotient spaces and quotient manifolds. In this section, we first present the general theory of quotient manifolds, then we return to the special case of subspaces and their representations.

3.4.1 Theory of quotient manifolds

Let $\mathcal{M}$ be a manifold equipped with an *equivalence relation* $\sim$, i.e., a relation that is

1. reflexive: $x \sim x$ for all $x \in \mathcal{M}$,
2. symmetric: $x \sim y$ if and only if $y \sim x$ for all $x, y \in \mathcal{M}$,
3. transitive: if $x \sim y$ and $y \sim z$ then $x \sim z$ for all $x, y, z \in \mathcal{M}$.

The set

$$[x] := \{y \in \mathcal{M} : y \sim x\}$$

of all elements that are equivalent to a point x is called the *equivalence class* containing x. The set

$$\mathcal{M}/\sim := \{[x] : x \in \mathcal{M}\}$$

of all equivalence classes of $\sim$ in $\mathcal{M}$ is called the *quotient* of $\mathcal{M}$ by $\sim$. Notice that the points of $\mathcal{M}/\sim$ are subsets of $\mathcal{M}$. The mapping $\pi : \mathcal{M} \to \mathcal{M}/\sim$ defined by $x \mapsto [x]$ is called the *natural projection* or *canonical projection*. Clearly, $\pi(x) = \pi(y)$ if and only if $x \sim y$, so we have $[x] = \pi^{-1}(\pi(x))$. We will use $\pi(x)$ to denote $[x]$ viewed as a point of $\mathcal{M}/\sim$, and $\pi^{-1}(\pi(x))$ for $[x]$ viewed as a subset of $\mathcal{M}$. The set $\mathcal{M}$ is called the *total space* of the quotient $\mathcal{M}/\sim$.

Let $(\mathcal{M}, \mathcal{A}^+)$ be a manifold with an equivalence relation $\sim$ and let $\mathcal{B}^+$ be a manifold structure on the set $\mathcal{M}/\sim$. The manifold $(\mathcal{M}/\sim, \mathcal{B}^+)$ is called a *quotient manifold* of $(\mathcal{M}, \mathcal{A}^+)$ if the natural projection π is a submersion.

Proposition 3.4.1 *Let $\mathcal{M}$ be a manifold and let $\mathcal{M}/\sim$ be a quotient of $\mathcal{M}$. Then $\mathcal{M}/\sim$ admits at most one manifold structure that makes it a quotient manifold of $\mathcal{M}$.*

Given a quotient $\mathcal{M}/\sim$ of a manifold $\mathcal{M}$, we say that the set $\mathcal{M}/\sim$ *is* a quotient manifold if it admits a (unique) quotient manifold structure. In this case, we say that the equivalence relation $\sim$ is *regular*, and we refer to the set $\mathcal{M}/\sim$ endowed with this manifold structure as the manifold $\mathcal{M}/\sim$.

The following result gives a characterization of regular equivalence relations. Note that the *graph* of a relation $\sim$ is the set

$$\mathrm{graph}(\sim) := \{(x, y) \in \mathcal{M} \times \mathcal{M} : x \sim y\}.$$

Proposition 3.4.2 *An equivalence relation $\sim$ on a manifold $\mathcal{M}$ is regular (and thus $\mathcal{M}/\sim$ is a quotient manifold) if and only if the following conditions hold together:*

(i) The graph of $\sim$ is an embedded submanifold of the product manifold $\mathcal{M} \times \mathcal{M}$.
(ii) The projection $\pi_1 : \mathrm{graph}(\sim) \to \mathcal{M}$, $\pi_1(x, y) = x$ is a submersion.
(iii) The graph of $\sim$ is a closed subset of $\mathcal{M} \times \mathcal{M}$ (where $\mathcal{M}$ is endowed with its manifold topology).

The dimension of $\mathcal{M}/\sim$ is given by

$$\dim(\mathcal{M}/\sim) = 2\dim(\mathcal{M}) - \dim(\mathrm{graph}(\sim)). \tag{3.5}$$

The next proposition distinguishes the role of the three conditions in Proposition 3.4.2.

Proposition 3.4.3 *Conditions (i) and (ii) in Proposition 3.4.2 are necessary and sufficient for $\mathcal{M}/\sim$ to admit an atlas that makes π a submersion. Such an atlas is unique, and the atlas topology of $\mathcal{M}/\sim$ is identical to its quotient topology. Condition (iii) in Proposition 3.4.2 is necessary and sufficient for the quotient topology to be Hausdorff.*

The following result follows from Proposition 3.3.3 by using the fact that the natural projection to a quotient manifold is by definition a submersion.

Proposition 3.4.4 *Let $\mathcal{M}/\sim$ be a quotient manifold of a manifold $\mathcal{M}$ and let π denote the canonical projection. If $\dim(\mathcal{M}/\sim) < \dim(\mathcal{M})$, then each equivalence class $\pi^{-1}(\pi(x))$, $x \in \mathcal{M}$, is an embedded submanifold of $\mathcal{M}$ of dimension $\dim(\mathcal{M}) - \dim(\mathcal{M}/\sim)$.*

If $\dim(\mathcal{M}/\sim) = \dim(\mathcal{M})$, then each equivalence class $\pi^{-1}(\pi(x))$, $x \in \mathcal{M}$, is a discrete set of points. From now on we consider only the case $\dim(\mathcal{M}/\sim) < \dim(\mathcal{M})$.

When $\mathcal{M}$ is $\mathbb{R}^{n\times p}$ or a submanifold of $\mathbb{R}^{n\times p}$, we call $\mathcal{M}/\sim$ a *matrix quotient manifold*. For ease of reference, we will use the generic name *structure space* both for embedding spaces (associated with embedded submanifolds) and for total spaces (associated with quotient manifolds). We call a *matrix manifold* any manifold that is constructed from $\mathbb{R}^{n\times p}$ by the operations of taking embedded submanifolds and quotient manifolds. The major matrix manifolds that appear in this book are the noncompact Stiefel manifold (defined in Section 3.1.5), the orthogonal Stiefel manifold (Section 3.3.2), and the Grassmann manifold (Section 3.4.4). Other important matrix manifolds are the *oblique manifold*

$$\{X \in \mathbb{R}^{n\times p} : \operatorname{diag}(X^T X) = I_p\},$$

where $\operatorname{diag}(M)$ denotes the matrix M with all its off-diagonal elements assigned to zero; the generalized Stiefel manifold

$$\{X \in \mathbb{R}^{n\times p} : X^T B X = I\}$$

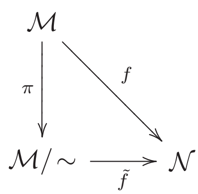
where B is a symmetric positive-definite matrix; the *flag manifolds*, which are quotients of $\mathbb{R}_*^{n\times p}$ where two matrices are equivalent when they are related by a right multiplication by a block upper triangular matrix with prescribed block size; and the manifold of *symplectic matrices*

$$\{X \in \mathbb{R}^{2n\times 2n} : X^T J X = J\},$$

where $J = \left[\begin{smallmatrix} 0_n & I_n \\ -I_n & 0_n \end{smallmatrix}\right]$.

3.4.2 Functions on quotient manifolds

A function f on $\mathcal{M}$ is termed *invariant under* $\sim$ if $f(x) = f(y)$ whenever $x \sim y$, in which case the function f induces a unique function $\tilde{f}$ on $\mathcal{M}/\sim$, called the *projection* of f, such that $f = \tilde{f} \circ \pi$.



The smoothness of $\tilde{f}$ can be checked using the following result.

Proposition 3.4.5 *Let $\mathcal{M}/\sim$ be a quotient manifold and let $\tilde{f}$ be a function on $\mathcal{M}/\sim$. Then $\tilde{f}$ is smooth if and only if $f := \tilde{f} \circ \pi$ is a smooth function on $\mathcal{M}$.*

3.4.3 The real projective space $\mathbb{RP}^{n-1}$

The real projective space $\mathbb{RP}^{n-1}$ is the set of all directions in $\mathbb{R}^n$, i.e., the set of all straight lines passing through the origin of $\mathbb{R}^n$. Let $\mathbb{R}^n_* := \mathbb{R}^n - \{0\}$ denote the Euclidean space $\mathbb{R}^n$ with the origin removed. Note that $\mathbb{R}^n_*$ is the $p = 1$ particularization of the noncompact Stiefel manifold $\mathbb{R}^{n\times p}_*$ (Section 3.1.5); hence $\mathbb{R}^n_*$ is an open submanifold of $\mathbb{R}^n$. The real projective space $\mathbb{RP}^{n-1}$ is naturally identified with the quotient $\mathbb{R}^n_*/\sim$, where the equivalence relation is defined by

$$x \sim y \qquad \Leftrightarrow \qquad \exists t \in \mathbb{R}_* : y = xt,$$

and we write

$$\mathbb{RP}^{n-1} \simeq \mathbb{R}^n_*/\sim$$

to denote the identification of the two sets.

The proof that $\mathbb{R}^n_*/\sim$ is a quotient manifold follows as a special case of Proposition 3.4.6 (stating that the Grassmann manifold is a matrix quotient manifold). The letters $\mathbb{RP}$ stand for "real projective", while the superscript $(n-1)$ is the dimension of the manifold. There are also complex projective spaces and more generally projective spaces over more abstract vector spaces.

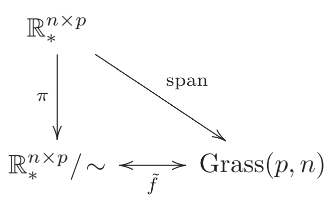
3.4.4 The Grassmann manifold $\mathrm{Grass}(p, n)$

Let n be a positive integer and let p be a positive integer not greater than n. Let $\mathrm{Grass}(p, n)$ denote the set of all p-dimensional subspaces of $\mathbb{R}^n$. In this section, we produce a one-to-one correspondence between $\mathrm{Grass}(p, n)$ and a quotient manifold of $\mathbb{R}^{n\times p}$, thereby endowing $\mathrm{Grass}(p, n)$ with a matrix manifold structure.

Recall that the noncompact Stiefel manifold $\mathbb{R}^{n\times p}_*$ is the set of all $n \times p$ matrices with full column rank. Let $\sim$ denote the equivalence relation on $\mathbb{R}^{n\times p}_*$ defined by

$$X \sim Y \qquad \Leftrightarrow \qquad \mathrm{span}(X) = \mathrm{span}(Y), \tag{3.6}$$

where $\mathrm{span}(X)$ denotes the subspace $\{X\alpha : \alpha \in \mathbb{R}^p\}$ spanned by the columns of $X \in \mathbb{R}^{n\times p}_*$. Since the fibers of $\mathrm{span}(\cdot)$ are the equivalence classes of $\sim$ and since $\mathrm{span}(\cdot)$ is onto $\mathrm{Grass}(p, n)$, it follows that $\mathrm{span}(\cdot)$ induces a one-to-one correspondence between $\mathrm{Grass}(p, n)$ and $\mathbb{R}^{n\times p}_*/\sim$.



Before showing that the set $\mathbb{R}^{n\times p}_*/\sim$ is a quotient manifold, we introduce some notation and terminology. If a matrix X and a subspace $\mathcal{X}$ satisfy

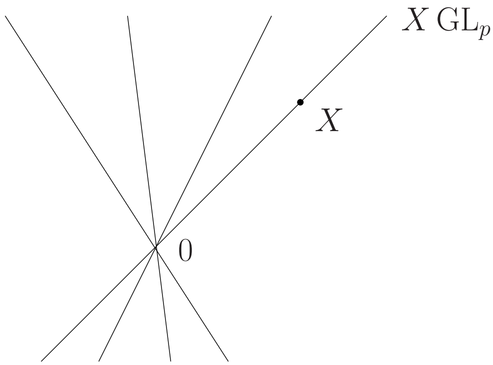


Figure 3.3 Schematic illustration of the representation of $\mathrm{Grass}(p, n)$ as the quotient space $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$. Each point is an n-by-p matrix. Each line is an equivalence class of the matrices that have the same span. Each line corresponds to an element of $\mathrm{Grass}(p, n)$. The figure corresponds to the case $n = 2$, $p = 1$.

$\mathcal{X} = \mathrm{span}(X)$, we say that $\mathcal{X}$ is the *span* of X, that X spans $\mathcal{X}$, or that X is a *matrix representation* of $\mathcal{X}$. The set of all matrix representations of $\mathrm{span}(X)$ is the equivalence class $\pi^{-1}(\pi(X))$. We have $\pi^{-1}(\pi(X)) = \{XM : M \in \mathrm{GL}_p\} =: X\mathrm{GL}_p$; indeed, the operations $X \mapsto XM$, $M \in \mathrm{GL}_p$, correspond to all possible changes of basis for $\mathrm{span}(X)$. We will thus use the notation $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ for $\mathbb{R}_*^{n\times p}/\sim$. Therefore we have

$$\mathrm{Grass}(p, n) \simeq \mathbb{R}_*^{n\times p}/\mathrm{GL}_p.$$

A schematic illustration of the quotient $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ is given in Figure 3.3.

The identification of $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ with the set of p-dimensional subspaces (p-planes) in $\mathbb{R}^n$ makes this quotient particularly worth studying. Next, the quotient $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ is shown to be a quotient manifold.

Proposition 3.4.6 (Grassmann manifold) *The quotient set $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ (i.e., the quotient of $\mathbb{R}_*^{n\times p}$ by the equivalence relation defined in* (3.6)*) admits a (unique) structure of quotient manifold.*

Proof. We show that the conditions in Proposition 3.4.2 are satisfied. We first prove condition (ii). Let (X_0, Y_0) be in $\mathrm{graph}(\sim)$. Then there exists M such that $Y_0 = X_0 M$. Given any V in $\mathbb{R}^{n\times p}$, the curve $t \mapsto (X_0 + tV, (X_0 + tV)M)$ is into $\mathrm{graph}(\sim)$ and satisfies $\frac{\mathrm{d}}{\mathrm{d}t}(\pi_1(\gamma(t)))\big|_{t=0} = V$. This shows that π_1 is a submersion. For condition (iii), observe that the graph of $\sim$ is closed as it is the preimage of the closed set $\{0_{n\times p}\}$ under the continuous function $\mathbb{R}_*^{n\times p} \times \mathbb{R}_*^{n\times p} \to \mathbb{R}^{n\times p} : (X, Y) \mapsto (I - X(X^TX)^{-1}X^T)Y$. For condition (i), the idea is to produce submersions F_i with open domain $\Omega_i \subset (\mathbb{R}_*^{n\times p} \times \mathbb{R}_*^{n\times p})$ such that $\mathrm{graph}(\sim) \cap \Omega_i$ is the zero-level set of F_i and that the Ω_i's cover $\mathrm{graph}(\sim)$. It then follows from Proposition 3.3.3 that $\mathrm{graph}(\sim)$ is an embedded submanifold of $\mathbb{R}_*^{n\times p} \times \mathbb{R}_*^{n\times p}$. To this end, assume for a moment that we have a smooth function

$$\mathbb{R}_*^{n\times p} \to \mathrm{St}(n-p,n) : X \mapsto X_\perp \tag{3.7}$$

such that $X^T X_\perp = 0$ for all X in an open domain $\tilde{\Omega}$ and consider

$$F : \tilde{\Omega} \times \mathbb{R}_*^{n\times p} \to \mathbb{R}^{(n-p)\times p} : (X,Y) \mapsto X_\perp^T Y.$$

Then $F^{-1}(0) = \mathrm{graph}(\sim) \cap \mathrm{dom}(F)$. Moreover, F is a submersion on its domain since for any $V \in \mathbb{R}^{(n-p)\times p}$,

$$\mathrm{D}F(X,Y)[0, X_\perp V] = X_\perp^T (X_\perp V) = V.$$

It remains to define the smooth function (3.7). Depending on n and p, it may or may not be possible to define such a function on the whole $\mathbb{R}_*^{n\times p}$. However, there are always such functions, constructed as follows, whose domain $\tilde{\Omega}$ is open and dense in $\mathbb{R}_*^{n\times p}$. Let $E \in \mathbb{R}^{n\times(n-p)}$ be a constant matrix of the form

$$E = \left[e_{i_1} | \cdots | e_{i_{n-p}}\right],$$

where the e_i's are the canonical vectors in $\mathbb{R}^n$ (unit vectors with a 1 in the ith entry), and define $X_\perp$ as the orthonormal matrix obtained by taking the last $n-p$ columns of the Gram-Schmidt orthogonalization of the matrix $[X|E]$. This function is smooth on the domain $\tilde{\Omega} = \{X \in \mathbb{R}_*^{n\times p} : [X|E] \text{ full rank}\}$, which is an open dense subset of $\mathbb{R}_*^{n\times p}$. Consequently, $F(X,Y) = X_\perp^T Y$ is smooth (and submersive) on the domain $\Omega = \tilde{\Omega} \times \mathbb{R}_*^{n\times p}$. This shows that $\mathrm{graph}(\sim) \cap \Omega$ is an embedded submanifold of $(\mathbb{R}_*^{n\times p} \times \mathbb{R}_*^{n\times p})$. Taking other matrices E yields other domains Ω which together cover $(\mathbb{R}_*^{n\times p} \times \mathbb{R}_*^{n\times p})$, so $\mathrm{graph}(\sim)$ is an embedded submanifold of $(\mathbb{R}_*^{n\times p} \times \mathbb{R}_*^{n\times p})$, and the proof is complete. □

Endowed with its quotient manifold structure, the set $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ is called the *Grassmann manifold* of p-planes in $\mathbb{R}^n$ and denoted by $\mathrm{Grass}(p,n)$. The particular case $\mathrm{Grass}(1,n) = \mathbb{RP}^n$ is the real projective space discussed in Section 3.4.3. From Proposition 3.3.3, we have that $\dim(\mathrm{graph}(\sim)) = 2np - (n-p)p$. It then follows from (3.5) that

$$\dim(\mathrm{Grass}(p,n)) = p(n-p).$$

3.5 TANGENT VECTORS AND DIFFERENTIAL MAPS

There are several possible approaches to generalizing the notion of a *directional derivative*

$$\mathrm{D}f(x)[\eta] = \lim_{t\to 0} \frac{f(x+t\eta) - f(x)}{t} \tag{3.8}$$

to a real-valued function f defined on a manifold. A first possibility is to view η as a *derivation at* x, that is, an object that, when given a real-valued function f defined on a neighborhood of $x \in \mathcal{M}$, returns a real ηf, and that satisfies the properties of a derivation operation: linearity and the Leibniz rule (see Section 3.5.5). This "axiomatization" of the notion of a directional derivative is elegant and powerful, but it gives little intuition as to how a tangent vector could possibly be represented as a matrix array in a computer.

A second, perhaps more intuitive approach to generalizing the directional derivative (3.8) is to replace $t \mapsto (x+t\eta)$ by a smooth curve γ on $\mathcal{M}$ through x (i.e., $\gamma(0) = x$). This yields a well-defined directional derivative $\frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t}\Big|_{t=0}$. (Note that this is a classical derivative since the function $t \mapsto f(\gamma(t))$ is a smooth function from $\mathbb{R}$ to $\mathbb{R}$.) Hence we have an operation, denoted by $\dot{\gamma}(0)$, that takes a function f, defined locally in a neighbourhood of x, and returns the real number $\frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t}\Big|_{t=0}$.

These two approaches are reconciled by showing that every derivative along a curve defines a pointwise derivation and that every pointwise derivation can be realized as a derivative along a curve. The first claim is direct. The second claim can be proved using a local coordinate representation, a third approach used to generalize the notion of a directional derivative.

3.5.1 Tangent vectors

Let $\mathcal{M}$ be a manifold. A smooth mapping $\gamma : \mathbb{R} \to \mathcal{M}$: $t \mapsto \gamma(t)$ is termed a *curve in* $\mathcal{M}$. The idea of defining a derivative $\gamma'(t)$ as

$$\gamma'(t) := \lim_{\tau \to 0} \frac{\gamma(t+\tau) - \gamma(t)}{\tau} \tag{3.9}$$

requires a vector space structure to compute the difference $\gamma(t+\tau)-\gamma(t)$ and thus fails for an abstract nonlinear manifold. However, given a smooth real-valued function f on $\mathcal{M}$, the function $f \circ \gamma : t \mapsto f(\gamma(t))$ is a smooth function from $\mathbb{R}$ to $\mathbb{R}$ with a well-defined classical derivative. This is exploited in the following definition. Let x be a point on $\mathcal{M}$, let γ be a curve through x at $t = 0$, and let $\mathfrak{F}_x(\mathcal{M})$ denote the set of smooth real-valued functions defined on a neighborhood of x. The mapping $\dot{\gamma}(0)$ from $\mathfrak{F}_x(\mathcal{M})$ to $\mathbb{R}$ defined by

$$\dot{\gamma}(0)f := \frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t}\Big|_{t=0}, \qquad f \in \mathfrak{F}_x(\mathcal{M}), \tag{3.10}$$

is called the *tangent vector to the curve* γ *at* $t = 0$.

We emphasize that $\dot{\gamma}(0)$ is defined as a mapping from $\mathfrak{F}_x(\mathcal{M})$ to $\mathbb{R}$ and not as the time derivative $\gamma'(0)$ as in (3.9), which in general is meaningless. However, when $\mathcal{M}$ is (a submanifold of) a vector space $\mathcal{E}$, the mapping $\dot{\gamma}(0)$ from $\mathfrak{F}_x(\mathcal{M})$ to $\mathbb{R}$ and the derivative $\gamma'(0) := \lim_{t\to 0} \frac{1}{t}(\gamma(t)-\gamma(0))$ are closely related: for all functions $\overline{f}$ defined in a neighborhood $\mathcal{U}$ of $\gamma(0)$ in $\mathcal{E}$, we have

$$\dot{\gamma}(0)f = \mathrm{D}\overline{f}\,(\gamma(0))\,[\gamma'(0)]\,,$$

where f denotes the restriction of $\overline{f}$ to $\mathcal{U}\cap\mathcal{M}$; see Sections 3.5.2 and 3.5.7 for details. It is useful to keep this interpretation in mind because the derivative $\gamma'(0)$ is a more familiar mathematical object than $\dot{\gamma}(0)$.

We can now formally define the notion of a tangent vector.

Definition 3.5.1 (tangent vector) *A* tangent vector ξ_x *to a manifold* $\mathcal{M}$ *at a point* x *is a mapping from* $\mathfrak{F}_x(\mathcal{M})$ *to* $\mathbb{R}$ *such that there exists a curve* γ *on* $\mathcal{M}$ *with* $\gamma(0) = x$*, satisfying*

$$\xi_x f = \dot{\gamma}(0) f := \left.\frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t}\right|_{t=0}$$

for all $f \in \mathfrak{F}_x(\mathcal{M})$*. Such a curve* γ *is said to* realize *the tangent vector* ξ_x*.*

The point x is called the *foot* of the tangent vector ξ_x. We will often omit the subscript indicating the foot and simply write ξ for ξ_x.

Given a tangent vector ξ to $\mathcal{M}$ at x, there are infinitely many curves γ that realize ξ (i.e., $\dot{\gamma}(0) = \xi$). They can be characterized as follows in local coordinates.

Proposition 3.5.2 *Two curves* γ_1 *and* γ_2 *through a point* x *at* $t = 0$ *satisfy* $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$ *if and only if, given a chart* $(\mathcal{U}, \varphi)$ *with* $x \in \mathcal{U}$*, it holds that*

$$\left.\frac{\mathrm{d}(\varphi(\gamma_1(t)))}{\mathrm{d}t}\right|_{t=0} = \left.\frac{\mathrm{d}(\varphi(\gamma_2(t)))}{\mathrm{d}t}\right|_{t=0}.$$

Proof. The "only if" part is straightforward since each component of the vector-valued φ belongs to $\mathfrak{F}_x(\mathcal{M})$. For the "if" part, given any $f \in \mathfrak{F}_x(\mathcal{M})$, we have

$$\begin{aligned}\dot{\gamma}_1(0) f &= \left.\frac{\mathrm{d}(f(\gamma_1(t)))}{\mathrm{d}t}\right|_{t=0} = \left.\frac{\mathrm{d}((f \circ \varphi^{-1})(\varphi(\gamma_1(t))))}{\mathrm{d}t}\right|_{t=0} \\ &= \left.\frac{\mathrm{d}((f \circ \varphi^{-1})(\varphi(\gamma_2(t))))}{\mathrm{d}t}\right|_{t=0} = \dot{\gamma}_2(0) f.\end{aligned}$$

□

The *tangent space* to $\mathcal{M}$ at x, denoted by $T_x\mathcal{M}$, is the set of all tangent vectors to $\mathcal{M}$ at x. This set admits a structure of *vector space* as follows. Given $\dot{\gamma}_1(0)$ and $\dot{\gamma}_2(0)$ in $T_x\mathcal{M}$ and a, b in $\mathbb{R}$, define

$$(a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0))\, f := a\,(\dot{\gamma}_1(0) f) + b\,(\dot{\gamma}_2(0) f).$$

To show that $(a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0))$ is a well-defined tangent vector, we need to show that there exists a curve γ such that $\dot{\gamma}(0) = a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0)$. Such a curve is obtained by considering a chart $(\mathcal{U}, \varphi)$ with $x \in \mathcal{U}$ and defining $\gamma(t) = \varphi^{-1}(a\varphi(\gamma_1(t) + b\varphi(\gamma_2(t))$. It is readily checked that this γ satisfies the required property.

The property that the tangent space $T_x\mathcal{M}$ is a vector space is very important. In the same way that the derivative of a real-valued function provides a local linear approximation of the function, the tangent space $T_x\mathcal{M}$ provides a local vector space approximation of the manifold. In particular, in Section 4.1, we define mappings, called *retractions*, between $\mathcal{M}$ and $T_x\mathcal{M}$, which can be used to locally transform an optimization problem on the manifold $\mathcal{M}$ into an optimization problem on the more friendly vector space $T_x\mathcal{M}$.

Using a coordinate chart, it is possible to show that the dimension of the vector space $T_x\mathcal{M}$ is equal to d, the dimension of the manifold $\mathcal{M}$: given a chart $(\mathcal{U}, \varphi)$ at x, a basis of $T_x\mathcal{M}$ is given by $(\dot\gamma_1(0), \ldots, \dot\gamma_d(0))$, where $\gamma_i(t) := \varphi^{-1}(\varphi(x) + te_i)$, with e_i denoting the ith canonical vector of $\mathbb{R}^d$. Notice that $\dot\gamma_i(0)f = \partial_i(f \circ \varphi^{-1})(\varphi(x))$, where ∂_i denotes the partial derivative with respect to the ith component:

$$\partial_i h(x) := \lim_{t\to 0} \frac{h(x+te_i) - h(x)}{t}.$$

One has, for any tangent vector $\dot\gamma(0)$, the decomposition

$$\dot\gamma(0) = \sum_i (\dot\gamma(0)\varphi_i)\dot\gamma_i(0),$$

where φ_i denotes the ith component of φ. This provides a way to define the coordinates of tangent vectors at x using the chart $(\mathcal{U}, \varphi)$, by defining the element of $\mathbb{R}^d$

$$\begin{pmatrix} \dot\gamma(0)\varphi_1 \\ \vdots \\ \dot\gamma(0)\varphi_d \end{pmatrix}$$

as the representation of the tangent vector $\dot\gamma(0)$ in the chart $(\mathcal{U}, \varphi)$.

3.5.2 Tangent vectors to a vector space

Let $\mathcal{E}$ be a vector space and let x be a point of $\mathcal{E}$. As pointed out in Section 3.1.4, $\mathcal{E}$ admits a linear manifold structure. Strictly speaking, a tangent vector ξ to $\mathcal{E}$ at x is a mapping

$$\xi : \mathfrak{F}_x(\mathcal{E}) \to \mathbb{R} : f \mapsto \xi f = \left.\frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t}\right|_{t=0},$$

where γ is a curve in $\mathcal{E}$ with $\gamma(0) = x$. Defining $\gamma'(0) \in \mathcal{E}$ as in (3.9), we have

$$\xi f = \mathrm{D} f\,(x)\,[\gamma'(0)]\,.$$

Moreover, $\gamma'(0)$ does not depend on the curve γ that realizes ξ. This defines a canonical linear one-to-one correspondence $\xi \mapsto \gamma'(0)$, which identifies $T_x\mathcal{E}$ with $\mathcal{E}$:

$$T_x\mathcal{E} \simeq \mathcal{E}. \tag{3.11}$$

Since tangent vectors are local objects (a tangent vector at a point x acts on smooth real-valued functions defined in any neighborhood of x), it follows that if $\mathcal{E}_*$ is an open submanifold of $\mathcal{E}$, then

$$T_x\mathcal{E}_* \simeq \mathcal{E} \tag{3.12}$$

for all $x \in \mathcal{E}_*$. A schematic illustration is given in Figure 3.4.

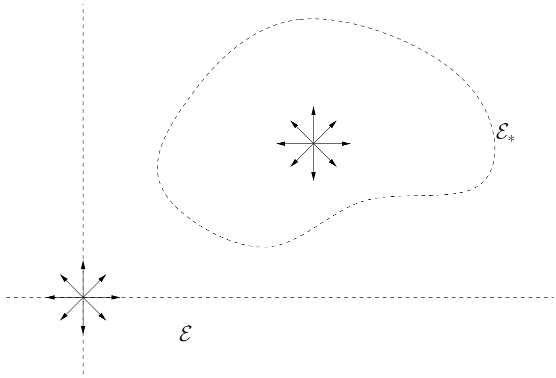


Figure 3.4 Tangent vectors to an open subset $\mathcal{E}_*$ of a vector space $\mathcal{E}$.

3.5.3 Tangent bundle

Given a manifold $\mathcal{M}$, let $T\mathcal{M}$ be the set of all tangent vectors to $\mathcal{M}$:

$$T\mathcal{M} := \bigcup_{x\in\mathcal{M}} T_x\mathcal{M}.$$

Since each $\xi \in T\mathcal{M}$ is in one and only one tangent space $T_x\mathcal{M}$, it follows that $\mathcal{M}$ is a quotient of $T\mathcal{M}$ with natural projection

$$\pi : T\mathcal{M} \to \mathcal{M} : \xi \in T_x\mathcal{M} \;\mapsto\; x,$$

i.e., $\pi(\xi)$ is the foot of ξ. The set $T\mathcal{M}$ admits a natural manifold structure as follows. Given a chart $(\mathcal{U}, \varphi)$ of $\mathcal{M}$, the mapping

$$\xi \in T_x\mathcal{M} \;\mapsto\; (\varphi_1(x), \ldots, \varphi_d(x), \xi\varphi_1, \ldots, \xi\varphi_d)^T$$

is a chart of the set $T\mathcal{M}$ with domain $\pi^{-1}(\mathcal{U})$. It can be shown that the collection of the charts thus constructed forms an atlas of the set $T\mathcal{M}$, turning it into a manifold called the *tangent bundle* of $\mathcal{M}$.

3.5.4 Vector fields

A *vector field* ξ on a manifold $\mathcal{M}$ is a smooth function from $\mathcal{M}$ to the tangent bundle $T\mathcal{M}$ that assigns to each point $x \in \mathcal{M}$ a tangent vector $\xi_x \in T_x\mathcal{M}$. On a submanifold of a vector space, a vector field can be pictured as a collection of arrows, one at each point of $\mathcal{M}$. Given a vector field ξ on $\mathcal{M}$ and a (smooth) real-valued function $f \in \mathfrak{F}(\mathcal{M})$, we let ξf denote the real-valued function on $\mathcal{M}$ defined by

$$(\xi f)(x) := \xi_x(f)$$

for all x in $\mathcal{M}$. The addition of two vector fields and the multiplication of a vector field by a function $f \in \mathfrak{F}(\mathcal{M})$ are defined as follows:

$$(f\xi)_x := f(x)\xi_x,$$
$$(\xi + \zeta)_x := \xi_x + \zeta_x \qquad \text{for all } x \in \mathcal{M}.$$

Smoothness is preserved by these operations. We let $\mathfrak{X}(\mathcal{M})$ denote the set of smooth vector fields endowed with these two operations.

Let $(\mathcal{U}, \varphi)$ be a chart of the manifold $\mathcal{M}$. The vector field E_i on $\mathcal{U}$ defined by

$$(E_i f)(x) := \partial_i (f \circ \varphi^{-1})(\varphi(x)) = \mathrm{D}(f \circ \varphi^{-1})\,(\varphi(x))\,[e_i]$$

is called the *ith coordinate vector field* of $(\mathcal{U}, \varphi)$. These coordinate vector fields are smooth, and every vector field ξ admits the decomposition

$$\xi = \sum_i (\xi \varphi_i) E_i$$

on $\mathcal{U}$. (A pointwise version of this result was given in Section 3.5.1.)

If the manifold is an n-dimensional vector space $\mathcal{E}$, then, given a basis $(e_i)_{i=1,\dots,d}$ of $\mathcal{E}$, the vector fields E_i, $i = 1, \dots, n$, defined by

$$(E_i f)(x) := \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t} = \mathrm{D}f\,(x)\,[e_i]$$

form a basis of $\mathfrak{X}(\mathcal{E})$.

3.5.5 Tangent vectors as derivations*

Let x and η be elements of $\mathbb{R}^n$. The derivative mapping that, given a real-valued function f on $\mathbb{R}^n$, returns the real $\mathrm{D}f\,(x)\,[\eta]$ can be axiomatized as follows on manifolds. Let $\mathcal{M}$ be a manifold and recall that $\mathfrak{F}(\mathcal{M})$ denotes the set of all smooth real-valued functions on $\mathcal{M}$. Note that $\mathfrak{F}(\mathcal{M}) \subset \mathfrak{F}_x(\mathcal{M})$ for all $x \in \mathcal{M}$. A *derivation at* $x \in \mathcal{M}$ is a mapping ξ_x from $\mathfrak{F}(\mathcal{M})$ to $\mathbb{R}$ that is

1. $\mathbb{R}$-linear: $\xi_x(af + bg) = a\xi_x(f) + b\xi_x(g)$, and
2. Leibnizian: $\xi_x(fg) = \xi_x(f)g(x) + f(x)\xi_x(g)$, for all $a,\ b \in \mathbb{R}$ and $f,\ g \in \mathfrak{F}(\mathcal{M})$.

With the operations

$$(\xi_x + \zeta_x)f := \xi_x(f) + \zeta_x(f),$$
$$(a\xi_x)f := a\xi_x(f) \qquad \text{for all } f \in \mathfrak{F}(\mathcal{M}),\ a \in \mathbb{R},$$

the set of all derivations at x becomes a vector space. It can also be shown that a derivation ξ_x at x is a local notion: if two real-valued functions f and g are equal on a neighborhood of x, then $\xi_x(f) = \xi_x(g)$.

The concept of a tangent vector at x, as defined in Section 3.5.1, and the notion of a derivation at x are equivalent in the following sense: (i) Given a curve γ on $\mathcal{M}$ through x at $t = 0$, the mapping $\dot\gamma(0)$ from $\mathfrak{F}(\mathcal{M}) \subseteq \mathfrak{F}_x(\mathcal{M})$ to $\mathbb{R}$, defined in (3.10), is a derivation at x. (ii) Given a derivation ξ at x, there exists a curve γ on $\mathcal{M}$ through x at $t = 0$ such that $\dot\gamma(0) = \xi$. For example, the curve γ defined by $\gamma(t) = \varphi^{-1}\left(\varphi(0) + t\sum_i(\xi(\varphi_i)e_i)\right)$ satisfies the property.

A (global) *derivation* on $\mathfrak{F}(\mathcal{M})$ is a mapping $\mathcal{D} : \mathfrak{F}(\mathcal{M}) \to \mathfrak{F}(\mathcal{M})$ that is

1. $\mathbb{R}$-linear: $\mathcal{D}(af + bg) = a\,\mathcal{D}(f) + b\,\mathcal{D}(g)$, $(a,\ b \in \mathbb{R})$, and
2. Leibnizian: $\mathcal{D}(fg) = \mathcal{D}(f)g + f\,\mathcal{D}(g)$.

Every vector field $\xi \in \mathfrak{X}(\mathcal{M})$ defines a derivation $f \mapsto \xi f$. Conversely, every derivation on $\mathfrak{F}(\mathcal{M})$ can be realized as a vector field. (Viewing vector fields as derivations comes in handy in understanding Lie brackets; see Section 5.3.1.)

3.5.6 Differential of a mapping

Let $F : \mathcal{M} \to \mathcal{N}$ be a smooth mapping between two manifolds $\mathcal{M}$ and $\mathcal{N}$. Let ξ_x be a tangent vector at a point x of $\mathcal{M}$. It can be shown that the mapping $\mathrm{D}F\,(x)\,[\xi_x]$ from $\mathfrak{F}_{F(x)}(\mathcal{N})$ to $\mathbb{R}$ defined by

$$(\mathrm{D}F\,(x)\,[\xi])\,f := \xi(f \circ F) \tag{3.13}$$

is a tangent vector to $\mathcal{N}$ at $F(x)$. The tangent vector $\mathrm{D}F\,(x)\,[\xi_x]$ is realized by $F \circ \gamma$, where γ is any curve that realizes ξ_x. The mapping

$$\mathrm{D}F(x) : T_x\mathcal{M} \to T_{F(x)}\mathcal{N} : \xi \mapsto \mathrm{D}F\,(x)\,[\xi]$$

is a linear mapping called the *differential* (or *differential map*, *derivative*, or *tangent map*) of F at x (see Figure 3.5).

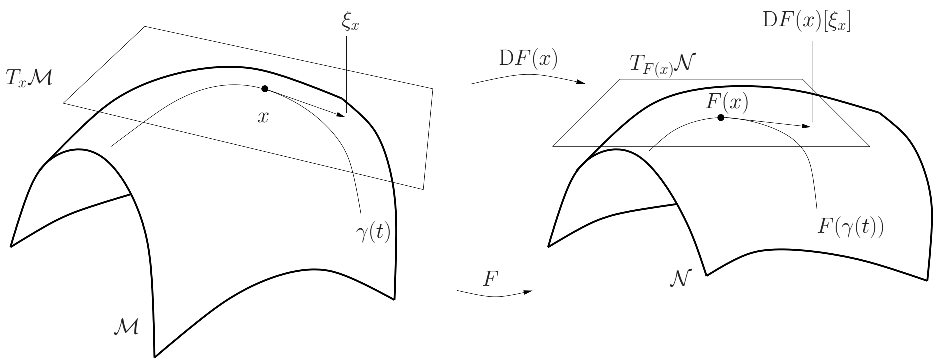


Figure 3.5 Differential map of F at x.

Note that F is an immersion (respectively, submersion) if and only if $\mathrm{D}F(x) : T_x\mathcal{M} \to T_{F(x)}\mathcal{N}$ is an injection (respectively, surjection) for every $x \in \mathcal{M}$.

If $\mathcal{N}$ is a vector space $\mathcal{E}$, then the canonical identification $T_{F(x)}\mathcal{E} \simeq \mathcal{E}$ yields

$$\mathrm{D}F\,(x)\,[\xi_x] = \sum_i (\xi_x F^i)e_i, \tag{3.14}$$

where $F(x) = \sum_i F^i(x)e_i$ is the decomposition of $F(x)$ in a basis $(e_i)_{i=1,\ldots,n}$ of $\mathcal{E}$.

If $\mathcal{N} = \mathbb{R}$, then $F \in \mathfrak{F}_x(\mathcal{M})$, and we simply have

$$\mathrm{D}F\,(x)\,[\xi_x] = \xi_x F \tag{3.15}$$

using the identification $T_x\mathbb{R} \simeq \mathbb{R}$. We will often use $\mathrm{D}F(x)[\xi_x]$ as an alternative notation for $\xi_x F$, as it better emphasizes the derivative aspect.

If $\mathcal{M}$ and $\mathcal{N}$ are linear manifolds, then, with the identification $T_x\mathcal{M} \simeq \mathcal{M}$ and $T_y\mathcal{N} \simeq \mathcal{N}$, $\mathrm{D}F(x)$ reduces to its classical definition

$$\mathrm{D}F(x)[\xi_x] = \lim_{t\to 0} \frac{F(x+t\xi_x) - F(x)}{t}. \tag{3.16}$$

Given a differentiable function $F : \mathcal{M} \mapsto \mathcal{N}$ and a vector field ξ on $\mathcal{M}$, we let $\mathrm{D}F[\xi]$ denote the mapping

$$\mathrm{D}F[\xi] : \mathcal{M} \to T\mathcal{N} : x \mapsto \mathrm{D}F(x)[\xi_x].$$

In particular, given a real-valued function f on $\mathcal{M}$ and a vector field ξ on $\mathcal{M}$,

$$\mathrm{D}f[\xi] = \xi f.$$

3.5.7 Tangent vectors to embedded submanifolds

We now investigate the case where $\mathcal{M}$ is an embedded submanifold of a vector space $\mathcal{E}$. Let γ be a curve in $\mathcal{M}$, with $\gamma(0) = x$. Define

$$\gamma'(0) := \lim_{t\to 0} \frac{\gamma(t) - \gamma(0)}{t},$$

where the subtraction is well defined since $\gamma(t)$ belongs to the vector space $\mathcal{E}$ for all t. (Strictly speaking, one should write $i(\gamma(t)) - i(\gamma(0))$, where i is the natural inclusion of $\mathcal{M}$ in $\mathcal{E}$; the inclusion is omitted to simplify the notation.) It follows that $\gamma'(0)$ thus defined is an element of $T_x\mathcal{E} \simeq \mathcal{E}$ (see Figure 3.6). Since γ is a curve in $\mathcal{M}$, it also induces a tangent vector $\dot\gamma(0) \in T_x\mathcal{M}$. Not

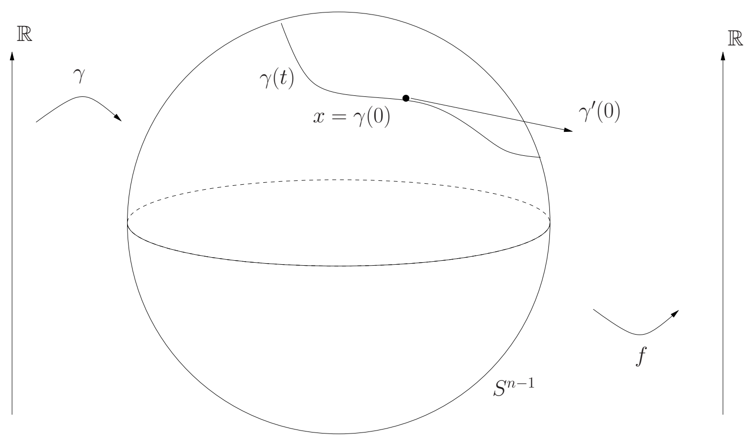


Figure 3.6 Curves and tangent vectors on the sphere. Since S^{n-1} is an embedded submanifold of $\mathbb{R}^n$, the tangent vector $\dot\gamma(0)$ can be pictured as the directional derivative $\gamma'(0)$.

surprisingly, $\gamma'(0)$ and $\dot\gamma(0)$ are closely related: If $\overline{f}$ is a real-valued function in a neighborhood $\mathcal{U}$ of x in $\mathcal{E}$ and f denotes the restriction of $\overline{f}$ to $\mathcal{U} \cap \mathcal{M}$ (which is a neighborhood of x in $\mathcal{M}$ since $\mathcal{M}$ is embedded), then we have

$$\dot{\gamma}(0)f = \left.\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\right|_{t=0} = \left.\frac{\mathrm{d}}{\mathrm{d}t} \overline{f}(\gamma(t))\right|_{t=0} = \mathrm{D}\overline{f}\,(x)\,[\gamma'(0)]\,. \tag{3.17}$$

This yields a natural identification of $T_x\mathcal{M}$ with the set

$$\{\gamma'(0) : \gamma \text{ curve in } \mathcal{M},\ \gamma(0) = x\}, \tag{3.18}$$

which is a linear subspace of the vector space $T_x\mathcal{E} \simeq \mathcal{E}$. In particular, when $\mathcal{M}$ is a matrix submanifold (i.e., the embedding space is $\mathbb{R}^{n\times p}$), we have $T_x\mathcal{E} = \mathbb{R}^{n\times p}$, hence the tangent vectors to $\mathcal{M}$ are naturally represented by $n \times p$ matrix arrays.

Graphically, a tangent vector to a submanifold of a vector space can be thought of as an "arrow" tangent to the manifold. It is convenient to keep this intuition in mind when dealing with more abstract manifolds; however, one should bear in mind that the notion of a tangent arrow cannot always be visualized meaningfully in this manner, in which case one must return to the definition of tangent vectors as objects that, given a real-valued function, return a real number, as stated in Definition 3.5.1.

In view of the identification of $T_x\mathcal{M}$ with (3.18), we now write $\dot{\gamma}(t)$, $\gamma'(t)$, and $\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)$ interchangeably. We also use the equality sign, such as in (3.19) below, to denote the identification of $T_x\mathcal{M}$ with (3.18).

When $\mathcal{M}$ is (locally or globally) defined as a level set of a constant-rank function $F : \mathcal{E} \mapsto \mathbb{R}^n$, we have

$$T_x\mathcal{M} = \ker(\mathrm{D}F(x)). \tag{3.19}$$

In other words, the tangent vectors to $\mathcal{M}$ at x correspond to those vectors ξ that satisfy $\mathrm{D}F\,(x)\,[\xi] = 0$. Indeed, if γ is a curve in $\mathcal{M}$ with $\gamma(0) = x$, we have $F(\gamma(t)) = 0$ for all t, hence

$$\mathrm{D}F\,(x)\,[\dot{\gamma}(0)] = \left.\frac{\mathrm{d}(F(\gamma(t)))}{\mathrm{d}t}\right|_{t=0} = 0,$$

which shows that $\dot{\gamma}(0) \in \ker(\mathrm{D}F(x))$. By counting dimensions using Proposition 3.3.4, it follows that $T_x\mathcal{M}$ and $\ker(\mathrm{D}F(x))$ are two vector spaces of the same dimension with one included in the other. This proves the equality (3.19).

Example 3.5.1 ***Tangent space to a sphere***

Let $t \mapsto x(t)$ be a curve in the unit sphere S^{n-1} through x_0 at $t = 0$. Since $x(t) \in S^{n-1}$ for all t, we have

$$x^T(t)x(t) = 1$$

for all t. Differentiating this equation with respect to t yields

$$\dot{x}^T(t)x(t) + x^T(t)\dot{x}(t) = 0,$$

hence $\dot{x}(0)$ is an element of the set

$$\{z \in \mathbb{R}^n : x_0^T z = 0\}. \tag{3.20}$$

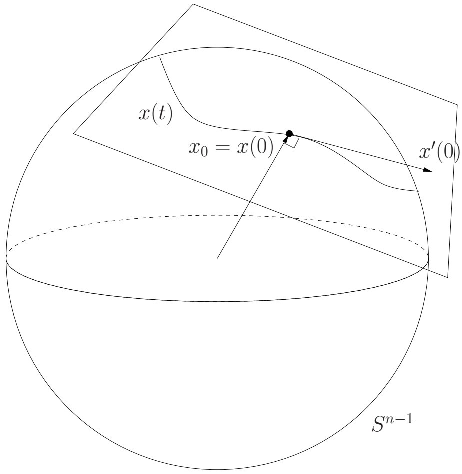


Figure 3.7 Tangent space on the sphere. Since S^{n-1} is an embedded submanifold of $\mathbb{R}^n$, the tangent space $T_x S^{n-1}$ can be pictured as the hyperplane tangent to the sphere at x, with origin at x.

This shows that $T_{x_0} S^{n-1}$ *is a subset of* (3.20). *Conversely, let* z *belong to the set* (3.20). *Then the curve* $t \mapsto x(t) := (x_0 + tz)/\|x_0 + tz\|$ *is on* S^{n-1} *and satisfies* $\dot{x}(0) = z$. *Hence* (3.20) *is a subset of* $T_{x_0} S^{n-1}$. *In conclusion,*

$$T_x S^{n-1} = \{z \in \mathbb{R}^n : x^T z = 0\}, \tag{3.21}$$

which is the set of all vectors orthogonal to x *in* $\mathbb{R}^n$*; see Figure 3.7.*

More directly, consider the function $F : \mathbb{R}^n \to \mathbb{R} : x \mapsto x^T x - 1$. *Since* $S^{n-1} = \{x \in \mathbb{R}^n : F(x) = 0\}$ *and since* F *is full rank on* S^{n-1}, *it follows from* (3.19) *that*

$$T_x S^{n-1} = \ker(\mathrm{D}F(x)) = \{z \in \mathbb{R}^n : x^T z + z^T x = 0\} = \{z \in \mathbb{R}^n : x^T z = 0\},$$

as in (3.21).

Example 3.5.2 ***Orthogonal Stiefel manifold***

We consider the orthogonal Stiefel manifold

$$\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}$$

as an embedded submanifold of the Euclidean space $\mathbb{R}^{n \times p}$ *(see Section 3.3.2). Let* X_0 *be an element of* $\mathrm{St}(p, n)$ *and let* $t \mapsto X(t)$ *be a curve in* $\mathrm{St}(p, n)$ *through* X_0 *at* $t = 0$*; i.e.,* $X(t) \in \mathbb{R}^{n \times p}$, $X(0) = X_0$, *and*

$$X^T(t) X(t) = I_p \tag{3.22}$$

for all t. *It follows by differentiating* (3.22) *that*

$$\dot{X}^T(t) X(t) + X^T(t) \dot{X}(t) = 0. \tag{3.23}$$

We deduce that $\dot{X}(0)$ *belongs to the set*

$$\{Z \in \mathbb{R}^{n\times p} : X_0^T Z + Z^T X_0 = 0\}. \tag{3.24}$$

We have thus shown that $T_{X_0}\,\mathrm{St}(p,n)$ *is a subset of* (3.24). *It is possible to conclude, as in the previous example, by showing that for all* Z *in* (3.24) *there is a curve in* $\mathrm{St}(p,n)$ *through* X_0 *at* t *such that* $\dot{X}(0) = Z$. *A simpler argument is to invoke* (3.19) *by pointing out that* (3.24) *is the kernel of* $\mathrm{D}F(X_0)$, *where* $F : X \mapsto X^T X$, *so that* I_p *is a regular value of* F *and* $F^{-1}(I_p) = \mathrm{St}(p,n)$. *In conclusion, the set described in* (3.24) *is the tangent space to* $\mathrm{St}(p,n)$ *at* X_0. *That is,*

$$T_X\,\mathrm{St}(p,n) = \{Z \in \mathbb{R}^{n\times p} : X^T Z + Z^T X = 0\}.$$

We now propose an alternative characterization of $T_X\,\mathrm{St}(p,n)$. *Without loss of generality, since* $\dot{X}(t)$ *is an element of* $\mathbb{R}^{n\times p}$ *and* $X(t)$ *has full rank, we can set*

$$\dot{X}(t) = X(t)\Omega(t) + X_\perp(t)K(t), \tag{3.25}$$

where $X_\perp(t)$ *is any* $n \times (n-p)$ *matrix such that* $\mathrm{span}(X_\perp(t))$ *is the orthogonal complement of* $\mathrm{span}(X(t))$. *Replacing* (3.25) *in* (3.23) *yields*

$$\Omega(t)^T + \Omega(t) = 0;$$

i.e., $\Omega(t)$ *is a skew-symmetric matrix. Counting dimensions, we deduce that*

$$T_X\,\mathrm{St}(p,n) = \{X\Omega + X_\perp K : \Omega^T = -\Omega,\ K \in \mathbb{R}^{(n-p)\times p}\}.$$

Observe that the two characterizations of $T_X\,\mathrm{St}(p,n)$ *are facilitated by the embedding of* $\mathrm{St}(p,n)$ *in* $\mathbb{R}^{n\times p}$: $T_X\,\mathrm{St}(p,n)$ *is identified with a linear subspace of* $\mathbb{R}^{n\times p}$.

Example 3.5.3 ***Orthogonal group***

Since the orthogonal group O_n *is* $\mathrm{St}(p,n)$ *with* $p = n$, *it follows from the previous section that*

$$T_U O_n = \{Z = U\Omega : \Omega^T = -\Omega\} = U\mathcal{S}_{\mathrm{skew}}(n), \tag{3.26}$$

where $\mathcal{S}_{\mathrm{skew}}(n)$ *denotes the set of all skew-symmetric* $n \times n$ *matrices.*

3.5.8 Tangent vectors to quotient manifolds

We have seen that tangent vectors of a submanifold embedded in a vector space $\mathcal{E}$ can be viewed as tangent vectors to $\mathcal{E}$ and pictured as arrows in $\mathcal{E}$ tangent to the submanifold. The situation of a quotient $\mathcal{E}/\sim$ of a vector space $\mathcal{E}$ is more abstract. Nevertheless, the structure space $\mathcal{E}$ also offers convenient representations of tangent vectors to the quotient.

For generality, we consider an abstract manifold $\overline{\mathcal{M}}$ and a quotient manifold $\mathcal{M} = \overline{\mathcal{M}}/\sim$ with canonical projection π. Let ξ be an element of $T_x\mathcal{M}$ and let $\overline{x}$ be an element of the equivalence class $\pi^{-1}(x)$. Any element $\overline{\xi}$ of

$T_{\overline{x}}\overline{\mathcal{M}}$ that satisfies $\mathrm{D}\pi(\overline{x})[\overline{\xi}] = \xi$ can be considered a representation of ξ. Indeed, for any smooth function $f : \mathcal{M} \to \mathbb{R}$, the function $\overline{f} := f \circ \pi : \overline{\mathcal{M}} \to \mathbb{R}$ is smooth (Proposition 3.4.5), and one has

$$\mathrm{D}\overline{f}(\overline{x})[\overline{\xi}] = \mathrm{D}f(\pi(\overline{x}))[\mathrm{D}\pi(\overline{x})[\overline{\xi}]] = \mathrm{D}f(x)[\xi].$$

A difficulty with this approach is that there are infinitely many valid representations $\overline{\xi}$ of ξ at $\overline{x}$.

It is desirable to identify a unique "lifted" representation of tangent vectors of $T_x\mathcal{M}$ in $T_{\overline{x}}\overline{\mathcal{M}}$ in order that we can use the lifted tangent vector representation unambiguously in numerical computations. Recall from Proposition 3.4.4 that the equivalence class $\pi^{-1}(x)$ is an embedded submanifold of $\overline{\mathcal{M}}$. Hence $\pi^{-1}(x)$ admits a tangent space

$$\mathcal{V}_{\overline{x}} = T_{\overline{x}}(\pi^{-1}(x))$$

called the *vertical space* at $\overline{x}$. A mapping $\mathcal{H}$ that assigns to each element $\overline{x}$ of $\overline{\mathcal{M}}$ a subspace $\mathcal{H}_{\overline{x}}$ of $T_{\overline{x}}\overline{\mathcal{M}}$ complementary to $\mathcal{V}_{\overline{x}}$ (i.e., such that $\mathcal{H}_{\overline{x}} \oplus \mathcal{V}_{\overline{x}} = T_{\overline{x}}\mathcal{M}$) is called a *horizontal distribution* on $\overline{\mathcal{M}}$. Given $\overline{x} \in \overline{\mathcal{M}}$, the subspace $\mathcal{H}_{\overline{x}}$ of $T_{\overline{x}}\overline{\mathcal{M}}$ is then called the *horizontal space* at $\overline{x}$; see Figure 3.8. Once $\overline{\mathcal{M}}$ is endowed with a horizontal distribution, there exists one and only one element $\overline{\xi}_{\overline{x}}$ that belongs to $\mathcal{H}_{\overline{x}}$ and satisfies $\mathrm{D}\pi(\overline{x})[\overline{\xi}_{\overline{x}}] = \xi$. This unique vector $\overline{\xi}_{\overline{x}}$ is called the *horizontal lift* of ξ at $\overline{x}$.

In particular, when the structure space is (a subset of) $\mathbb{R}^{n\times p}$, the horizontal lift $\overline{\xi}_{\overline{x}}$ is an $n \times p$ matrix, which lends itself to representation in a computer as a matrix array.

Example 3.5.4 ***Real projective space***

Recall from Section 3.4.3 that the projective space $\mathbb{RP}^{n-1}$ is the quotient $\mathbb{R}^n_*/\sim$, where $x \sim y$ if and only if there is an $\alpha \in \mathbb{R}_*$ such that $y = x\alpha$. The equivalence class of a point x of $\mathbb{R}^n_*$ is

$$[x] = \pi^{-1}(\pi(x)) = x\mathbb{R}_* := \{x\alpha : \alpha \in \mathbb{R}_*\}.$$

The vertical space at a point $x \in \mathbb{R}^n_*$ is

$$\mathcal{V}_x = x\mathbb{R} := \{x\alpha : \alpha \in \mathbb{R}\}.$$

A suitable choice of horizontal distribution is

$$\mathcal{H}_x := (\mathcal{V}_x)^\perp := \{z \in \mathbb{R}^n : x^T z = 0\}. \tag{3.27}$$

(This horizontal distribution will play a particular role in Section 3.6.2 where the projective space is turned into a Riemannian quotient manifold.)

A tangent vector $\xi \in T_{\pi(x)}\mathbb{RP}^{n-1}$ is represented by its horizontal lift $\overline{\xi}_x \in \mathcal{H}_x$ at a point $x \in \mathbb{R}^n_*$. It would be equally valid to use another representation $\overline{\xi}_y \in \mathcal{H}_y$ of the same tangent vector at another point $y \in \mathbb{R}^n_*$ such that $x \sim y$. The two representations $\overline{\xi}_x$ and $\overline{\xi}_y$ are not equal as vectors in $\mathbb{R}^n$ but are related by a scaling factor, as we now show. First, note that $x \sim y$ if and only if there exists a nonzero scalar α such that $y = \alpha x$. Let $f : \mathbb{RP}^{n-1} \to \mathbb{R}$ be an arbitrary smooth function and define $\overline{f} := f \circ \pi : \mathbb{R}^n_* \to \mathbb{R}$. Consider the function $g : x \mapsto \alpha x$, where α is an arbitrary nonzero scalar. Since

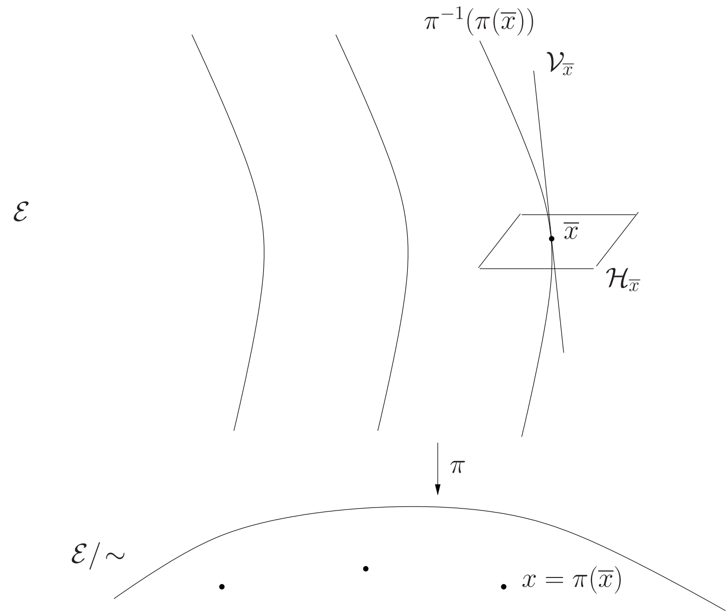


Figure 3.8 Schematic illustration of a quotient manifold. An equivalence class $\pi^{-1}(\pi(\overline{x}))$ is pictured as a subset of the total space $\mathcal{E}$ and corresponds to the single point $\pi(\overline{x})$ in the quotient manifold $\mathcal{E}/\sim$. At $\overline{x}$, the tangent space to the equivalence class is the vertical space $\mathcal{V}_{\overline{x}}$, and the horizontal space $\mathcal{H}_{\overline{x}}$ is chosen as a complement of the vertical space.

$\pi(g(x)) = \pi(x)$ for all x, we have $\overline{f}(g(x)) = \overline{f}(x)$ for all x, and it follows by taking the differential of both sides that

$$\mathrm{D}\overline{f}(g(x))[\mathrm{D}g(x)[\overline{\xi}_x]] = \mathrm{D}\overline{f}(x)[\overline{\xi}_x]. \tag{3.28}$$

By the definition of $\overline{\xi}_x$, we have $\mathrm{D}\overline{f}(x)[\overline{\xi}_x] = \mathrm{D}f(\pi(x))[\xi]$. Moreover, we have $\mathrm{D}g(x)[\overline{\xi}_x] = \alpha\overline{\xi}_x$. Thus (3.28) yields $\mathrm{D}\overline{f}(\alpha x)[\alpha\overline{\xi}_x]] = \mathrm{D}f(\pi(\alpha x))[\xi]$. This result, since it is valid for any smooth function f, implies that $\mathrm{D}\pi(\alpha x)[\alpha\overline{\xi}_x] = \xi$. This, along with the fact that $\alpha\overline{\xi}_x$ is an element of $\mathcal{H}_{\alpha x}$, implies that $\alpha\overline{\xi}_x$ is the horizontal lift of ξ at αx, i.e.,

$$\overline{\xi}_{\alpha x} = \alpha\overline{\xi}_x.$$

Example 3.5.5 ***Grassmann manifolds***

Tangent vectors to the Grassmann manifolds and their matrix representations are presented in Section 3.6.

3.6 RIEMANNIAN METRIC, DISTANCE, AND GRADIENTS

Tangent vectors on manifolds generalize the notion of a directional derivative. In order to characterize which direction of motion from x produces the steepest increase in f, we further need a notion of length that applies to tangent vectors. This is done by endowing every tangent space $T_x\mathcal{M}$ with an *inner product* $\langle \cdot, \cdot \rangle_x$, i.e., a bilinear, symmetric positive-definite form. The inner product $\langle \cdot, \cdot \rangle_x$ induces a norm,

$$\|\xi_x\|_x := \sqrt{\langle \xi_x, \xi_x \rangle_x},$$

on $T_x\mathcal{M}$. (The subscript x may be omitted if there is no risk of confusion.) The (normalized) direction of steepest ascent is then given by

$$\arg\max_{\xi \in T_x\mathcal{M} : \|\xi_x\|=1} \mathrm{D} f(x)[\xi_x].$$

A manifold whose tangent spaces are endowed with a smoothly varying inner product is called a *Riemannian manifold*. The smoothly varying inner product is called the *Riemannian metric*. We will use interchangeably the notation

$$g(\xi_x, \zeta_x) = g_x(\xi_x, \zeta_x) = \langle \xi_x, \zeta_x \rangle = \langle \xi_x, \zeta_x \rangle_x$$

to denote the inner product of two elements ξ_x and ζ_x of $T_x\mathcal{M}$. Strictly speaking, a Riemannian manifold is thus a couple $(\mathcal{M}, g)$, where $\mathcal{M}$ is a manifold and g is a Riemannian metric on $\mathcal{M}$. Nevertheless, when the Riemannian metric is unimportant or clear from the context, we simply talk about "the Riemannian manifold $\mathcal{M}$". A vector space endowed with an inner product is a particular Riemannian manifold called *Euclidean space*. Any (second-countable Hausdorff) manifold admits a Riemannian structure.

Let $(\mathcal{U}, \varphi)$ be a chart of a Riemannian manifold $(\mathcal{M}, g)$. The components of g in the chart are given by

$$g_{ij} := g(E_i, E_j),$$

where E_i denotes the ith coordinate vector field (see Section 3.5.4). Thus, for vector fields $\xi = \sum_i \xi^i E_i$ and $\zeta = \sum_i \zeta^i E_i$, we have

$$g(\xi, \zeta) = \langle \xi, \eta \rangle = \sum_{i,j} g_{ij} \xi^i \zeta^j.$$

Note that the g_{ij}'s are real-valued functions on $\mathcal{U} \subseteq \mathcal{M}$. One can also define the real-valued functions $g_{ij} \circ \varphi^{-1}$ on $\varphi(\mathcal{U}) \subseteq \mathbb{R}^d$; we use the same notation g_{ij} for both. We also use the notation $G : \hat{x} \mapsto G_{\hat{x}}$ for the matrix-valued function such that the (i,j) element of $G_{\hat{x}}$ is $g_{ij}|_{\hat{x}}$. If we let $\hat{\xi}_{\hat{x}} = \mathrm{D}\varphi\left(\varphi^{-1}(\hat{x})\right)[\xi_x]$ and $\hat{\zeta}_{\hat{x}} = \mathrm{D}\varphi\left(\varphi^{-1}(\hat{x})\right)[\zeta_x]$, with $\hat{x} = \varphi(x)$, denote the representations of ξ_x and ζ_x in the chart, then we have, in matrix notation,

$$g(\xi_x, \zeta_x) = \langle \xi_x, \zeta_x \rangle = \hat{\xi}_{\hat{x}}^T G_{\hat{x}} \hat{\zeta}_{\hat{x}}. \tag{3.29}$$

Note that G is a symmetric, positive definite matrix at every point.

The *length of a curve* $\gamma : [a, b] \to \mathcal{M}$ on a Riemannian manifold $(\mathcal{M}, g)$ is defined by

$$L(\gamma) = \int_a^b \sqrt{g(\dot{\gamma}(t), \dot{\gamma}(t))}\, \mathrm{d}t.$$

The *Riemannian distance* on a connected Riemannian manifold $(\mathcal{M}, g)$ is

$$\mathrm{dist} : \mathcal{M} \times \mathcal{M} \to \mathbb{R} : \mathrm{dist}(x, y) = \inf_{\Gamma} L(\gamma) \tag{3.30}$$

where Γ is the set of all curves in $\mathcal{M}$ joining points x and y. Assuming (as usual) that $\mathcal{M}$ is Hausdorff, it can be shown that the Riemannian distance defines a *metric*; i.e.,

1. $\mathrm{dist}(x, y) \geq 0$, with $\mathrm{dist}(x, y) = 0$ if and only if $x = y$ (positive-definiteness);
2. $\mathrm{dist}(x, y) = \mathrm{dist}(y, x)$ (symmetry);
3. $\mathrm{dist}(x, z) + \mathrm{dist}(z, y) \geq \mathrm{dist}(x, y)$ (triangle inequality).

Metrics and Riemannian metrics should not be confused. A metric is an abstraction of the notion of distance, whereas a Riemannian metric is an inner product on tangent spaces. There is, however, a link since any Riemannian metric induces a distance, the Riemannian distance.

Given a smooth scalar field f on a Riemannian manifold $\mathcal{M}$, the *gradient* of f at x, denoted by $\mathrm{grad}\, f(x)$, is defined as the unique element of $T_x\mathcal{M}$ that satisfies

$$\langle \mathrm{grad}\, f(x), \xi \rangle_x = \mathrm{D} f\,(x)\,[\xi]\,, \quad \forall \xi \in T_x\mathcal{M}. \tag{3.31}$$

The coordinate expression of $\mathrm{grad}\, f$ is, in matrix notation,

$$\widehat{\mathrm{grad}\, f}(\hat{x}) = G_{\hat{x}}^{-1}\, \mathrm{Grad}\, \widehat{f}(\hat{x}), \tag{3.32}$$

where G is the matrix-valued function defined in (3.29) and Grad denotes the *Euclidean gradient* in $\mathbb{R}^d$,

$$\mathrm{Grad}\, \widehat{f}(\hat{x}) := \begin{pmatrix} \partial_1 \widehat{f}(\hat{x}) \\ \vdots \\ \partial_d \widehat{f}(\hat{x}) \end{pmatrix}.$$

(Indeed, from (3.29) and (3.32), we have $\langle \mathrm{grad}\, f, \xi \rangle = \hat{\xi}^T G(G^{-1}\, \mathrm{Grad}\, \widehat{f}) = \hat{\xi}^T\, \mathrm{Grad}\, \widehat{f} = \mathrm{D}\widehat{f}[\hat{\xi}] = \mathrm{D}f[\xi]$ for any vector field ξ.)

The gradient of a function has the following remarkable steepest-ascent properties (see Figure 3.9):

- The direction of $\mathrm{grad}\, f(x)$ is the steepest-ascent direction of f at x:
$$\frac{\mathrm{grad}\, f(x)}{\|\mathrm{grad}\, f(x)\|} = \underset{\xi \in T_x\mathcal{M} : \|\xi\| = 1}{\arg\max}\ \mathrm{D} f\,(x)\,[\xi]\,.$$
- The norm of $\mathrm{grad}\, f(x)$ gives the steepest slope of f at x:
$$\|\mathrm{grad}\, f(x)\| = \mathrm{D} f\,(x) \left[\frac{\mathrm{grad}\, f(x)}{\|\mathrm{grad}\, f(x)\|}\right].$$

These two properties are important in the scope of optimization methods.

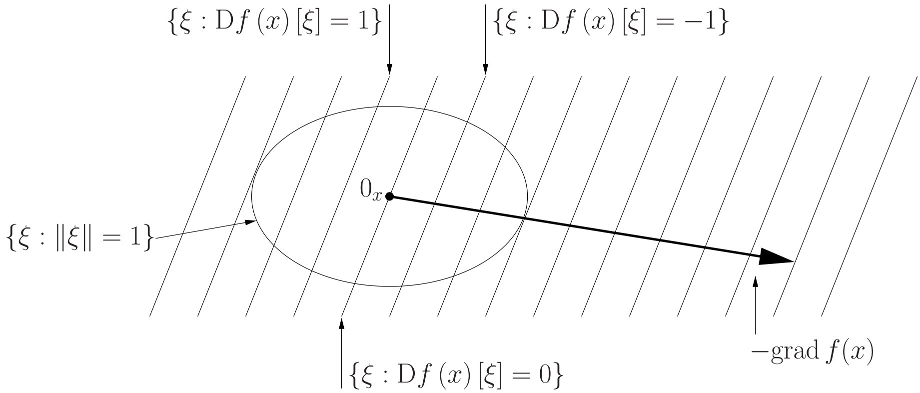


Figure 3.9 Illustration of steepest descent.

3.6.1 Riemannian submanifolds

If a manifold $\overline{\mathcal{M}}$ is endowed with a Riemannian metric, one would expect that manifolds generated from $\overline{\mathcal{M}}$ (such as submanifolds and quotient manifolds) can inherit a Riemannian metric in a natural way. This section considers the case of embedded submanifolds; quotient manifolds are dealt with in the next section.

Let $\mathcal{M}$ be an embedded submanifold of a Riemannian manifold $\overline{\mathcal{M}}$. Since every tangent space $T_x\mathcal{M}$ can be regarded as a subspace of $T_x\overline{\mathcal{M}}$, the Riemannian metric $\overline{g}$ of $\overline{\mathcal{M}}$ induces a Riemannian metric g on $\mathcal{M}$ according to

$$g_x(\xi, \zeta) = \overline{g}_x(\xi, \zeta), \quad \xi,\ \zeta \in T_x\mathcal{M},$$

where ξ and ζ on the right-hand side are viewed as elements of $T_x\overline{\mathcal{M}}$. This turns $\mathcal{M}$ into a Riemannian manifold. Endowed with this Riemannian metric, $\mathcal{M}$ is called a *Riemannian submanifold* of $\overline{\mathcal{M}}$. The orthogonal complement of $T_x\mathcal{M}$ in $T_x\overline{\mathcal{M}}$ is called the *normal space to* $\mathcal{M}$ *at* x and is denoted by $(T_x\mathcal{M})^\perp$:

$$(T_x\mathcal{M})^\perp = \{\xi \in T_x\overline{\mathcal{M}} : \overline{g}_x(\xi, \zeta) = 0 \text{ for all } \zeta \in T_x\mathcal{M}\}.$$

Any element $\xi \in T_x\overline{\mathcal{M}}$ can be uniquely decomposed into the sum of an element of $T_x\mathcal{M}$ and an element of $(T_x\mathcal{M})^\perp$:

$$\xi = \mathrm{P}_x\xi + \mathrm{P}_x^\perp\xi,$$

where P_x denotes the orthogonal projection onto $T_x\mathcal{M}$ and $\mathrm{P}_x^\perp$ denotes the orthogonal projection onto $(T_x\mathcal{M})^\perp$.

Example 3.6.1 ***Sphere***

On the unit sphere S^{n-1} considered a Riemannian submanifold of $\mathbb{R}^n$, the inner product inherited from the standard inner product on $\mathbb{R}^n$ is given by

$$\langle \xi, \eta \rangle_x := \xi^T \eta. \tag{3.33}$$

The normal space is

$$(T_x S^{n-1})^\perp = \{x\alpha : \alpha \in \mathbb{R}\},$$

and the projections are given by

$$\mathrm{P}_x \xi = (I - xx^T)\xi, \qquad \mathrm{P}_x^\perp \xi = xx^T \xi$$

for $x \in S^{n-1}$.

Example 3.6.2 ***Orthogonal Stiefel manifold***

Recall that the tangent space to the orthogonal Stiefel manifold $\mathrm{St}(p, n)$ *is*

$$T_X \, \mathrm{St}(p, n) = \{X\Omega + X_\perp K : \Omega^T = -\Omega, \ K \in \mathbb{R}^{(n-p)\times p}\}.$$

The Riemannian metric inherited from the embedding space $\mathbb{R}^{n\times p}$ *is*

$$\langle \xi, \eta \rangle_X := \mathrm{tr}(\xi^T \eta). \tag{3.34}$$

If $\xi = X\Omega_\xi + X_\perp K_\xi$ *and* $\eta = X\Omega_\eta + X_\perp K_\eta$, *then* $\langle \xi, \eta \rangle_X = \mathrm{tr}(\Omega_\xi^T \Omega_\eta + K_\xi^T K_\eta)$. *In view of the identity* $\mathrm{tr}(S^T\Omega) = 0$ *for all* $S \in \mathcal{S}_{\mathrm{sym}}(p)$, $\Omega \in \mathcal{S}_{\mathrm{skew}}(p)$, *the normal space is*

$$(T_X \, \mathrm{St}(p, n))^\perp = \{XS : S \in \mathcal{S}_{\mathrm{sym}}(p)\}.$$

The projections are given by

$$\mathrm{P}_X \xi = (I - XX^T)\xi + X \,\mathrm{skew}(X^T\xi), \tag{3.35}$$

$$\mathrm{P}_X^\perp \xi = X \,\mathrm{sym}(X^T \xi), \tag{3.36}$$

where $\mathrm{sym}(A) := \frac{1}{2}(A + A^T)$ *and* $\mathrm{skew}(A) := \frac{1}{2}(A - A^T)$ *denote the components of the decomposition of* A *into the sum of a symmetric term and a skew-symmetric term.*

Let $\overline{f}$ be a cost function defined on a Riemannian manifold $\overline{\mathcal{M}}$ and let f denote the restriction of $\overline{f}$ to a Riemannian submanifold $\mathcal{M}$. The gradient of f is equal to the projection of the gradient of $\overline{f}$ onto $T_x\mathcal{M}$:

$$\mathrm{grad}\, f(x) = \mathrm{P}_x \,\mathrm{grad}\, \overline{f}(x). \tag{3.37}$$

Indeed, $\mathrm{P}_x \,\mathrm{grad}\,\overline{f}(x)$ belongs to $T_x\mathcal{M}$ and (3.31) is satisfied since, for all $\zeta \in T_x\mathcal{M}$, we have $\langle \mathrm{P}_x \,\mathrm{grad}\,\overline{f}(x), \zeta\rangle = \langle \mathrm{grad}\,\overline{f}(x) - \mathrm{P}_x^\perp \,\mathrm{grad}\,\overline{f}(x), \zeta\rangle = \langle \mathrm{grad}\,\overline{f}(x), \zeta\rangle = \mathrm{D}\overline{f}\,(x)\,[\zeta] = \mathrm{D}f\,(x)\,[\zeta]$.

3.6.2 Riemannian quotient manifolds

We now consider the case of a quotient manifold $\mathcal{M} = \overline{\mathcal{M}}/\sim$, where the structure space $\overline{\mathcal{M}}$ is endowed with a Riemannian metric $\overline{g}$. The horizontal space $\mathcal{H}_{\overline{x}}$ at $\overline{x} \in \overline{\mathcal{M}}$ is canonically chosen as the orthogonal complement in $T_{\overline{x}}\overline{\mathcal{M}}$ of the vertical space $\mathcal{V}_{\overline{x}} = T_{\overline{x}}\pi^{-1}(x)$, namely,

$$\mathcal{H}_{\overline{x}} := (T_{\overline{x}}\mathcal{V}_{\overline{x}})^\perp = \{\eta_{\overline{x}} \in T_{\overline{x}}\overline{\mathcal{M}} : \overline{g}(\chi_{\overline{x}}, \eta_{\overline{x}}) = 0 \text{ for all } \chi_{\overline{x}} \in \mathcal{V}_{\overline{x}}\}.$$

Recall that the horizontal lift at $\overline{x} \in \pi^{-1}(x)$ of a tangent vector $\xi_x \in T_x\mathcal{M}$ is the unique tangent vector $\overline{\xi}_{\overline{x}} \in \mathcal{H}_{\overline{x}}$ that satisfies $\mathrm{D}\pi(\overline{x})[\overline{\xi}_{\overline{x}}]$. If, for every

$x \in \mathcal{M}$ and every $\xi_x,\ \zeta_x \in T_x\mathcal{M}$, the expression $\overline{g}_{\overline{x}}(\overline{\xi}_{\overline{x}}, \overline{\zeta}_{\overline{x}})$ does *not* depend on $\overline{x} \in \pi^{-1}(x)$, then

$$g_x(\xi_x, \zeta_x) := \overline{g}_{\overline{x}}(\overline{\xi}_{\overline{x}}, \overline{\zeta}_{\overline{x}}) \tag{3.38}$$

defines a Riemannian metric on $\mathcal{M}$. Endowed with this Riemannian metric, $\mathcal{M}$ is called a *Riemannian quotient manifold* of $\overline{\mathcal{M}}$, and the natural projection $\pi : \overline{\mathcal{M}} \to \mathcal{M}$ is a *Riemannian submersion.* (In other words, a Riemannian submersion is a submersion of Riemannian manifolds such that $\mathrm{D}\pi$ preserves inner products of vectors normal to fibers.)

Riemannian quotient manifolds are interesting because several differential objects on the quotient manifold can be represented by corresponding objects in the structure space in a natural manner (see in particular Section 5.3.4). Notably, if $\overline{f}$ is a function on $\overline{\mathcal{M}}$ that induces a function f on $\mathcal{M}$, then one has

$$\overline{\operatorname{grad} f}_{\overline{x}} = \operatorname{grad} \overline{f}(\overline{x}). \tag{3.39}$$

Note that $\operatorname{grad} \overline{f}(\overline{x})$ belongs to the horizontal space: since $\overline{f}$ is constant on each equivalence class, it follows that $\overline{g}_{\overline{x}}(\operatorname{grad} \overline{f}(\overline{x}), \xi) \equiv \mathrm{D}\overline{f}\,(\overline{x})\,[\xi] = 0$ for all vertical vectors ξ, hence $\operatorname{grad} \overline{f}(\overline{x})$ is orthogonal to the vertical space.

We use the notation $\mathrm{P}^h_{\overline{x}}\xi_{\overline{x}}$ and $\mathrm{P}^v_{\overline{x}}\xi_{\overline{x}}$ for the projection of $\xi_{\overline{x}} \in T_{\overline{x}}\overline{\mathcal{M}}$ onto $\mathcal{H}_{\overline{x}}$ and $\mathcal{V}_{\overline{x}}$.

Example 3.6.3 ***Projective space***

On the projective space $\mathbb{RP}^{n-1}$*, the definition*

$$\langle \xi, \eta \rangle_{x\mathbb{R}} := \frac{1}{x^T x} \overline{\xi}_x^{\,T} \overline{\eta}_x$$

turns the canonical projection $\pi : \mathbb{R}^n_* \to \mathbb{RP}^{n-1}$ *into a Riemannian submersion.*

Example 3.6.4 ***Grassmann manifolds***

We show that the Grassmann manifold $\operatorname{Grass}(p, n) = \mathbb{R}^{n\times p}_*/\mathrm{GL}_p$ *admits a structure of a Riemannian quotient manifold when* $\mathbb{R}^{n\times p}_*$ *is endowed with the Riemannian metric*

$$\overline{g}_Y(Z_1, Z_2) = \operatorname{tr}\left((Y^TY)^{-1}Z_1^TZ_2\right).$$

The vertical space at Y *is by definition the tangent space to the equivalence class* $\pi^{-1}(\pi(Y)) = \{YM : M \in \mathbb{R}^{p\times p}_*\}$*, which yields*

$$\mathcal{V}_Y = \{YM : M \in \mathbb{R}^{p\times p}\}.$$

The horizontal space at Y *is then defined as the orthogonal complement of the vertical space with respect to the metric* $\overline{g}$*. This yields*

$$\mathcal{H}_Y = \{Z \in \mathbb{R}^{n\times p} : Y^TZ = 0\}, \tag{3.40}$$

and the orthogonal projection onto the horizontal space is given by

$$\mathrm{P}^h_Y Z = (I - Y(Y^TY)^{-1}Y^T)Z. \tag{3.41}$$

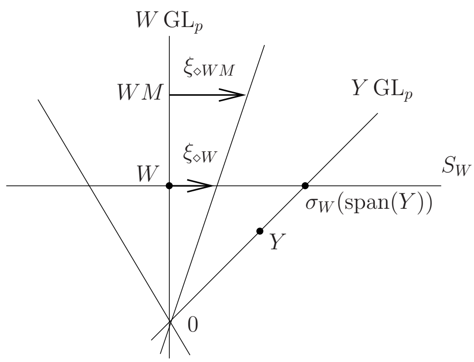


Figure 3.10 $\mathrm{Grass}(p,n)$ is shown as the quotient $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ for the case $p=1$, $n=2$. Each point, the origin excepted, is an element of $\mathbb{R}_*^{n\times p}=\mathbb{R}^2-\{0\}$. Each line is an equivalence class of elements of $\mathbb{R}_*^{n\times p}$ that have the same span. So each line through the origin corresponds to an element of $\mathrm{Grass}(p,n)$. The affine subspace S_W is an affine cross section as defined in (3.43). The relation (3.42) satisfied by the horizontal lift $\overline{\xi}$ of a tangent vector $\xi \in T_{\mathcal{W}}\,\mathrm{Grass}(p,n)$ is also illustrated. This figure can help to provide insight into the general case, however, one nonetheless has to be careful when drawing conclusions from it. For example, in general there does not exist a submanifold of $\mathbb{R}^{n\times p}$ that is orthogonal to the fibers $Y\,\mathrm{GL}_p$ at each point, although it is obviously the case for $p=1$ (any centered sphere in $\mathbb{R}^n$ will do).

Given $\xi \in T_{\mathrm{span}(Y)}\,\mathrm{Grass}(p,n)$, *there exists a unique horizontal lift* $\overline{\xi}_Y \in T_Y\mathbb{R}_*^{n\times p}$ *satisfying*

$$\mathrm{D}\pi(Y)[\overline{\xi}_Y]=\xi.$$

In order to show that $\mathrm{Grass}(p,n)$ *admits a structure of a Riemannian quotient manifold of* $(\mathbb{R}_*^{n\times p},\overline{g})$, *we have to show that*

$$\overline{g}(\overline{\xi}_{YM},\overline{\zeta}_{YM})=\overline{g}(\overline{\xi}_Y,\overline{\zeta}_Y)$$

for all $M\in\mathbb{R}_*^{p\times p}$. *This relies on the following result.*

Proposition 3.6.1 *Given* $Y\in\mathbb{R}_*^{n\times p}$ *and* $\xi\in T_{\mathrm{span}(Y)}\,\mathrm{Grass}(p,n)$, *we have*

$$\overline{\xi}_{YM}=\overline{\xi}_Y\cdot M \tag{3.42}$$

for all $M\in\mathbb{R}_*^{p\times p}$, *where the center dot (usually omitted) denotes matrix multiplication.*

Proof. Let $W\in\mathbb{R}_*^{n\times p}$. Let $\mathcal{U}_W=\{\mathrm{span}(Y): W^TY \text{ invertible}\}$. Notice that $\mathcal{U}_W$ is the set of all the p-dimensional subspaces $\mathcal{Y}$ of $\mathbb{R}^n$ that do not contain any direction orthogonal to $\mathrm{span}(W)$. Consider the mapping

$$\sigma_W:\mathcal{U}_W\to\mathbb{R}_*^{n\times p}:\mathrm{span}(Y)\mapsto Y(W^TY)^{-1}W^TW;$$

see Figure 3.10. One has $\pi(\sigma_W(\mathcal{Y})) = \mathrm{span}(\sigma_W(\mathcal{Y})) = \mathcal{Y}$ *for all* $\mathcal{Y} \in \mathcal{U}_W$*; i.e.,* σ_W *is a right inverse of* π*. Consequently,* $\mathrm{D}\pi(\sigma_W(\mathcal{Y})) \circ \mathrm{D}\sigma_W(\mathcal{Y}) = \mathrm{id}$*. Moreover, the range of* σ_W *is*

$$\mathcal{S}_W := \{Y \in \mathbb{R}_*^{n\times p} : W^T(Y - W) = 0\}, \tag{3.43}$$

from which it follows that the range of $\mathrm{D}\sigma_W(\mathcal{Y}) = \{Z \in \mathbb{R}^{n\times p} : W^T Z = 0\} = \mathcal{H}_W$*. In conclusion,*

$$\mathrm{D}\sigma_W(\mathcal{W})[\xi] = \overline{\xi}_W.$$

Now, $\sigma_{WM}(\mathcal{Y}) = \sigma_W(\mathcal{Y})M$ *for all* $M \in \mathbb{R}_*^{p\times p}$ *and all* $\mathcal{Y} \in \mathcal{U}_W$*. It follows that*

$$\overline{\xi}_{WM} = \mathrm{D}\sigma_{WM}(\mathcal{W})[\xi] = \mathrm{D}(\sigma_W \cdot M)(\mathcal{W})[\xi] = \mathrm{D}\sigma_W(\mathcal{W})[\xi] \cdot M = \overline{\xi}_W \cdot M,$$

where the center dot denotes the matrix multiplication. □

Using this result, we have

$$\begin{aligned}
\overline{g}_{YM}(\overline{\xi}_{YM}, \overline{\zeta}_{YM}) &= \overline{g}_{YM}(\overline{\xi}_Y M, \overline{\zeta}_Y M) \\
&= \mathrm{tr}\left(((YM)^T YM)^{-1}(\overline{\xi}_Y M)^T(\overline{\zeta}_Y M)\right) \\
&= \mathrm{tr}\left(M^{-1}(Y^TY)^{-1}M^{-T}M^T\overline{\xi}_Y{}^T\overline{\zeta}_Y M\right) \\
&= \mathrm{tr}\left((Y^TY)^{-1}\overline{\xi}_Y{}^T\overline{\zeta}_Y\right) \\
&= \overline{g}_Y(\overline{\xi}_Y, \overline{\zeta}_Y).
\end{aligned}$$

This shows that $\mathrm{Grass}(p, n)$*, endowed with the Riemannian metric*

$$g_{\mathrm{span}(Y)}(\xi, \zeta) := \overline{g}_Y(\overline{\xi}_Y, \overline{\zeta}_Y), \tag{3.44}$$

is a Riemannian quotient manifold of $(\mathbb{R}_*^{n\times p}, \overline{g})$*. In other words, the canonical projection* $\pi : \mathbb{R}_*^{n\times p} \to \mathrm{Grass}(p, n)$ *is a Riemannian submersion from* $(\mathbb{R}_*^{n\times p}, \overline{g})$ *to* $(\mathrm{Grass}(p, n), g)$*.*

3.7 NOTES AND REFERENCES

Differential geometry textbooks that we have referred to when writing this book include Abraham *et al.* [AMR88], Boothby [Boo75], Brickell and Clark [BC70], do Carmo [dC92], Kobayashi and Nomizu [KN63], O'Neill [O'N83], Sakai [Sak96], and Warner [War83]. Some material was also borrowed from the course notes of M. De Wilde at the University of Liège [DW92]. Do Carmo [dC92] is well suited for engineers, as it does not assume any background in abstract topology; the prequel [dC76] on the differential geometry of curves and surfaces makes the introduction even smoother. Abraham *et al.* [AMR88] and Brickell and Clark [BC70] cover global analysis questions (submanifolds, quotient manifolds) at an introductory level. Brickell and Clark [BC70] has a detailed treatment of the topology of manifolds. O'Neill [O'N83] is an excellent reference for Riemannian connections of submanifolds and quotient manifolds (Riemannian submersions). Boothby [Boo75] provides an excellent introduction to differential geometry with a perspective on Lie theory, and Warner [War83] covers more advanced material in this direction. Other references on differential geometry include the classic works of Kobayashi and Nomizu [KN63], Helgason [Hel78], and Spivak [Spi70]. We also mention Darling [Dar94], which introduces abstract manifold theory only after covering Euclidean spaces and their submanifolds.

Several equivalent ways of defining a manifold can be found in the literature. The definition in do Carmo [dC92] is based on local parameterizations. O'Neill [O'N83, p. 22] points out that for a Hausdorff manifold (with countably many components), being second-countable is equivalent to being paracompact. (In abstract topology, a space X is *paracompact* if every open covering of X has a locally finite open refinement that covers X.) A differentiable manifold $\mathcal{M}$ admits a partition of unity if and only if it is paracompact [BC70, Th. 3.4.4]. The material on the existence and uniqueness of atlases has come chiefly from Brickell and Clark [BC70]. A function with constant rank on its domain is called a *subimmersion* in most textbooks. The terms "canonical immersion" and "canonical submersion" have been borrowed from Guillemin and Pollack [GP74, p. 14]. The manifold topology of an immersed submanifold is always finer than its topology as a subspace [BC70], but they need not be the same topology. (When they are, the submanifold is called *embedded*.) Examples of subsets of a manifold that do not admit a submanifold structure, and examples of immersed submanifolds that are not embedded, can be found in most textbooks on differential geometry, such as do Carmo [dC92]. Proposition 3.3.1, on the uniqueness of embedded submanifold structures, is proven in Brickell and Clark [BC70] and O'Neill [O'N83]. Proposition 3.3.3 can be found in several textbooks without the condition $d_1 > d_2$. In the case where $d_1 = d_2$, $F^{-1}(y)$ is a discrete set of points [BC70, Prop. 6.2.1]. In several references, embedded submanifolds are called *regular submanifolds* or simply *submanifolds*. Proposition 3.3.2, on coordinate slices, is sometimes used to define the notion of an embedded submanifold, such as in Abraham *et al.* [AMR88]. Our definition of a regular equivalence relation follows that of Abraham *et al.* [AMR88]. The characterization of quotient manifolds in Proposition 3.4.2 can be found in Abraham *et al.* [AMR88, p. 208]. A shorter proof of Proposition 3.4.6 (showing that $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ admits a structure of quotient manifold, the Grassmann manifold) can be given using the theory of homogeneous spaces, see Boothby [Boo75] or Warner [War83].

Most textbooks define tangent vectors as derivations. Do Carmo [dC92] introduces tangent vectors to curves, as in Section 3.5.1. O'Neill [O'N83] proposes both definitions. A tangent vector at a point x of a manifold can also be defined as an equivalence class of all curves that realize the same derivation: $\gamma_1 \sim \gamma_2$ if and only if, in a chart (U, φ) around $x = \gamma_1(0) = \gamma_2(0)$, we have $(\varphi \circ \gamma_1)'(0) = (\varphi \circ \gamma_2)'(0)$. This notion does not depend on the chart since, if $(\mathcal{V}, \psi)$ is another chart around x, then

$$(\psi \circ \gamma)'(0) = (\psi \circ \varphi^{-1})'(\varphi(m)) \cdot (\varphi \circ \gamma)'(0).$$

This is the approach taken, for example, by Gallot *et al.* [GHL90].

The notation $\mathrm{D}F\,(x)\,[\xi]$ is not standard. Most textbooks use $dF_x\xi$ or $F_{*x}\xi$. Our notation is slightly less compact but makes it easier to distinguish the three elements F, x, and ξ of the expression and has proved more flexible when undertaking explicit computations involving matrix manifolds.

An alternative way to define smoothness of a vector field is to require that the function ξf be smooth for every $f \in \mathfrak{F}(\mathcal{M})$; see O'Neill [O'N83]. In the parlance of abstract algebra, the set $\mathfrak{F}(\mathcal{M})$ of all smooth real-valued functions on $\mathcal{M}$, endowed with the usual operations of addition and multiplication, is a *commutative ring*, and the set $\mathfrak{X}(\mathcal{M})$ of vector fields is a *module* over $\mathfrak{F}(\mathcal{M})$ [O'N83]. Formula (3.26) for the tangent space to the orthogonal group can also be obtained by treating O_n as a Lie group: the operation of left multiplication by U, $L_U : X \mapsto UX$, sends the neutral element I to U, and the differential of L_U at I sends $T_I O_n = \mathfrak{o}(n) = \mathcal{S}_{\mathrm{skew}}(n)$ to $U\mathcal{S}_{\mathrm{skew}}(n)$; see, e.g., Boothby [Boo75] or Warner [War83]. For a proof that the Riemannian distance satisfies the three axioms of a metric, see O'Neill [O'N83, Prop. 5.18]. The axiom that fails to hold in general for non-Hausdorff manifolds is that $\mathrm{dist}(x, y) = 0$ if and only if $x = y$. An example can be constructed from the material in Section 4.3.2. Riemannian submersions are covered in some detail in Cheeger and Ebin [CE75], do Carmo [dC92], Klingenberg [Kli82], O'Neill [O'N83], and Sakai [Sak96]. The term "Riemannian quotient manifold" is new.

The Riemannian metric given in (3.44) is the essentially unique rotation-invariant Riemannian metric on the Grassmann manifold [Lei61, AMS04]. More information on Grassmann manifolds can be found in Ferrer *et al.* [FGP94], Edelman *et al.* [EAS98], Absil *et al.* [AMS04], and references therein.

In order to define the steepest-descent direction of a real-valued function f on a manifold $\mathcal{M}$, it is enough to endow the tangent spaces to $\mathcal{M}$ with a norm. Under smoothness assumptions, this turns $\mathcal{M}$ into a *Finsler manifold*. Finsler manifolds have received little attention in the literature in comparison with the more restrictive notion of Riemannian manifolds. For recent work on Finsler manifolds, see Bao *et al.* [BCS00].